

119

M 1996 DCB 45 Vol. 2

Varon Eric  
DCB-5

Notes de lecture, propositions de recherches et bibliographie  
concernant les Postes de Lecture Assistée par Ordinateur à la  
Bibliothèque Nationale de France.

BIBLIOTHEQUE DE L'ENSIB



810247

Les Postes de Lecture Assistée par Ordinateur  
à la Bibliothèque Nationale de France :  
aspects théoriques et expérimentation

## Table des matières :

### Avant-propos

<b>1. Introduction</b>	<b>1</b>
1.1. Le programme de numérisation	9
1.2. Conséquences des réflexions sur le choix des textes à numériser et sur les fonctions des postes de lecture	10
1.3. Les fonctions centrales : la lecture assistée par ordinateur	13
1.4. L'annotation dynamique	14
1.5. Pratique et définition de la lecture assistée par ordinateur	18
1.6. Personnalisation du document	24
1.7. La lecture / écriture, la différence phénoménologique et la transduction	26
1.8. La question de l'hylémorphisme	28
<b>2. Objectif : l'assistance au travail intellectuel</b>	<b>30</b>
2.1. Marketing des Postes de Lecture Assistée par Ordinateur	33
2.2. L'aspect institutionnel	36
2.3. Relation avec d'autres recherches actuelles : la notion d'hypertexte	41
<b>3. Nouvelles technologies et culture humaine : hypertexte, mémoire et érudition contemporaine</b>	<b>45</b>
3.1. PLAO et hypertexte	46
3.2. Condition de production de l'hypertexte	51
3.2.1. Le modèle hypertexte propose un mode particulier d'accès au texte et à l'information	52
3.3. PLAO et études historiques	58
<b>4. L'insertion des Postes de Lecture Assistée par Ordinateur dans l'univers des réseaux internationaux</b>	<b>60</b>
4.1. Document électronique	62

4.1.1. Hypertexte : l'aspect historique	64
4.1.2. Les principaux points de repères historiques	65
4.2. Implémentation des Postes de Lecture Assistée par Ordinateur	68
4.3. Postes de Lecture Assistée par Ordinateur et édition électronique	70
4.3.1 La sémantique de présentation de documents et le langage de spécification	74
5. Poste de Lecture Assistée par Ordinateur et culture post-moderne	81
5.1. Rationalisation des liens et lecture assistée par ordinateur, normes SGML et langage HTML	84
5.1.1. Méthodes littéraires et traitement automatique du texte	87
a) mise en abyme	
b) déconstruction	
6. État actuel du projet	91
6.1. La coopération avec les institutions et les éditeurs	92
7. Cartographier et analyser les textes numérisés : Bibliométrie, Scientométrie, Infométrie et Postes de Lecture Assistée par Ordinateur	93
<u>Conclusion et propositions de recherches</u>	94

## Avant-propos :

La numérisation d'ouvrages en texte intégral entraîne avec elle une nouvelle conception de la lecture et de l'écriture. Les textes sont conservés et transmis sur des supports autres que le livre ou le périodique. L'historien de la culture écrite doit être attentif à ce phénomène qui s'emplifie rapidement, modifiant les modes de transmission de la pensée et les habitudes du travail intellectuel.

Le projet de la Bibliothèque Nationale de France, comprend depuis les origines (*Propositions pour une grande bibliothèque* de Messieurs Patrice Cahart et Michel Melot, 1989), la constitution d'un important fonds de textes numérisés qui pourront être l'objet d'un nouveau type de lecture, rendue possible par les Postes de Lecture Assistée par Ordinateur (PLAO). L'introduction de grandes quantités de textes électroniques, modifiera les habitudes de lecture. Lire sur écran n'est pas lire dans un codex. Si elle ouvre des possibilités neuves et immenses, la représentation électronique des textes modifie totalement leur utilisation : à la matérialité du livre, elle substitue l'immatérialité du texte sans lieu propre ; aux relations de contiguïté établie dans l'objet imprimé, elle oppose la libre composition de fragments indéfiniment manipulables ; à la saisie immédiate de la totalité de l'oeuvre, rendue visible par l'objet qui la contient, elle fait succéder la navigation au long cours dans des archipels textuels. Ces mutations commandent, inévitablement, impérativement, de nouvelles manières de lire, de nouveaux rapports à l'écrit, de nouvelles techniques intellectuelles. Parmi celles-ci : l'hypertexte mérite une attention particulière. Le domaine d'un hypertexte représente l'espace informationnel sur lequel l'auteur se propose de travailler : constituer un domaine d'information consiste à travailler avec l'implicite inhérent à toute connaissance. Avec les textes numérisés traités comme base d'hypertextes, on peut à partir d'une page en cours de lecture, accéder de façon non-linéaire à un ensemble d'autres informations présentes, par exemple à des commentaires, à des informations historiques, géographiques, linguistiques, créer ses propres espaces d'annotation ou de commentaire.

Contrairement à ce qui existe dans la plupart des autres médias (livres, films, bandes sonores, disques, etc...), l'information dans les hypertextes, n'a pas besoin d'être stockée de façon massivement linéaire. Au contraire son principe fondateur est qu'une information peut être conservée, sous forme d'enregistrements séparés, de blocs (chunks), n'importe où dans l'espace de mémorisation et lue suivant des parcours multiples.

Le réseau Internet utilise l'hypertexte par l'intermédiaire de divers outils de navigation (*Uniform Resource Locator, hypertexte transfer protocol, Hypertext Markup Language*), qui permettent de rechercher l'information à distance avec des mots-clés. Dans le domaine de la littérature grise, c'est-à-dire de l'ensemble des documents comme les rapports, les thèses, les actes de congrès qui échappe aux circuits éditoriaux commerciaux, et représente une source extrêmement riche d'informations inédites, l'assistance au travail intellectuel est mise en oeuvre dans le cadre du programme *Griseli* par huit pôles associés utilisant des logiciels éditeurs SGML qui permettent une DTD (définition de type de document normalisée sur le plan international, norme ISO 12083), et utilisent le langage Hytime qui inclut la norme de balisage et de description SGML dans un contexte hypermédia, ouvrant l'utilisation de ces documents aux applications multimédia.

Les précédentes révolutions de la lecture sont advenues alors que ne changeaient pas les structures fondamentales du livre. La révolution entamée est avant tout, une révolution des supports et des formes qui transmettent l'écrit.

La numérisation des documents répond à plusieurs besoins. On peut numériser pour archiver des documents, améliorer l'accès à certains documents, protéger les collections. Après numérisation, il est possible d'effectuer sur le document un certain nombre de traitements sur lesquels nous reviendrons.

Cette révolution dans le domaine des supports n'a qu'un précédent dans le domaine du travail intellectuel : la substitution du codex au volumen, du livre composé de cahiers assemblés, au livre en forme de rouleau, aux premiers siècles de l'ère chrétienne. Cette révolution rencontre une idée essentielle : la lecture de recherche n'est guère dissociable de l'écriture. Tel est le sens des études sur les Postes de Lecture Assistée par Ordinateur à la Bibliothèque Nationale de France qui conduisent à l'idée d'annotation dynamique. Des logiciels tenant compte de ces données fondamentales, conçus comme moyen d'exploitation des corpus numérisés, mais aussi en tant qu'aide dynamique au travail intellectuel, seront utilisés en liaison avec les bases de données, permettant au lecteur d'effectuer depuis son Poste de Lecture Assistée par Ordinateur de multiples opérations : indexation, annotation, copie.

Des réflexions ont été conduites depuis 1990, portant sur le choix des textes à numériser, et sur la notion de mode de lecture et les logiciels à utiliser. (En ce qui concerne le choix des textes numérisés l'étude devrait être poursuivie sur place).

Un premier groupe de sujets de recherche concerne :

- la présentation des Postes de Lecture Assistée par Ordinateur
- l'assistance au travail intellectuel
- l'accès facilité à l'information

Ensuite se pose la question des publics : qui est concerné, d'autres types de bibliothèques peuvent-elles utiliser ce type de poste ?

(cf. : 2.1 page 28 : "Marketing des Postes de Lecture Assistée par Ordinateur"). La constitution de collections numérisées et l'étude de postes de lecture, par des groupes d'experts issus du monde de la culture écrite, qui adaptent les nouvelles technologies au travail intellectuel et de recherche (et non l'inverse), constituent une nouvelle segmentation de l'offre, et une application du principe de servuction, service mis au point suivant le principe de la concertation entre usager et agent du service public (sur la notion de servuction, cf. : Monsieur Jean-Michel Salaün, *Le Marketing des bibliothèques*, pp. 62-63).

La pratique actuelle de transmissions des documents par photocopies et envoi postal reste une technique de transition. La transmission électronique de documents conduit dans le cadre du document structuré sous SGML à la réalisation d'une chaîne de traitement des documents de bout en bout. Les documents sont stockés sur serveurs WWW (World Wide Web), et accessibles par Internet. Leur adresse électronique est impérativement inscrite sur la notice. Dans le cadre de l'assistance au travail intellectuel proposée par Les Postes de Lecture Assistée par Ordinateur, la question de la transmission électronique de documents de recherche et d'étude, mais aussi des messages, des interventions dans les forums, les pré-publications électroniques, se pose. Or, ces éléments beaucoup plus fluides que les documents traditionnels, ne subissent aucune validation et échappent encore davantage à tout contrôle ou identification.

On notera que ce processus de servuction concernant l'assistance au travail intellectuel en est à son début et que la partie concernant les résultats de l'activité, l'évaluation de la satisfaction de l'utilisateur reste à venir. Ce secteur de recherche n'est pas un sujet d'étude pour les entreprises commerciales qui investissent dans ce qu'il est convenu d'appeler le *multimedia*. Il convient de combler une lacune qui risque d'approfondir le fossé des pratiques culturelles entre le monde des chercheurs, des savants, et celui des consommateurs d'images et de sons.

Un autre sujet d'étude porte sur l'historique de l'hypertexte et sur les difficultés liées à la conception des liens et des noeuds qui induisent des parcours de lecture.

Les difficultés liées à la conception et à l'utilisation de l'hypertexte évoquées dans :

-*Les catalogues en ligne : enquête à la Médiathèque de la Cité des sciences et de l'industrie : panorama des recherches*, coordination Monsieur Mohamed Hassoun, Madame Danielle Roger, ENSSIB-CERCI, 1994),

-Monsieur Hervé Le Crosnier dans sa thèse: *Systèmes d'accès à des ressources documentaires vers des antéserveurs intelligents* (Aix-Marseille III 1990) et dans plusieurs articles du B.B.F. (tome 36, n°4, 1991 et tome 40, n°2, 1995), sont pris en compte. La rationalisation des liens d'hypertexte peut être étudiée à partir du logiciel GRIF de l'INRIA (cf. : 3.3. PLAO et étude historique ).

Les Postes de Lecture Assistée par Ordinateur se situent dans un contexte où un programme central unique appelé *serveur*, gère tous les périphériques d'entrée et de sortie, et permet la création et la manipulation de fenêtres sur l'écran, l'affichage de textes ou de graphiques. Il se distingue du livre électronique, texte numérisé, dont la lecture reste essentiellement linéaire, dans la mesure où il permet d'appeler simultanément d'autres informations stockées sur le serveur, de pratiquer l'interactivité c'est-à-dire la possibilité offerte au lecteur de faire ses choix de parcours localement et d'annoter ses lectures. La différence est notable avec l'interrogation de bases de données, qui consiste à enregistrer une requête de départ, demande qui est alors traitée globalement par l'ordinateur. Pour le PLAO c'est en fonction de la lecture qui est en train de se faire, et donc en suivant le cheminement de la réflexion de l'utilisateur, que les informations sont appelées, d'où la notion fondamentale de lecture non-linéaire, insensible à la nature de l'information (texte, image, son) et à sa dispersion : l'information n'est plus nécessairement stockée de manière linéaire, le principe fondateur est qu'une information peut être conservée, sous forme d'enregistrements séparés, dans l'espace de mémorisation et lue suivant des parcours multiples. La lecture non linéaire parcourt les blocs du document initial dans un ordre indépendant de celui de leur enregistrement. Les blocs d'une base de données peuvent être lus suivant l'ordre de leur numéro d'enregistrement ou suivant des parcours de lecture différents utilisant des tables d'index et un langage de requêtes structuré de type SQL, les textes numérisés peuvent être utilisés sous forme d'hypertexte, à condition qu'un système de description du document (DOD), de balisage (tags) de la structure interne (auteur, chapitre, paragraphe, index de mots significatifs) soit mis en oeuvre. Les langages de balisage les plus courants sont : SGML utilisé par les éditeurs et imprimeurs, et HTML sous-ensemble de SGML défini par les chercheurs sur le réseau Internet (CERN de Genève à l'origine, et adresses *http*) et devenu une norme de fait dans le monde .

Le texte électronique autorise la communication des textes à distance, qui annule la distinction, jusque là irrémédiable, entre le lieu du texte et celui du lecteur. Délocalisé, accessible simultanément pour plusieurs utilisateurs, le texte en sa représentation électronique peut atteindre n'importe quel lecteur doté du matériel nécessaire pour le recevoir. A supposer que tous les textes existants, manuscrits ou imprimés soient numérisés, c'est l'universelle disponibilité du patrimoine écrit qui devient possible. 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 du codex mais directement le concept d'une immense bibliothèque dans laquelle chaque bloc d'information serait accessible sous forme d'hypertextes permettant de consulter les livres numérisés.

Cette conception du document électronique qui émerge du principe de la digitalisation du livre pose deux types de problèmes au bibliothécaire :

- l'organisation du mode d'utilisation des livres numérisés au travers des Postes de Lecture Assistée par Ordinateur, mais aussi des réseaux ( Internet, Renater),
- il s'agit d'une nouvelle conception de l'organisation de l'information, qui permet à la fois de repenser la bibliothèque mais aussi qui doit hériter du savoir-faire et des réflexions multiséculaires des bibliothécaires, pour ne pas entraîner le lecteur vers une perte de sens par surabondance et désorganisation de l'information.

Les Postes de Lecture Assistée par Ordinateur constituent un outil d'observation privilégié pour l'étude de la mise au point de nouveaux objets techniques, textes numérisés, normes de balisage, hypertexte, annotations dynamiques, architecture client-serveur et réseau, documents multi-médias, qui s'inscrivent dans le champ des sciences de l'information.

Comment la recherche dans le domaine de la bibliométrie, de l'infométrie, de la scientométrie se conjugera-t-elle avec la pratique des Postes de Lecture Assistée par Ordinateur ? La constitution de vastes corpus numérisés permettra, sans doute, de représenter le contenu de l'information par des cartes thématiques (soulignant la topologie des relations entre des disciplines ou des thèmes de recherche), visant à utiliser les traces et indices produits par des utilisateurs divers, au cours de leurs pratiques, de leurs interactions, et ce afin de faire apparaître de nouveaux objets d'analyse. Ces nouveaux outils s'appuieront probablement sur les nouvelles mémoires électroniques, sur les capacités offertes par le vaste processus de numérisation du signe et la plasticité qui en découle.

Les Postes de Lecture Assistée par Ordinateur manifestation des besoins de la recherche et des méthodes de la lecture savante, ne contiennent pas tous les savoirs, mais divers moyens d'accéder aux savoirs. Cette offre peut être étudiée pour être adaptée aux publics et aux types de bibliothèques.

## 1. Introduction

*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.*

*Toute cette conception de la lecture savante a été, on le sait, élaborée par Jean Gattegno, Alain Giffard, Bernard Stiegler, Jacques Virbel (voir bibliographie) et de nombreux participants associés en tant que "grands lecteurs". La gageure n'était ici pas seulement ni essentiellement technique. Elle supposait d'abord d'avoir une claire vision des possibilités et des limites attendues de l'assistance informatique. Le poste de lecture assistée par ordinateur, par l'acuité de la méthode de conception à partir des usages, a montré ce qu'on pouvait attendre de l'informatique de façon prémonitoire, puisque, quatre ou cinq ans après, ces fonctions tendent à se retrouver de façon banalisée sur le marché des logiciels. (Yannick Maignien *La Bibliothèque virtuelle ou de l'Ars memoria à Xanadu*, B.B.F.t.40 n°2, 1995).*

L'informatique à la Bibliothèque nationale de France commencée en 1970 avec la diffusion de la bibliographie nationale sur bande magnétique est arrivée à maturité en 1987, lors du démarrage de la base BN-OPALE, c'est-à-dire du catalogue informatisé des imprimés. Le dépôt légal, les acquisitions françaises et étrangères, le catalogue, l'inscription des lecteurs, la recherche bibliographique, la communication des ouvrages, le récolement, sont gérés par des systèmes aujourd'hui en exploitation. La Bibliothèque nationale de France prépare également le proche avenir : elle étend sa gamme de CD-ROM bibliographiques, mène un ambitieux programme de numérisation, ouvre aux bibliothèques françaises et étrangères ses bases BN-OPALE et BN-OPALINE via le réseau RENATER / INTERNET. Si la numérisation des fonds (largement engagée au sein de la Bibliothèque Nationale de France, par le programme des 100 000 volumes et des 300 000 images numérisées), et la conception de Poste de

Lecture Assistée par Ordinateur vont de pair, l'informatisation du catalogue, au sens classique du terme (gestion de bases de données portant sur les attributs externes des documents : auteur, titre, édition, description physique), n'utilise que très partiellement les possibilités de l'informatique. L'hypertexte, le balisage des documents, les index électroniques, la création de supermémoires électroniques (IBM développe un CD multicouche à laser bleu, avec 6,5 gigaoctets de mémoire et l'équipe du professeur Misawa de l'Université de Tokushima au Japon espère mettre au point rapidement des CD à quartz contenant des téraoctets de mémoire sur 12 cm de diamètre), permettent de créer des outils de lecture tout à fait nouveaux, qui peuvent être reliés à des réseaux.

Avec la numérisation des textes et des images, la Bibliothèque Nationale de France anticipe l'évolution de la mémoire électronique. La numérisation permet de diminuer la pression sur les collections et à ce titre intéresse les bibliothèques : la communication d'un document numérisé évite d'avoir à prêter l'original ; le texte numérisé est communicable simultanément à plusieurs personnes et peut même, sous certaines conditions, être accessible à distance.

Pour le lecteur, la numérisation facilite et élargit l'accès aux collections. Elle lui apporte de nouvelles possibilités de lecture savante assistée par ordinateur. Il peut se constituer sur son ordinateur un corpus de textes numérisés. Il peut, pour éclairer le texte faire appel à des documents s'y rapportant (dictionnaires, éditions critiques, biographies articles ...). Dans ce cadre les Postes de Lecture Assistée par Ordinateur (PLAO), éléments indispensables des services d'accès aux collections, permettront la consultation des fonds numérisés, mais également l'approche de nouvelles pratiques intellectuelles de la lecture et de l'écriture (annotation dynamique et constitution d'un corpus personnel déterminé librement par le chercheur grâce à de larges facilités d'accès). Ils permettront de travailler les textes directement sur écran, d'effectuer des recherches dans les bases et banques de données distantes sans se déplacer ce qui représente un gain de temps considérable, de télécharger (protocoles FTP *file transfer protocol* ) sur son micro-ordinateur tout ou partie d'un texte numérisé et accéder aux collections de diverses bibliothèques. Autant de perspectives riches de

promesses qui doivent toutefois se développer dans le respect des droits des auteurs et des éditeurs. Elles s'inscrivent dans l'évolution de la culture humaine et suscitent des résistances parce qu'elles introduisent un objet technique dans la culture savante : le philosophe Gilbert Simondon a remarqué que dans notre civilisation un hiatus se manifeste entre les attitudes suscitées en l'homme par l'objet technique et la vraie nature de ces objets ; de ce rapport inadéquat résulte chez l'acheteur, chez le constructeur, chez l'opérateur, un ensemble de valorisations et de dévaluations mythologiques ; pour remplacer ce rapport inadéquat par une véritable relation, il faut opérer une prise de conscience du mode d'existence des objets techniques.

En prolongement de cette pensée Monsieur Bernard Stiegler professeur de philosophie à l'Université de Compiègne a souligné la place de mémoire conservée dans les objets, dans le processus d'évolution de l'homme. Les avancées technologiques et informatiques, l'hypertexte notamment rappellent que les grands concepts informatiques sont nés dans les bibliothèques ; il revient à la société industrielle d'avoir rendu possible la théorisation de l'économie de l'information. Aujourd'hui, c'est la question de l'accès au savoir et de sa démocratisation qui est essentielle. (Bernard Stiegler, intervention au Congrès de l'ABF, Chambéry, 1993).

L'utilisation des nouvelles technologies et plus particulièrement des techniques de numérisation, s'était imposée aux rédacteurs des *Propositions pour une grande bibliothèque* Messieurs Patrice Cahart et Michel Melot qui avaient indiqué la place que ces pratiques "d'un type entièrement nouveau" devaient occuper au sein du nouvel établissement, et particulièrement la mise à la disposition des chercheurs des "postes de lecture assistée par ordinateur" (PLAO), permettant la consultation de textes numérisés. Dans son numéro spécial consacré à la mémoire, la revue *La Recherche* n° 267 juillet-août 1994 volume 25, dans un article intitulé "La chaîne du savoir" remarque que : "le catalogue national informatisé lancé parallèlement à la construction de la nouvelle Bibliothèque Nationale de France, et qui comprendra plus de dix millions de notices bibliographiques est l'un des plus impressionnants projets de recensement et de documentation d'un patrimoine. Dans le même temps, le programme de numérisation, sous la direction scientifique de Monsieur Y. Maignien,

met sous forme informatique un premier volet de cent mille ouvrages (trois cent mille assez rapidement). L'objectif est de favoriser la diffusion de textes difficiles d'accès dans les bibliothèques courantes, et de proposer un large échantillon représentatif des savoirs contemporains. Ce difficile projet encyclopédique brosse une vision moderne de la mémoire collective. Partant des questions "comment se souvenir? De quoi doit-on se souvenir? Que faut-il oublier ? Comment éviter le piège de l'exhaustivité ? qui surgissent dans toute réflexion sur la mise en place de nouvelles technologies destinées à conserver les traces de savoirs et de savoir-faire", les auteurs examinent les choix de la Bibliothèque Nationale de France, notamment la constitution d'une équipe d'une dizaine de personnes, assistées par un groupe d'une centaine de scientifiques de toutes disciplines, chargées de dresser une sorte de bibliothèque idéale. Cette dernière comprendra dix mille ouvrages de référence (encyclopédies, dictionnaires, etc.) numérisés en mode texte, sur lesquels pourront être appliqués les outils modernes de la bibliométrie (statistiques sur les formes lexicales) et quatre-vingt-dix mille textes en mode image ; avec ces derniers sera donc aussi conservée une mémoire visuelle des choix graphiques et typographiques, qui font partie intégrante de l'histoire du livre. En pariant en outre sur des techniques modernes de diffusion de cette bibliothèque idéale, c'est un véritable effort de prospective qui a été fait sur ce que pourront être, demain, les moyens de conservation, de gestion et de diffusion de la mémoire collective" (Philippe Blasco, Philippe Pataud-Celerier). L'impact de l'omniprésence des réseaux sur les activités de la communauté universitaire via Internet, conduit à la notion de collègues invisibles virtuels. Comme le souligne Monsieur Michel Melot Président du Conseil Supérieur des bibliothèques dans son rapport pour l'année 1994, disponible en texte intégral sur Internet (<http://www.grenet.fr/redoc/csb>), dans son chapitre sur la documentation électronique faisant le point sur les progrès techniques (codage électronique des documents et norme SGML, numérisation, compression de données, protocole Internet) et leurs conséquences dans le domaine de l'édition et des bibliothèques, pour préparer le nouveau rôle éditorial de ces dernières en partenariat avec les éditeurs, le contrôle de la circulation électronique des documents, l'aménagement des

modes d'application du droit d'auteur, le numéro d'identification des documents électroniques, il est indispensable de développer une véritable "veille juridique et technique". (Rapport du président pour l'année 1994 / Conseil supérieur des bibliothèques, Paris, Association du Conseil Supérieur des bibliothèques, 1995).

La Bibliothèque Nationale de France, a reçu dans ses missions la mise en oeuvre des technologies contemporaines de l'informatique et des télécommunications au service de la lecture, de la conservation du patrimoine, de l'assistance au travail intellectuel, de conservation et d'utilisation de la mémoire. Elle comportera un catalogue entièrement informatisé, ainsi que de vastes fonds numérisés : son, images fixes et animées, textes ; ce sera une bibliothèque consultable à distance. Lors des journées d'études consacrées à "l'arrivée d'Internet dans les bibliothèques et centres de documentation : conséquences et pratiques" par l'Association des conservateurs de bibliothèques le 6 avril 1995, Monsieur André Zysberg directeur du Département informatique et nouvelles technologies à la Bibliothèque nationale de France a précisé l'état du projet. On le sait la BNF met au point un nouveau système d'information qui sera "le seul avec celui de la Bibliothèque nationale du Canada" à être basé sur un système client/serveur", cent mille documents numérisés et trois cent mille images fixes seront accessibles. Le catalogue unique multimédia proposera huit millions de notices avec un temps maximum de deux secondes. L'accès à distance (aujourd'hui proposé gratuitement à une centaine de bibliothèques et donnant accès à Opale et Opaline) sera étendu en 1997, offrant des possibilités élargies, obéissant à la norme Z 39-50, permettant d'aller chercher des références dans le monde entier, de les stocker sur un antéserveur et de les restituer sur l'écran de la bibliothèque qui interroge.

L'assistance à la lecture, à la recherche, à l'accès à l'information et aux progrès dans la pratique du travail intellectuel est au coeur des préoccupations des pouvoirs publics. La numérisation de grands corpus de textes, l'apparition de nouvelles industries de la mémoire, leur mise en réseau au niveau mondial entraînent une nouvelle conception de l'encyclopédisme, une nécessaire réflexion sur la manière dont les mémoires de masse doivent être indexées, les flux d'informations analysés, pour

rendre l'accès du public aisé et fructueux. Marshall McLuhan invitait à "surfer" sur la vague des nouveaux médias. Mais la planche de surf n'est pas une interface commode pour la navigation en haute mer : la masse grandissante des informations à traiter, l'importance acquise par la recherche et le travail intellectuel dans les sociétés postindustrielles imposent une politique mettant au service de tous les moyens d'information et de culture modernes. La Bibliothèque Nationale de France, héritière des traditions de la Bibliothèque Nationale, s'est vue attribuer des missions nouvelles liées à la numérisation des fonds anciens (politique de préservation et de communication des collections), et à la réflexion sur la place des nouvelles technologies: pratique de l'hypertexte et accès au réseau mondial (Internet), au moyen des Postes de Lecture Assistée par ordinateur.

Le travail que nous nous proposons d'effectuer s'inscrit dans le cadre d'une recherche concernant les Postes de Lecture Assistée par Ordinateur de la Bibliothèque Nationale de France. Les stations de lecture sont un élément d'un programme comprenant d'importants fonds de documents numérisés, reliés à un serveur fournissant des informations bibliographiques ou textuelles. Le projet est basé sur l'analyse de l'émergence des besoins des lecteurs spécialisés et l'intégration des fonctions diverses d'accès aux textes numérisés , d'annotation, d'analyse, de gestion et d'administration de bases de données individuelles. Comme le rappelait Madame Marcelle Baudiquez, directrice du Développement scientifique et des réseaux à la Bibliothèque Nationale de France lors d'un colloque de l'ABF (24 février 1995), il s'agit de mettre en place : " une collection de recherche , de sauvegarde et de conservation, une large diffusion simultanée complémentaire pour la recherche et le libre accès, une amélioration des flux de communication". L'outil informatique d'accès permettra une protection de l'intégralité de l'oeuvre, un travail personnel du lecteur sur les textes - par l'usage de postes de lecture assistée par ordinateur (PLAO) -, le stockage et l'impression des résultats de travail, la coopération et la communication à distance par une mise en réseau de la communication virtuelle avec les établissements partenaires et les ressources du système documentaire français. Les travaux sur la définition de ces stations de lecture concernent les chercheurs en informatique, linguistique

quantitative, information scientifique, ergonomie, les bibliothécaires, les spécialistes des conditions de formation du texte littéraire, les critiques, les auteurs et les praticiens des sciences humaines aussi bien que les fournisseurs de matériel et les lecteurs. Il faut également noter que le projet du Vice-président des États-Unis Albert Gore, (High-Performance Computing Act) propose la création d'une bibliothèque digitale, basée sur les collections de la Bibliothèque du Congrès et accessible par le réseau National Research and Education Network, et que dans le même temps des projets de bibliothèques de textes digitalisés voient le jour dans différents États : Projet Gutenberg à l'Université Benedicte en Illinois (qui propose les oeuvres de Lewis Carroll, ainsi que *The Book of Mormon*, *The Song of Hiawata*, *The Federalist Papers*, *Moby Dick*) Brown University, Princeton (National Center for Machine-Readable Texts in the Humanities), Oxford Text Archive, Georgetown Center for Text and Technology, projet de la British Library et que tout acheteur d'une machine NExT reçoit l'intégralité des oeuvres de William Shakespeare numérisées, alors que l'intégralité de la littérature grecque ancienne est disponible sur CD-ROM, et que l'offre privée s'étend (cf. l'article paru dans la revue *La Recherche* numéro 273 février 1995 volume 26 *Littérature et poésie sur livre électronique* consacré aux éditions IliAS qui publient sous forme d'*hyperlivres* les classiques de la littérature et de la poésie françaises). Ce mouvement mondial de numérisation du patrimoine écrit (on peut également citer l'entreprise de numérisation des Archives générales des Indes, à Séville, soit 45 millions de documents concernant l'Amérique, le programme de l'Unesco "Mémoire du monde", ouvert en juin 1992 avec le traitement des manuscrits de la Grande mosquée de Sanaa, au Yémen), conduit naturellement à étudier la mise en place de stations de travail adaptées à l'écriture et à la lecture, à la recherche et à l'annotation dynamique qui ne peuvent être disjointes, ainsi qu'au contrôle de l'origine des sources de la qualité des éditions numérisées, du respect du droit d'auteur.

On a étudié comment l'innovation s'est transformée en marchandise au début du siècle et comment aujourd'hui, la concentration industrielle a modifié les modalités de production et de diffusion de la technologie dans les industries de la culture. (cf : Jean-Michel Salaün *A qui appartient la télévision ?* Éditions Aubier, 1989).

Aujourd'hui l'industrie de la mémoire et des réseaux constitue un élément dynamique de l'économie mondiale pour des firmes comme Microsoft de Bill Gates, IBM, Apple, Intel, Motorola, ATT. Face à cet investissement de la mémoire artificielle collective, les pouvoirs publics doivent maintenir la prééminence du texte écrit, des droits du créateur et de l'auteur, ainsi que les possibilités d'accès pour tous à la culture. C'est pourquoi la notion de Poste de Lecture assistée par Ordinateur est importante : il s'agit d'apporter une aide active au travail intellectuel vivant, à la recherche et à l'étude, indépendamment des flux économiques et des effets de mode.

On peut appliquer à la lecture assistée par ordinateur la règle qu'Émile Durkheim édictait pour l'observation de tous les faits sociaux :

“Quand on entreprend d'expliquer un phénomène social, il faut rechercher séparément la cause efficiente qui le produit et la fonction qu'il remplit. “ (Règles de la méthode sociologique).

Il ne manque pas d'acteurs intéressés par cette demande nouvelle. Les industries culturelles, presse, édition peuvent développer des services concurrents dans le domaine de la lecture de textes numérisés et de l'accès par CD-ROM ou réseaux. Les bibliothèques se distinguent par l'importance de leurs fonds numérisés, et la mise au point d'outils de lecture spécialisés étudiés pour le public des chercheurs.

Après une présentation des Postes de Lecture assistée par Ordinateur, nous étudierons leur place dans la recherche culturelle de notre temps en relation avec les pratiques intellectuelles, et le monde des médias, en tentant de dégager une particularité forte, correspondant aux stratégies de l'offre propre aux bibliothèques. Celles-ci peuvent insérer la notion de Poste de Lecture Assistée par Ordinateur dans une analyse des relations entre l'organisation et son milieu.

La numérisation d'un corpus d'ouvrage en texte intégral réintroduit les notions de collection et de patrimoine dans les systèmes électroniques. Alors que l'Information Scientifique et Technique avait privilégié l'interrogation ponctuelle de bases de données distantes, pour obtenir des bibliographies ou des informations, l'idée de lecture assistée par ordinateur redonne vigueur à "l'activité de base arrière" définie dans le *Marketing des bibliothèques* (Jean-Michel Salaün, éditions de Cercle de la

Librairie, 1992 ) et liée à l'idée de patrimoine : il faut choisir les ouvrages à numériser, étudier les suggestions et les remarques des utilisateurs, collaborer avec eux dans la mise à jour des outils liés au PLAO (logiciel de création de base de données personnelles, scanner etc...), mener une étude critique des éditions informatisées (origine qualité), et faire respecter les droits d'auteur.

Il ne s'agit plus de créer des répertoires inspirés de la technique du bulletin signalétique, mais d'une collection d'oeuvres constituant un fonds, dont la nature est déterminée par des commissions de spécialistes. Il faut donc sélectionner ces fonds et examiner les outils informatiques utiles pour leur utilisation par les lecteurs. Les nouvelles formes de lecture générées par ces dispositifs sont un sujet d'étude qui peut se révéler riche d'enseignements.

### **1.1. Le programme de numérisation**

Créée par décret du 3 janvier 1994, la Bibliothèque nationale de France résulte de la fusion de la Bibliothèque nationale située à Paris rue de Richelieu et de l'Établissement public chargé de la construction de la Bibliothèque de France sur le site de Tolbiac. Cette construction a démarré à la fin de l'année 1990 et s'est achevée en 1995. Le bâtiment sera ouvert au public fin 1996. Cent mille textes et trois cents mille images seront consultables par Poste de Lecture Assistée par Ordinateur dès la deuxième moitié de 1997. A cette collection s'ajoutent les éditions numériques soumises au Dépôt légal, les produits multimédias et les documents informatiques.

La B.N.F. mettra en oeuvre les technologies contemporaines de l'information et des télécommunications. Elle comportera outre le catalogue entièrement informatisé déjà évoqué, de vastes fonds numérisés consultables à distance, et des postes de travail d'où l'on pourra interroger d'autres bibliothèques et serveurs, ainsi que ses fonds locaux .

Les textes numérisés accessibles proviendront de différentes sources. La Bibliothèque Nationale de France envisage la numérisation d'environ 350 000 ouvrages, en mode bit-map ou alpha-numérique ; mais l'utilisateur pourra par ailleurs

accéder à d'autres fonds, soit sur place, par exemple des éditions réalisées sur disque compact (CD-ROM) par différents éditeurs, soit via des réseaux, auprès des serveurs accessibles dans le monde. L'accès aux collections d'images numérisées se fera via le système informatique audiovisuel et sur des postes de travail spécifiques. Cependant la consultation de ces images numérisées sera également possible sur les postes de travail destinés à la collection de textes numérisés avec une qualité d'image satisfaisante puisque les originaux auront été numérisés avec une haute définition.

Il est essentiel que des accords avec les éditeurs pour les droits liés à la consultation et à la reproduction des documents numérisés soient conclus, tout en n'obligeant pas l'établissement à pratiquer des prix prohibitifs pour les utilisateurs. La numérisation est appliquée largement aux collections patrimoniales libres de droits, mais pour une raison de cohérence documentaire, les documents datés d'avant la fin du XIX<sup>ème</sup> siècle ne représentent que 60 % des 100 000 documents actuellement numérisés. Les 40 000 ouvrages restants, qui ne sont pas libres de droits, font l'objet de négociations avec les éditeurs et les auteurs.

### **1.2. Conséquences des réflexions sur le choix des textes à numériser et sur les fonctions des Postes de Lecture :**

Les différents usages possibles d'un texte numérisé à l'intérieur d'une bibliothèque entraînent différentes logiques de sélection des documents. Elles ne sont pas contradictoires, elles doivent être intégrées de manière à définir un programme d'acquisition.

Dans une logique de préservation ou de transmission à distance, c'est le critère de rareté (dans les bibliothèques) qui prévaut, mais, pour décider de numériser, il faut croiser ce critère avec celui de la demande potentielle. Pour la gestion de la communication ce sont les documents les plus demandés qu'il faut privilégier, en tenant compte toutefois de la nature du document et de sa pertinence pour le travail informatique.

Ces analyses ont conduit à une première approximation du fonds électronique et de sa place dans la Bibliothèque Nationale de France.

Que ce soit pour faciliter la transmission à distance ou pour mettre à la disposition du public des Postes de Lecture Assistée par Ordinateur, le fonds électronique doit répondre à des critères de cohérence intellectuelle permettant d'offrir la plus grande visibilité. En ce sens la réponse de la demande de numérisation ponctuelle d'un ouvrage de la part d'un usager devra être appréciée en fonction de la pertinence d'intégrer ou non l'ouvrage dans l'ensemble de la collection. Toutefois la mise à disposition de scanners dans le cadre du PLAO permet à l'utilisateur de numériser des textes qu'il garde sur son propre support de mémoire, détachés de la base commune.

Les domaines ayant fait l'objet d'une exploration approfondie sont les suivants : littérature et linguistique française, science du langage, philosophie, histoire des sciences et sciences exactes, histoire du Moyen Âge et histoire contemporaine, Antiquité (histoire, littérature, philosophie). Au coeur de la collection on trouvera, pour toutes les disciplines, les "classiques", "les textes de référence" ou les textes fondateurs de la discipline. Pour répondre à une demande unanime des experts consultés, un fonds important de revues sera intégré à la collection. Celles-ci constituent en effet des outils de référence pour toutes les disciplines, et parfois l'essentiel des documents souhaités pour certains champs (ethnologie, linguistique). De plus, certaines collections de revues peuvent être d'un accès très difficile (revue de littérature et d'art du début du XX ème siècle par exemple). 80 % des textes numérisés ont été sélectionnés ; 35 % d'entre eux sont déjà numérisés et stockés. 40 000 images fixes environ sont actuellement numérisées à partir de collections extérieures à l'établissement et constituent des lots thématiques.

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 ;

- 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 : télédéchargement 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) ;

-entre utilisateurs et administration de la bibliothèque ;

-entre la bibliothèque et les autres bibliothèques et centres serveurs.

### **1.3. Les fonctions centrales : la lecture 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 assistée 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. Celle-ci est en effet une lecture de fréquentation et d'appropriation qui dans son principe, et malgré quelques analogies de surface, ne revient pas à la consultation classique 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

structures 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,

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.)

#### **1.4. 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 ( Nielsen, Jakob *Online Documentation and Reader Annotation. International Conference on Work with Display Units*, Stockholm, 12-15 May 1986 ; Virbel Jacques et al ... *Outils personnalisés de modélisation de l'utilisateur dans la gestion électronique de documents. Note, Institut de Recherche en Informatique de Toulouse*, 1991) . 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 générés 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. Par exemple, pour en rester aux plus classiques :

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

- modifier le texte (corrections, suppressions, insertions, interversions, 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;

  - expliciter 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, etc.

Quand 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 parcouru 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.).

Ces interventions ont été modélisées dans une maquette informatique, LECAO, réalisée à l'Université de Compiègne. Ce logiciel (Lecture et Écriture Critiques Assistées par Ordinateur) est une maquette simulant une interface de lecture assistée par ordinateur. Elle a pu être réalisée avec l'aide du Ministère de la Recherche, département Sciences de l'Homme et de la Société, et en collaboration avec la Bibliothèque Nationale de France. Ce programme se poursuit sous un nouveau nom OPEN (Outil Personnalizable d'Édition Numérique), dans le cadre du programme PARINFO. LECAO 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 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.

#### **1.5. Pratique et définition de la lecture assistée par ordinateur :**

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 destinés à la publication assistée par ordinateur, écrans couleurs à cristaux liquides des ordinateurs portables). 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, un outil ergonomique pour la navigation textuelle et l'annotation).

Ces 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. 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. La numérisation permet de faire se rejoindre la plupart des outils de création et d'aide au travail intellectuel.

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.

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 coeur 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.

La réflexion sur les propriétés souhaitables des stations de lecture à la Bibliothèque Nationale de France a débuté dès 1990 sous la direction de Bernard Stiegler (on peut également citer Jacques Virbel chercheur à l'Université Paul

Sabatier de Toulouse, et l'helléniste Christian Jacob chercheur au CNRS). 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* : 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 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).

On peut tenter de rapprocher la lecture/écriture de recherche, de l'idée d'intentionnalité de Husserl : la conscience est par nature conscience de quelque chose, elle est intentionnelle. L'intentionnalité n'est pas une propriété surajoutée, mais la détermination intrinsèque et essentielle de la conscience. La conscience est dirigée sur quelque chose, sur un objet. La conscience vise un objet réel extérieur à elle, qui n'est pas une image mentale. Husserl introduit la notion de noème qui est associé à l'acte de la conscience de viser un objet réel. C'est la pensée de l'objet en tant que pensé dans l'acte de viser l'objet réel. Le noème est une abstraction opérée à partir des actes : des actes visant le même objet sous les mêmes déterminations, avec les mêmes croyances possèdent le même noème ; mais les actes s'appuient sur des instances temporelles différentes du noème. La mise en oeuvre du noème par la pensée est une noèse, une action de la pensée. L'intentionnalité de la conscience

repose sur le noème de l'acte, et non sur l'objet réel. A cette conception de Husserl on ajoutera les idées de Gilbert Simondon concernant l'individuation psychique et collective : la textualité d'un texte est essentiellement la genèse d'un sens tel qu'il n'est pas une signification (il est instable, incertain, processuel, dynamique). C'est une telle genèse que Simondon nomme un "processus d'individuation psychique et collective". Qu'elle doive être appréhendée comme relation transductive d'incertitude signifie que le sens d'un texte n'est pas derrière lui (il n'y a pas de vouloir dire du texte), mais entre le texte et son lecteur. Le lecteur se lit (lit son propre texte, le tissu textuel de sa propre mémoire) à travers ce qu'il lit, tel que cela a été mis en forme matériellement et peut par là même faire acte de création comme le donne à penser Marcel Proust :

*chaque lecteur est, quand il lit, le propre lecteur de soi-même. L'ouvrage de l'écrivain n'est qu'une espèce d'instrument optique qu'il offre au lecteur afin de lui permettre de discerner ce que dans ce livre, il n'eût peut-être pas vu en soi-même. La reconnaissance en soi-même, par le lecteur, de ce que dit le livre, est la preuve de la vérité de celui-ci et vice-versa, au moins dans une certaine mesure, la différence entre les deux textes pouvant être souvent imputée non à l'auteur mais au lecteur. (Marcel Proust le Temps retrouvé).*

Husserl analyse l'écriture comme condition de possibilité d'une communication omnitemporelle sans laquelle aucune réitération de l'idéalité géométrique ne serait possible. L'écriture linéaire est ce qui, pour la première fois (première fois du "savoir impersonnel", du livre ouvert par Hérodote, de la géométrie), donne un accès littéral au passage de la parole (à son présent passant) comme à son passé (à son présent comme passé)- accès à la lettre, exact, et dans cette mesure, condition même de l'idéalité visée par Husserl, et en premier lieu, de l'idéalité géométrique : c'est l'intelligibilité "pour tout le monde, indéfiniment perdurable" (J. Derrida, *Introduction à l'origine de la géométrie*, PUF 1962), qui est la condition de la réactivabilité du sens, elle même condition de l'idéalité. Pour le philosophe Bernard Stiegler, il s'agit d'une intelligibilité qui appelle une exactitude de l'enregistrement de la signification. (Bernard Stiegler étudie cette question en philosophe à partir de Husserl, Derrida, Leroi-Gourhan, Gilbert Simondon, mais aussi en praticien de

l'ordinateur, à la façon dont les Nouvelles Technologies de l'Information et de la communication modifient les catégories et les usages de la mémoire et du temps, depuis la barrette d'ivoire de l'alphabet cunéiforme d'Ugarit jusqu'aux microprocesseurs. Enseignant à l'Université de technologie de Compiègne, il a été commissaire à l'exposition "Mémoires du futur" au centre Georges-Pompidou en 1987, et il collabore à l'équipement des postes de lecture assistée par ordinateur (PLAO) de la future Bibliothèque Nationale de France ). Ce n'est pas n'importe quelle écriture qui rend possible la *communautisation* des énoncés idéaux de la science : c'est celle qui pose exactement , ortho-graphiquement, permettant d'*examiner à loisir*, selon les mots de Leibnitz, ce qui s'est *pensé* comme étant ce qui *s'est passé*, auto-examen de la pensée qui devient ainsi à elle-même son propre passé totalement accessible - ce qui ne veut pas dire transparent : penser littéralement, c'est accéder au jeu différentiel de la scripturalité originaire de la langue du fait même de pouvoir toujours y ré-accesser indistinctement, exactement, ortho-graphiquement.

La critique de l'analyse husserlienne (Derrida, Stiegler) a mis en lumière, la limitation du processus de télé-communication inhérent à toute géométrie reliée au seul domaine de l'intersubjectivité des géomètres. Pour Husserl, l'écriture est d'emblée nécessaire pour autoriser une géométrie, et cependant cette nécessité ne vient qu'après coup, comme étant celle d'une archivation pour les successeurs de l'inventeur. Or le loisir de l'examen est l'horizon de l'invention elle-même, son champ spatial aussi bien que temporel , ce qui se dispose sur les supports écrits, s'y mettant matériellement en forme, dans le laps de temps qu'ouvrent les boucles récursives de la rature et du repentir (avatars, pour Husserl, de la "réactivation"). L'écriture n'est donc pas seulement la répétition de l'invention ; c'est bien plutôt l'invention qui est la répétition. S'il est vrai que "la fonction de la sédimentation traditionnelle dans le monde communautaire de la culture sera de dépasser [par l'écriture] la finitude rétentionnelle " de la conscience individuelle (Jacques Derrida), le dépassement de la finitude rétentionnelle "dans le monde communautaire" (comme inscription ou enregistrement) est nécessaire dès le moment de l'invention. Mieux : il est ce moment par excellence.

Il faut évaluer la portée des analyses des instruments bibliographiques puis informatiques de navigation dans le savoir, proposées par Leroi-Gourhan, depuis cette situation originaire de finitude rétionnelle, telle qu'elle affecte toute mémoire anticipant un avenir dans la possibilité de revenir sur son passé, et telle qu'elle détermine les époques de cette possibilité selon la matérialité des mises en formes que libère la mnémotechnique dominante. Et c'est aussi dans cette perspective qu'il nous faut concevoir l'avenir du texte numérisé, dans la mesure où ce qui était vrai de la bibliothèque collective, patrimoniale, l'est aussi aujourd'hui de la bibliothèque privée, et au-delà, des formes les plus intimes de bibliothèques (du corpus scanné par son lecteur à son bénéfice exclusif, jusqu'à l'agenda électronique connectable sur la station de travail, en passant par le power-book) : *le développement des instruments micro-informatiques de gestion de données personnelles nous fait entrer dans l'ère des instruments de navigation dans la mémoire individuelle devenant par là même, assistée des automatismes de l'ordinateur, "réactivable" intégralement et à tout instant. Le sens du mot "réactivation" change ici de sens. Mais dans quelle mesure ? Et en quoi change-t-il du même coup le sens du mot chez Husserl lui-même ? C'est ce que je tenterai d'examiner à présent, en me penchant sur un travail commun au lecteur et à l'écrivain : l'annotation, en ses multiples dimensions .*

*Approcher le travail de l'écrivain, c'est fréquenter aussi ses outils et lieux d'études : stylo, papier, carnet, mais également chemises, dossiers, rayonnages où ils s'accumulent avec les ouvrages marqués, cornés, annotés, tiroirs où sont rangés des fichiers, et tant d'autres formes d'aide-mémoire composant les supports des codes de "métascriptions" et d'un processus global d'archi-annotation, auxquels aujourd'hui s'ajoutent évidemment les disquettes, mais encore les photos, bandes magnétiques audio et video et "échantillonnages" numériques que les extensions multimédia des systèmes micro-informatiques (tel quick-time sur le système 7 du Macintosh) font à leur tour entrer dans le champ d'une annotation électronique généralisée au sein de dispositifs dits hypermedia.*

*Bien avant l'hypertexte et l'électronique, tout un espace rétionnel et virtuel, statique et cependant parfaitement opératoire, encadre physiquement l'écrivain*

*travaillant à sa table. De subtiles techniques d'annotation, corrélation, classement, organisent dès son origine l'écriture des oeuvres, qui peuvent être aujourd'hui transposées dans le domaine des supports numériques au plus grand bénéfice du lecteur-scripteur : elles s'en trouvent incommensurablement dynamisées.* (Bernard Stiegler *Machine à écrire et matière à penser*, 1994).

Le mot "réactivation" change certes ici de sens. Bernard Stiegler analyse ce changement en se penchant sur un travail commun au lecteur et à l'écrivain : l'annotation, en ses multiples dimensions. De subtiles techniques d'annotation, corrélation, classement, organisent dès son origine l'écriture des oeuvres, qui peuvent être aujourd'hui transposées dans le domaine des supports numériques. Le lecteur-auteur fait appel à des capacités d'abstraction : il fait appel à des capacités métalinguistiques qui lui permettent de traiter le langage comme un objet propre et indépendant, autorisant l'exercice d'une pensée libre du contexte concret. Le lecteur/scripteur poursuit l'enchaînement de sa pensée à travers la lecture. Sa conscience est entièrement dominée par l'intentionnalité au sens de Husserl, et il traverse dans la confrontation au texte une transduction : "l'individu assimile une genèse et l'exerce à son tour" (Simondon).

#### 1.6. Personnalisation du document

Les procédés de personnalisation du document, tout en étant proches de cette notion, ne s'identifient pas complètement à la pratique de l'annotation dynamique. Ces méthodes de travail sont liées aux concepts de gestion électronique de documents et aux systèmes de production et de manipulation de documents. Sept approches différentes ont été étudiées :

- les systèmes *Computer-Supported Cooperative Work* (Winograd), recourant à des théories élaborées du dialogue et de la conversation et à la caractéristique précises des liens unissant des pièces textuelles constituant un dialogue ont été utilisés dans des systèmes comme TEXNET (Trigg) et GIBIS (Conklin). L'intérêt de ces recherches est qu'elles mettent en lumière l'importance de la définition précise de liens entre les blocs de texte constituant un dialogue, telles que "est une réplique à",

“est une illustration de”, “est un exemple de”, “se substitue à”, etc. Des outils d’annotation ont également été développés (Internote (Catlin) and Quilt (Fish, Leland), permettant l’association de commentaires reliés avec les blocs d’hypertexte).

- les systèmes “*Computer Aided Writing*”, fondés sur des modèles d’activités de rédaction, tel le *Writer’s Assistant* (Sharp, 1989), ou sur la formalisation de l’argumentation (Smolensky, 1988), ou des structures textuelles.

- l’archivage individualisé : la critique des systèmes documentaires à vocation universelle conduit à envisager la mise au point d’outils personnalisés d’indexation, directement dépendants des préoccupations documentaires particulières d’un utilisateur donné.

- les aides on-line : c’est donc une perspective d’individualisation et de personnalisation des relations homme-système qui se dégage des recherches actuelles sur les aides on-line, qu’elles se rapportent à des bases de données, des systèmes d’exploitation ou des logiciels.

- les Hypertextes utilisent deux concepts qui sont directement liés aux Postes de Lecture Assistée par Ordinateur la définitions de systèmes en terme de liens entre les blocs et la sémantique du *browsing* (butinage). Les ressources documentaires peuvent être consultées comme une encyclopédie électronique avec un accès direct et la navigation entre des structures non-linéaires.

- le livre électronique est un type d’application proche des Postes de Lecture Assistée par Ordinateur : le livre ou le manuel est considéré comme une base de données spécifique et dispose de différentes procédures de navigation et d’appropriation par le lecteur / utilisateur. L’augmentation du nombre d’éditions sur CD-ROM montre que ce secteur d’activité va en croissant rapidement.

Les Postes de Lecture Assistée par Ordinateur comporteront un ensemble d’utilsitaires pouvant intervenir à des niveaux ou des phases très variables d’une recherche, comme le sont les analyseurs lexicographiques, syntaxiques, stylométriques. Les stations seront ouvertes pour l’incorporation de logiciels externes nécessaires à la pratique du chercheur.

### 1.7. La lecture / écriture, la différence phénoménologique et la transduction :

La conscience est par nature conscience de quelque chose : elle est intentionnelle. L'intentionnalité n'est pas une propriété surajoutée, mais la détermination intrinsèque et essentielle de la conscience. Pour Husserl l'objet n'est pas préexistant à ma perception. C'est plutôt ma conscience qui le constitue dans les vécus qu'elle contient. Un vécu est un acte de la conscience. Un vécu est intentionnel quand il porte sur quelque chose, quand par son entremise, la conscience accède à un objet dit intentionnel qui est le corrélat du vécu. Qu'il soit écrivain, professeur, étudiant, "intellectuel", un lecteur "professionnel" utilise diverses techniques de balisage et d'orientation que matérialisent les "habitudes graphiques" et/ou spatiales, qu'il en systématise ou non l'usage : signes d'annotations de significations diverses (croix et traits en marges, soulignements au tracé divers dans le corps même du texte imprimé, etc), techniques de résumé et de synthèse, fichiers, dossiers, etc. Cette lecture suit des structures cognitives que l'on peut rapprocher de la notion de noèmes : une abstraction opérée à partir des actes. Des actes visant le même objet sous les mêmes déterminations avec les mêmes croyances possèdent le même noème ; mais les actes s'appuient sur des instances temporelles différentes du noème. La mise en oeuvre du noème par la pensée pour atteindre l'objet est une noèse, une action de la pensée. L'intentionnalité de la conscience repose sur le noème de l'acte, et non sur l'objet réel. Par exemple, quand j'aperçois une maison, j'ai conscience d'un objet possédant trois dimensions, avec un devant, un arrière, des côtés...Pourtant à l'examen, je m'aperçois que je n'ai réellement perçu que les façades visibles de la maison. J'ai dû utiliser ma connaissance préalable de ce qu'est une maison pour associer à ma perception effective la prédiction qu'il doit y avoir des façades arrières et latérales. Le noème de la maison permet de guider la perception que j'en ai. Le cinéma a largement utilisé cette faculté pour créer des villes imaginaires au moyen de décors en trompe l'oeil (c'est le spectateur qui reconstitue l'incendie d'Atlanta à partir de la combustion de quelques décors), et L'Intelligence artificielle a créé des structures cognitives, les frames de Minsky (programmation orientée-objet). Le lecteur savant, interprète les

textes en fonction de structures cognitives et il annote et écrit pour matérialiser le résultat de sa recherche tout en parcourant les documents. Il peut également faire appel à un logiciel spécialisé pour obtenir des statistiques sur les occurrences du vocabulaire ou pour mettre en ordre des données chiffrées. Une telle pratique conviendrait mal à la lecture de distraction et on n'imagine pas un roman photo en mode hypertexte (à moins de concevoir tout un univers imaginaire de vies entrecroisées selon des méthodes combinatoires à la manière de Georges Pérec).

Le lecteur glose les textes, les indexe, les met en rapport par des systèmes de corrélation (les fichiers), y extrait des passages pour des citations, utilise des instruments de recherche (bibliographies, revues spécialisées, dictionnaires, encyclopédies).

Toutes ces techniques de lecture visent à créer des liens qualifiés entre des documents ou des passages dans des documents. Les techniques hypertextuelles actuellement développées par le génie logiciel permettent d'intégrer rigoureusement, en les automatisant, ces opérations. L'avantage en est que la mémoire de la machine n'oublie pas. L'informatique est la science des symboles ininterprétés, la physique des signes. Elle étudie des mécanismes effectifs ou algorithmes de manipulation des symboles. Tout algorithme est un codage de transformation agissant sur les codages de données pour donner le codage d'un résultat. La logique mathématique nous enseigne que la contrainte d'effectivité entraîne le caractère dénombrable, arithmétique de l'informatique : ce qui revient à conclure que les symboles ininterprétés ne sont que des nombres.

Mais ainsi que le fait remarquer Newell (*Physical symbol systems, Cognitive Science*, 4, 135-183, 1980), il est possible de considérer que les ordinateurs manipulent des symboles qui accessoirement peuvent être des nombres. A l'intérieur d'une architecture dite de Von Neumann, des éléments sont utilisés pour désigner les données du problème et la fonction pour le résoudre. Par conséquent, ces éléments apparaissent non comme des nombres, mais comme des symboles.

Ces changements de perspective entraîne une nouvelle façon de résoudre informatiquement un problème : en effet pour résoudre un problème, il ne s'agit plus

de le coder, mais de le représenter. L'informatique devient alors manipulation formelle de représentation symbolique. Cette manipulation peut se prêter aux formes du travail intellectuel les plus variées par la constitution d'images numérisées, de sons, d'éléments de calcul et de statistique, l'annotation, la constitution de bases de données, de thesaurus iconographiques. Lire est toujours écrire et inversement dans la mesure où le texte n'existe que dans sa manifestation : l'oeuvre d'art ne s'accomplit que dans la manifestation qu'elle reçoit, la lecture est un processus d'individuation du texte (B. Stiegler). D'où l'importance attribuée à l'idée de transduction.

La transduction n'est pas seulement démarche de l'esprit; elle est aussi intuition, puisqu'elle est ce par quoi une structure apparaît dans un domaine de problématique comme apportant la résolution des problèmes posés. Mais 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, non par apport de quelque forme étrangère. Elle n'est pas non plus comparable à l'induction, car l'induction conserve bien les caractères des termes de réalité compris dans le domaine étudié, tirant les structures de l'analyse de ces termes eux-mêmes, mais elle ne concerne que ce qu'il y a de positif, c'est-à-dire ce qu'il y a de commun à tous les termes, éliminant ce qu'il y a de singulier; la transduction est au contraire, une découverte de dimension dont le système fait communiquer celles de chacun des termes, et telle que la réalité complète de chacun des termes du domaine puisse venir s'ordonner sans perte, sans réduction, dans les structures nouvelles découvertes.

### 1.8. La question de l'hylémorphisme :

Parler de lecture assistée par ordinateur paraît à certains étrange ou même iconoclaste. Pourtant 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. C'est seulement après le pictogramme qu'apparaît le phonogramme : il s'agit d'un signe écrit qui du fait qu'il représente un mot dont la prononciation contient tel ou tel son, finit par se spécialiser comme écriture de ce son pour noter tout mot ou toute partie du mot qui le contient. 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 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 : tige de jonc mâchonnées par le scribe en pinceau, encre au noir de fumé, papyrus, écritoires, plumes, crayons, règles, bécicles, constituent l'*hylé*(matière), et la forme (*morphé* : figure disposition) mais aussi les voies par lesquelles on accède au texte. Les écrans à tubes cathodiques à affichages alphanumériques (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 mémoire son parcours de recherche, ont modifié les conditions d'accès au texte.

L'anglais Jean de Garlande, qui enseigna à Paris et à Toulouse, énumère dans son *dictionnaire*, rédigé vers 1220, les instruments nécessaires aux clercs : livres, pupitre, lampe à suif, veilleuse, lanterne, corne à encre, plume, plomb, règle, tablettes, stylet, cathédre, bureau, craie, rasoir, pierre ponce. Son contemporain Alexandre Neckham fournit une liste analogue dans son *De Nominibus utensilium*, et Didier de Birstoff qui acheva le 20 juillet 1473 la transcription de l'actuel manuscrit 50 de la bibliothèque municipale de Saint-Mihiel précise qu'il a accompli ce travail avec ses

bésicles sur le nez. Leroi-Gourhan a montré qu'avec la multiplication des imprimés, dont la fracture du bon à tirer est une conséquence, l'accumulation des textes, où se trame le savoir qui les tisse s'accroît au point qu'une assistance à l'orientation dans la géologie de la mémoire, par la création de catalogues, d'index et de fichiers, devient indispensable, tandis que le livre se dote d'appareils de navigation facilitant le déplacement non-linéaire du lecteur (pagination, table de matières, index, etc.) .

## 2. Objectif : l'assistance au travail intellectuel :

Cette assistance au travail du clerc, du savant, du lecteur se devait d'être renforcée, perpétuée dans le cadre des nouvelles technologies de l'information, de la lecture, de l'écriture.

L'ordinateur s'inscrit dans la continuité de l'histoire humaine : au sens premier *a computer* désignait en anglais une personne qui calculait des tables de logarithmes, balistiques, de navigation, de fonctions, etc. En 1823 Charles Babbage persuada le gouvernement britannique de financer le développement d'une "machine différentielle" qui automatiserait ces calculs. Quand les ordinateurs électroniques sont nés à la fin de la Seconde guerre mondiale, une de leur première destination fut de calculer des tables de trajectoires pour des tirs d'artillerie. Des premières intuitions de Babbage à une période récente (une trentaine d'années), il semblait évident que l'ordinateur était un calculateur, un "croqueur de nombres" (en américain : "number cruncher"). Le projet Whirlwind, au début des années cinquante mit en avant l'affichage des résultats - il s'agissait alors de la position des avions repérée par radar - sur un écran cathodique (cf : William, Michael R. A History of Computing Technology. Englewood Cliffs : Prentice-Hall, 1985), pourtant Vannevar Bush, qui avait travaillé sur un projet de machine différentielle inspirée de Babbage et qui avait eu l'intuition de l'hypertexte s'était orienté vers la technologie du micro-film,

plutôt que vers le computer pour expliquer son projet de système d'information personnel (Memex).

C'est le texte écrit, typographique, qui s'impose avec la numérisation intégrale et exacte des textes, l'annotation dynamique, la transmission télématique et la réception par ordinateur de corpus représentatifs de la culture savante. Dans ce monde de la représentation électronique des textes, deux contraintes, tenues jusqu'ici comme impératives peuvent être levées. La première contrainte est celle qui limitait étroitement les possibles interventions du lecteur dans le livre imprimé. Depuis le XVI<sup>ème</sup> siècle, c'est-à-dire depuis le temps où l'imprimeur a pris à sa charge les signes, les marques et les titres, titres de chapitres ou titres courants, qui, au temps des incunables, étaient ajoutés à la main sur la page imprimée par le correcteur ou le possesseur du livre, le lecteur ne peut insinuer son écriture que dans les espaces vierges du livre. Stendhal commençait à noter ses pensées "sur de petits morceaux de papier" ( cf. : *Journal*, 9 août 1803) ; puis il recopiait les notes dans son cahier de travail et les développaient. Progressivement, ce journal "littéraire" dans lequel il accumulait les pensées que lui inspiraient ses lectures fut remplacé par les marges des livres eux-mêmes, qu'il remplissait directement. Il prit ensuite l'habitude, pour retrouver le fil de ses annotations, de confectionner, sur la couverture ou la page de garde des livres, une sorte d'index personnel de l'ouvrage où il répertoriait les passages commentés, et qui lui servait de point de départ pour des relectures. Il renouait ainsi avec l'usage des annotations marginales des copistes du Moyen-Age, pour lesquels les marges entourant le texte constituaient un espace de glose et de commentaires. Après la publication *d'Armance* il demanda qu'un exemplaire du roman soit relié "avec une feuille blanche entre chaque feuillet imprimé", et fit fabriquer des exemplaires de ses oeuvres dans lesquels des feuillets blancs étaient intercalés au milieu des feuillets imprimés, exemplaires qu'il donnait à ses amis pour qu'ils y notent remarques et commentaires.

Avec la numérisation le lecteur peut soumettre le texte à de multiples opérations (il peut l'indexer, l'annoter, le copier, le démembrer, le recomposer, le déplacer, etc.), mais, plus encore, il peut en devenir le co-auteur, la distinction, fortement visible dans

seul livre des oeuvres de nature fort diverses, rapprochées dans un même recueil, le lecteur de textes numérisés peut construire des ensembles textuels originaux dont l'existence et l'organisation ne dépendent que de lui et de la nature de ses travaux.

Toute information se transmet au moyen de signes : Charles Sanders Peirce les différenciaient en *indices*, *icônes* et *symboles*. Les signes ne peuvent assurer leur rôle informatif que par leur matérialisation. L'informatique modifie cette matérialisation du signe. A un tel point que certains penseurs contemporains n'hésitent pas à parler d'immatérialité : thème central de l'exposition *Les immatériaux* organisée par le Centre de Création Industrielle du Centre George Pompidou en 1985 (commissaires Thierry Chaput et Jean-François Lyotard). Sur les supports d'enregistrement des systèmes informatiques, les signes n'ont pas d'existence matérielle palpable, au sens où l'on peut dire qu'ils en ont une lorsqu'ils sont figés sur un support non-modifiable, sans destruction partielle ou totale liée à leur rigidité . Ils ne sont pas directement lisibles, mais stockés sous forme de code binaire ayant des propriétés de mobilité, générativité, instantanéité, interactivité, délocalisation : le texte électronique autorise pour la première fois de surmonter une contradiction qui a hanté les hommes d'occident, celle qui oppose, d'un côté le rêve d'une bibliothèque universelle, rassemblant tous les livres jamais publiés, tous les textes jamais écrits, voire, avec Borges, tous les livres qu'il est possible d'écrire en épuisant toutes les combinaisons de lettres de l'alphabet, et de l'autre la réalité, forcément décevante de collections qui aussi grandes soient-elles ne fournissent qu'une image partielle, lacunaire du savoir universel. Cette contradiction avait été perçue par les bibliothécaires Otlet et Lafontaine qui avaient proposé le Contrôle Bibliographique Universel fondé sur des catalogues collectifs, l'adoption de normes de description des documents, de la Classification Décimale Universelle, la coopération et l'échange des bibliographies nationales. La numérisation permet d'aller au-delà des systèmes de messageries et de prêt-interbibliothèques reposant sur les catalogues collectifs, en permettant d'expédier le code binaire constitutif du document dans sa typographie originale. L'exposition *Mémoire du futur* organisée en 1988 par la Bibliothèque Publique d'Information avait annoncé cette évolution. La communication des textes à distance,

prêt-interbibliothèques reposant sur les catalogues collectifs, en permettant d'expédier le code binaire constitutif du document dans sa typographie originale. L'exposition *Mémoire du futur* organisée en 1988 par la Bibliothèque Publique d'Information avait annoncé cette évolution. La communication des textes à distance, qui annule la distinction, jusqu'ici irrémédiable entre le lieu du texte et le lieu du lecteur, rend possible, accessible ce rêve ancien. Sans matérialité et sans localisation, le texte en sa représentation électronique peut atteindre n'importe quel lecteur doté du matériel nécessaire pour le recevoir. Associée par convention à seize bibliothèques et organismes documentaires français dont les fonds sont complémentaires des siens, constituant un vaste réseau d'échange et de partage des collections, la Bibliothèque Nationale de France, avec son catalogue collectif pourra relier entre eux, dans un système informatique intégré tous les catalogues existants (13 millions de notices). Ce système permettra de localiser les documents, mais aussi d'en solliciter la communication ou la reproduction à distance. Déjà, cinquante-quatre conventions de partenariats ont été passés avec des bibliothèques municipales et trente et une avec des bibliothèques universitaires.

### 2.1. Marketing des Postes de Lecture Assistée par Ordinateur :

Comment ce projet s'articule-t-il avec d'autres réalisations en cours ? Le projet de postes de lecture de la Bibliothèque Nationale de France est conçu autour de l'idée de bibliothèque personnelle et de bureaux électroniques. L'offre technique, dans le domaine de l'informatique documentaire, répond en partie aux besoins de la lecture assistée par ordinateur moyennant des adaptations ou des compléments.

La constitution de collections numérisées et l'étude de postes de lecture, par des groupes d'experts issus du monde de la culture écrite, qui adaptent les nouvelles technologies au travail intellectuel de recherche (et non l'inverse), constituent une nouvelle segmentation de l'offre, et une application du principe de *servuction*, service mis au point suivant le principe de la concertation entre usager et agent du service public (sur la notion de *servuction*, cf. : Monsieur Jean-Michel Salaün, *Le marketing de bibliothèques* pp. 62-63. On notera que ce processus de *servuction* en est à son

et que la partie concernant les résultats de l'activité, l'évaluation de la satisfaction de l'utilisateur reste à venir). Ce secteur de recherche n'est pas un sujet d'étude pour les entreprises commerciales qui investissent dans ce qu'il est convenu d'appeler le "multimedia". Il convient de combler une lacune qui risque d'approfondir le fossé des pratiques culturelles entre le monde des chercheurs, des savants, et celui des consommateurs d'images et de sons.

De nouveaux supports et de nouveaux modes de saisie de données, sont disponibles qu'il s'agisse de supports de masse réinscriptibles de très grande capacité (disques durs de 1 à 5 gigaoctets), ou de CD-ROM (disque compact à lecture laser non réinscriptible). Le formidable développement des dispositifs de mémoire artificielle incite le philosophe Bernard Stiegler, coordinateur des réflexions sur les postes de travail de la future Bibliothèque nationale de France, à parler de "nouvelles industries de la mémoire" (Bernard Stiegler : "La technique et le temps t.1, la faute d'Epiméthée, Ed. Galilée & Cité des Sciences et de l'industrie 1994). La démocratisation des micro-ordinateurs à forte capacité de traitement et de mémoire, la possibilité pour le grand public et les bibliothèques d'acquérir des systèmes de numérisation (scanner et logiciel de reconnaissance optique de caractère), la mise en réseau (INTERNET et RENATER), des ressources, créent un environnement favorable à la création de postes de lecture assistée par ordinateur dans le grand public. Se posent alors la question du respect du droit d'auteur et de la vérification de la qualité des éditions proposées et de l'origine des informations. Une réflexion sur la nouvelle approche du texte écrit et du travail intellectuel est indispensable. Pour Bernard Stiegler "la philosophie a refoulé la technique comme objet de pensée. La technique est l'impensé. Les réactions, immédiates ou médiates et médiatisées, "épidermiques" ou calculées, que provoquent les extraordinaires changements dont notre époque est le théâtre, et dont la technique constitue le facteur dynamique le plus puissant, doivent être impérativement surmontées". Pour lui le support dynamique est "un nouvel instrument spirituel -Husserl aurait plutôt parlé d'un nouvel "objet investi d'esprit" -où le texte du lecteur se constitue comme effectivité de sa lecture". Partant de l'expérience de LECAO (Lecture et Écriture Critique Critiques Assistées par Ordinateur

), maquette simulant une interface de lecture assistée par ordinateur, Bernard Stiegler écrit "qu'un processus maïeutique se noue entre le texte lu et le lecteur. Il se trame et se dépose entre eux comme un tissu au cours de la lecture, témoignant de la genèse où elle prend matériellement forme. Sur ce témoin, parce qu'il est dynamiquement supporté, le lecteur peut revenir aussi souvent que nécessaire, accumulant et affinant ses propres traces". (Bernard Stiegler "Machine à écrire et matière à penser").

L'ensemble des institutions éducatives et culturelles est concerné par la notion de Poste de Lecture Assistée par Ordinateur, et de collections numérisées. Il s'agit de redonner une nouvelle dimension aux idées de culture écrite et de collections. Comme nous l'avions noté dans l'introduction, la numérisation des textes peut répondre à plusieurs besoins. On peut numériser pour archiver les documents (économie d'espace de rangement), protéger les collections (conservation des documents rares et précieux, tout en conservant une possibilité d'accès pour le public), améliorer la communication à distance par exemple pour la littérature grise dans le programme Griseli qui permettra des recherches performantes par sujets et la communication du document en texte intégral sur le réseau électronique.

Après numérisation, il est possible d'effectuer sur le document un certain nombre de traitements. Par exemple, on peut gommer les parties du document qui ne sont pas pertinentes, saisir et coller du texte, modifier la taille du document afin qu'il puisse occuper une pleine page à l'écran ou à l'impression, pratiquer l'annotation dynamique déjà décrite. Dans ce cas, on garde le document sous sa forme image (fichier dans un format graphique).

Il est possible d'effectuer une reconnaissance optique des caractères. Le fichier image est alors converti en fichier texte (ASCII ou ANSI). On obtient la constitution d'une collection virtuelle qui n'est plus tributaire du manque de place sur les rayonnages mais pose d'autres problèmes, d'exhaustivité des champs de savoir (tout est-il accessible tout de suite ?), de respect du droit d'auteur, de la propriété industrielle, du secret de la défense nationale, de la remise en cause de l'espace fonctionnel attribué aux collections : " Le CD-ROM sur Beethoven mêlera les

informations écrites sur sa vie, des images, sa musique, pourquoi pas un film tiré de son oeuvre : la discothèque, la bibliothèque, la vidéothèque sont désarmés face à ce support unique. Par ailleurs, ces nouveaux supports sont de moins en moins négligeables dans une bibliothèque, dans la mesure où les informations qu'ils contiennent ne sont pas toujours disponibles sur papier, et l'édition électronique multimédia tend à se développer de façon autonome", ( André-Pierre Syren *La Bibliothèque localement universelle*, B.B.F.t. 40, n°3 1995). Les mémoires optiques numérisées bouleversent l'espace de la bibliothèque : le marketing des Postes de Lecture Assistée par Ordinateur n'est-il pas un moyen de repenser la bibliothèque, non en terme de localisation du document dans l'espace, mais en terme de segmentation de l'offre, après analyse de ce qu'est le travail intellectuel aujourd'hui, des attentes et des contraintes liées à l'assistance au travail de lecture , d'étude et de recherche ?

## 2.2. L'aspect institutionnel :

La Bibliothèque Nationale de France, Département Informatique et nouvelles technologies, se situe dans un schéma directeur qui fixe des missions, à l'ensemble des bibliothèques françaises. Celui-ci intègre :

- la constitution du Catalogue collectif de France :

Cet outil bibliographique qui sera le "langage commun de toutes les bibliothèques de France" embrassera la totalité du paysage documentaire français comprendra :

le catalogue de la Bibliothèque Nationale de France regroupant tous les documents disponibles (imprimés, périodiques, phonogrammes, vidéogrammes, documents électroniques) avec :

- la base BN-OPALE dont le démarrage a eut lieu en 1987. C'est le catalogue informatisé des imprimés. Elle comprend 1 800 000 références, plus de dix millions d'accès indexés , soit tous les livres entrés par dépôt légal, achat, don ou échange depuis 1970,

- les ouvrages anonymes des XVI ème et XVII ème siècles,
- les microformes de sauvegarde commercialisées par le Service de la reproduction,
- le catalogage partagé avec des bibliothèques universitaires ou spécialisées,
- la gestion partagée du dépôt légal avec les bibliothèques habilitées,
- l'accès au catalogue par 400 000 lecteurs chaque année,
- la reproduction de la Bibliographie nationale française (Livres, publications officielles, publication en série, au total 48 000 références par an),
- les bases bibliographiques des départements spécialisés : BN-OPALINE, 340 000 références, plus d'un million d'accès indexés,
- la base BN-SYCOMORE : gestion de la communication des livres et des périodiques aux lecteurs : 400 000 lecteurs par an, 700 000 volumes communiqués par an (à la Bibliothèque Nationale), gestion du récolement des collections de la Bibliothèque nationale de France : sept millions de documents inventoriés depuis 1991 (le récolement de la phonotèque est en cours d'implantation),

Des pôles associés :

- les catalogues des bibliothèques municipales associées, à l'origine au nombre de cinquante trois. Un plan de conversion rétrospective des notices concernant les ouvrages anciens, c'est-à-dire imprimés avant 1811, a été élaboré avec la direction du Livre et de la lecture du ministère de la Culture,

Les Bibliothèques Universitaires ou de recherche :

- le réseau des bibliothèques universitaires regroupant les applications informatiques utilisées actuellement : messagerie du Prêt entre bibliothèques, catalogues collectifs issus du PANCATALOGUE, et regroupant les bases issues des systèmes SIBIL , OCLC, BN-OPALE, Catalogue Collectif National concernant les périodiques, RAMEAU (répertoire encyclopédique concernant les vedettes matières utilisées pour indexer les documents).

Sous la direction de l'Agence Bibliographique de l'Enseignement Supérieur créée par le décret n°94-921 du 24 octobre 1994, établissement public à caractère administratif sous tutelle du Ministère de l'Education nationale, un plan de

développement du réseau des Bibliothèques Universitaires, doit accompagner une refonte des systèmes existants dans un ensemble cohérent : le SU (Système Universitaire), dont la mise en oeuvre devrait être amorcée en 1996. La création de ce système devrait grandement faciliter le partenariat avec le catalogue collectif de la BNF.

- Une coopération avec les bibliothèques qui ne participent pas aux réseaux de l'Education nationale ou de la Culture mais dont les fonds sont intéressants pour la recherche est également prévu.

L'informatique jouera un rôle vital dans le fonctionnement de la Bibliothèque de France, et se trouvera au coeur du système d'information, information étant prise dans le sens le plus large. C'est en effet grâce à des moyens et selon des procédures informatiques que le système d'information remplira sa mission : rendre accessible à l'extérieur comme à l'intérieur de l'établissement un très grand nombre de documents. Dans ce cadre les opérations de numérisation permettront la constitution d'un important corpus électronique consultable par le PLAO.

Le système informatique comprend actuellement 1000 terminaux connectés, plus 70 milliards de caractères en ligne avec :

le système GEAC base BN-OPALE qui comprend trois machines GEAC 9000 :

-un GEAC 9000, 7 processeurs-32 Mo de mémoire adressable - 15 Go de données (1 Go octet représente 1 milliard de caractères), 450 terminaux connectables,

-un GEAC 9000 , 4 processeurs- 24 Mo- 12 Go de données, 70 terminaux connectables,

-un GEAC 9000, 6 processeurs-24 Mo- 8 Go de données, 159 terminaux connectables

Pour le système Bull de la base BN-OPALINE :

un serveur utilisant une unité centrale BULL DPS 7000 / 705 2, processeurs-64 Mo- 22 Go octets de données, 210 terminaux connectables

Le système Sybase avec le logiciel SYCOMORE qui tourne sur :

-un serveur UNIX HP 9000 H40 avec 64 Mo de RAM et 10 Go de données, relié à des micro-ordinateurs sous UNIX, à des PC émulés comme postes multi-services et des terminaux Wyse VT 200 pour la consultation seule.

Le système ORACLE : base EVER-Reliure, LORIS-EVER :

un serveur BULL DPX-20 à base de processeur RISC avec 80 Mo de RAM et des PC 486 connectés en réseau, plus des terminaux.

Les objectifs du futur système informatique de la Bibliothèque Nationale de France à Tolbiac (en liaison avec les pôles associés et le système universitaire et de recherche), sont de proposer un service de communication performant pour les inscriptions, les réservations à distance de places de lecteurs et de documents, l'information sur les diverses activités de la BNF, la consultation conviviale du catalogue, la consultation et le travail sur les fonds numérisés (Poste de Lecture Assistée par Ordinateur). Un catalogue unique sera constitué pour tous les types de documents (1996), livres, périodiques, documents sonores, audiovisuels, informatiques et numérisés soit huit millions de notices bibliographiques. Le catalogue assurera le lien direct avec les fonctions vitales de la communication ou des acquisitions. Il sera la pièce maîtresse du Catalogue Collectif de France qui recensera les fonds les plus importants des bibliothèques françaises. L'utilisation des nouvelles technologies et, en particulier, des techniques de numérisation et de transfert sur support électronique, permettra :

-une meilleure préservation et une communication beaucoup plus large des originaux grâce aux copies électroniques de substitution fournies aux lecteurs,

-l'utilisation par le public d'outils performants de manipulation et d'exploitation des fonds de la BNF, tels les Postes de Lecture Assistée par Ordinateur,

-une transmission à distance, à d'autres institutions ou bibliothèques partenaires, en utilisant les réseaux de télécommunication.

Le système informatique gèrera autant les documents que les espaces et les services. Cette gestion globale a été décomposée en une vingtaine d'applications regroupées en trois grandes familles :

1°) Les services d'accès à la bibliothèque et à ses collections :

système d'information et d'orientation du public

gestion des accréditations des usagers

consultation du catalogue

assistance linguistique à la consultation du catalogue

gestion des demandes des lecteurs pour obtenir places, documents et services

gestion de la consultation automatisée des fonds numérisés (Postes de Lecture Assistée par Ordinateur)

contrôle de la circulation des documents.

2°) Les services d'assistance aux opérations de type scientifique ou technique :

gestion des acquisitions (sélection, achats, entrées, contrôle du dépôt légal)

production du catalogue

gestion de la conservation des collections

gestion du magasinage

gestion des ateliers de traitement des documents

gestion des moyens de transport

3°) Les services qui concourent au bon fonctionnement de l'établissement :

gestion des ressources humaines, financières, logistiques, des bâtiments, des colloques.

Ce système sera ouvert et évolutif, c'est-à-dire fondé sur les normes et standards nationaux et internationaux de l'informatique, ceci afin de faciliter son installation, son évolution, sa durabilité et son indépendance vis-à-vis du matériel (voir l'article de Claire Vayssade : Le dépôt légal des documents informatiques (B.B.F. t. 40 n°3, 1995) " la seule manière de préserver l'accès au contenu [des documents informatiques] est donc de les détacher des contraintes des supports et de l'environnement matériel, voué à l'obsolescence, et d'opérer une mutation sur de nouveaux supports. Il s'agit là d'archivage dynamique(...) A chaque génération technologique, les informations devront être stockées sur des supports durables normalisés (pour l'instant CD-ROM ISO 9660). Il faudra également prévoir la gestion des liens et des index permettant de retrouver les informations, c'est-à-dire la liaison entre le contenu et le logiciel d'interrogation. Cette gestion des liens et des index suppose une organisation logique

d'interrogation. Cette gestion des liens et des index suppose une organisation logique normalisée et stable dans le temps". Ce qui est vrai pour le dépôt légal des logiciels l'est aussi pour d'autres parties du système : la tendance actuelle de la recherche en informatique est de faire émerger un type de programmation indépendante des machines (programmation applicative, fonctionnelle, logique suivant les écoles de pensée), et dans ce cadre le système d'exploitation UNIX, standard des systèmes ouverts, en constante évolution permet d'éviter les *oukases* de telle ou telle firme commerciale.

### 2.3. Relation avec d'autres recherches actuelles : la notion d'hypertexte

L'hypertexte accompagné d'un système d'annotation est une composante des Postes de Lecture Assistés par Ordinateur : alors que l'écriture était une invention qui rendait "le langage visible", la lecture et l'écriture assistées par ordinateur tentent de permettre à l'activité du "lecteur idéal, atteint d'une isomnie idéale" (Umberto Eco), d'être presque instantanément inscriptible dans une mémoire dite "de masse". Tel est le sens des études du groupe de travail qui s'est constitué en septembre 1989, avec pour président, Bernard Stiegler. Cette recherche rejoint une question plus vaste relative aux implications de ces nouvelles technologies sur la lecture et l'écriture, de l'annotation structurée. "La lecture assistée par ordinateur pourrait bien induire un nouveau type de lecture, à l'instar de ce qui s'est produit avec l'apparition du livre", explique Jacques Virbel, l'un des membres du groupe d'étude, chercheur à l'Université Paul Sabatier de Toulouse. La notion d'OPAC, On line Public Access, reposait jusqu'ici sur la description des attributs externes des documents par ailleurs classés en accès libre sur des étagères où ils pouvaient être consultés. On connaît les critiques faites aux OPAC : d'une part malgré l'affinage des techniques d'indexation, seule une partie des thèmes contenus dans un livre apparaissent, d'autre part les modes de classement, le vocabulaire des indexeurs et du public peuvent être différents. En permettant à des chercheurs de travailler sur un corpus digitalisé en texte intégral et d'utiliser un logiciel comme "Phrasea", qui a pour fonction de générer des bases de données

par l'utilisateur (structure de l'unité documentaire, famille de synonymes etc.), permettant la constitution d'un dictionnaire et d'un index des mots du texte, on peut espérer qu'une nouvelle époque pour l'étude de l'indexation des attributs internes des documents va s'ouvrir. En effet si le travail d'indexation, d'annotation du chercheur est personnel, et disjoint de la base de données générale pour être chargé sur une mémoire de masse personnelle, à l'accès réservé, on peut néanmoins supposer que la multiplication des travaux de ce type amènera une augmentation des possibilités d'études des pratiques de lecture et d'indexation, d'annotation, savantes. A l'époque de la constitution de réserve de mémoire électronique, il est capital de savoir comment ces mémoires peuvent être indexées, segmentées, interrogées.. C'est une préoccupation qui rejoint la bibliothéconomie et les sciences de l'information, mais aussi la psychologie, la recherche en informatique, la sociologie (la "raison pratique" et les "habitus" de Pierre Bourdieu), les sciences cognitives et celles de la communication, la structuration d'un nouvel *espace public* de la communication via Internet (deux cents bibliothèques sont accessibles par ce réseaux et les scientifiques l'utilisent quotidiennement pour communiquer par des forums et des périodiques électroniques).

George P. Landow (Brown University) dans son ouvrage *The digital word : text-based computing in the humanities*, MIT press, 1993 attache une grande importance au projet de digitalisation de textes de la Bibliothèque nationale de France : pour lui un système intégré donnant accès à une masse de textes digitalisés pourrait donner à la France un avantage significatif dans la société du savoir du futur, avantage comparable à celui qu'Amsterdam a retiré au XVII<sup>ème</sup> siècle de la technologie de l'imprimerie et de l'édition.

Or, l'hypertexte, moyen de mettre à jour les connaissances en interrogeant les points cachés d'un texte par un système de renvoi vers d'autres textes, s'il connaît un développement considérable, est l'objet de quelques réserves :

- "Les techniques d'hypertexte permettent une circulation dans le catalogue plus proche de la démarche naturelle que les catalogues classiques, où elle est contrainte par une structure arborescente ; l'usager a plus d'initiative et plus de possibilités de

modifier sa démarche en cours de recherche. Elles permettent une plus grande flexibilité : des parcours, par exemple, adaptés selon les types d'utilisateurs. Elles sont particulièrement adaptées aux situations d'apprentissage de l'outil informatique.

En revanche, elles présentent des inconvénients, signalés par R. Rada :

- surcharge cognitive si le modèle est inadéquat,
- risque de désorientation de l'utilisateur après un certain nombre de bifurcations,
- la construction de liens est une opération très lourde. (Roy Rada, *Introduction and overview to Perspective on hypertext*, Journal of the American society of Information science, 40, 1989, 3, P.159-163).

Les OPAC proposant un mode de navigation relevant de l'hypertexte sont très peu nombreux. Différentes possibilités ont été identifiées (mais rarement implémentées) :

- définitions de liens entre les termes d'une notice et les notices contenant le même terme,
- construction de liens hypertexte par l'utilisation d'un thésaurus (c'est le cas du prototype Hypertext développé à Aberdeen),
- création de liens d'aide,
- établissement de liens aux références citées ou à des documents liés par des relations de nature bibliographique (éditions différentes, traductions, éditions critiques, relations d'inclusion).

L'hypertexte permet d'autres possibilités : offrir à l'utilisateur une fonction d'annotation ou définir des parcours différents selon les niveaux de lecture par exemple. (cf : *Les catalogues en ligne : enquête à la Médiathèque de la Cité des sciences et de l'industrie, panorama des recherches*, coord. Mohamed Hassoun, Danielle Roger, Enssib-Cersi, 1994).

Il y a deux possibilités d'usage de l'hypertexte : comme interface avec un OPAC, et comme moyen de travail personnel sur un texte. La création des liens est alors perçue différemment : dans l'OPAC cette création, affaire de spécialistes est préalable à l'interrogation, n'est pas modifiable directement par l'utilisateur. Dans le travail sur un texte numérisé il peut être envisagé que l'utilisateur crée des liens

-travail de création ou d'étude sur les textes

En interrogeant des expériences existantes : notamment Intermedia développé à Brown University par Norman Meyrowitz et le groupe interdisciplinaire de l'IRIS (Institute for Research in Information and Scholarship), décrit par George P. Landow professeur de cette université dans son ouvrage "HYPERTEXT The Convergence of Contemporary Critical Theory and Technology" Johns Hopkins University Press, Baltimore and London, ), et en cherchant les possibilités de rapprochement de la norme de description des documents ( DOD de SGML), et de la programmation en logique (PROLOG comme dans la base de donnée P/FDM de l'Université d'Aberdeen ), ce qui implique l'étude des normes en vigueur, SGML, HTML, HyTime.

Le projet de poste de lecture assisté par ordinateur (PLAO) jugé utopique et coûteux par certains à l'origine, se trouve encouragé par les progrès de la recherche mondiale : progrès rapide de la reconnaissance optique de caractère (OCR), qui permet la fabrication de scanners à des prix abordables et même l'interprétation exacte de l'écriture manuscrite par l'ordinateur, baisse générale des prix du matériel informatique à forte capacité de stockage (disques durs d'un ou plusieurs gigaoctets, CD-ROM de 500 à 600 megaoctets pressés pour des coûts de plus en plus faibles et donnant lieu à une activité éditoriale importante), et surtout normalisation des formats d'échange avec l'élaboration de méthodes permettant un accès universel à l'information : norme Z 39-50, code Ascii, normes SGML, HTML, HyTime, adoptées par les imprimeurs et éditeurs de CD-ROM, les chercheurs, les entreprises de presse scientifique, les concepteurs de réseaux informatiques. Nombre de concepts qui semblaient extraordinaires en 1990 sont aujourd'hui englobés dans des produits commerciaux de grande diffusion, ou dans des logiciels du domaine public mis en accès libre sur des serveurs tel celui du CERN de Genève en ce qui concerne le langage d'hypertexte HTML.

### 3. Nouvelles technologies et culture humaine : hypertexte, mémoire et érudition contemporaine.

Alors que les stations de travail étaient à l'origine utilisées massivement par les chercheurs des sciences dites "dures", tels les physiciens, les universitaires américains spécialisés dans les études littéraires intègrent l'hypertexte depuis de nombreuses années dans leurs études et leur enseignement.

Pourquoi ce mouvement est-il profond et s'emplifie-t-il depuis les premières intuitions de Theodore Nelson, étudiant en philosophie dans les années soixante ?

Dans le passé les discussions sur les commentaires appropriés aux diverses formes d'éditions , éditions historico-critiques, éditions d'études, étaient toujours dépendantes du support -livre. Comme l'espace offert était nécessairement limité le commentaire devait être coupé en conséquence. Bien qu'il existe un large consensus parmi les éditeurs sur le fait que l'édition n'est pas une science auxiliaire mais qu'elle est la base même de tout travail philologique et représente déjà une partie de l'interprétation, la distinction entre commentaire purement informatif et commentaire interprétatif, la distinction, donc, entre ce qui est à reprendre et ce qui est à exclure, subsiste dans la majeure partie des éditions historico-critiques.

Dans un domaine en apparence très éloigné des sciences, la musique, la lecture de la machine sans faille, sans délai, est devenue familière. Son objectivité est totale, comme dans le programme CHANT de l'IRCAM, qui produit une voix sans musicalité parce qu'absolument exacte. L'utilisation de l'ordinateur en tant que traitement de texte musical et outil de notation graphique est aujourd'hui devenu une réalité pour beaucoup de musiciens, il se révèle fondamentalement un outil d'élaboration et de manipulation de représentations multiples, de production, et non seulement le simple prolongement des pratiques existantes. Il s'agit d'un outil générique, qui peut englober et surtout unifier toutes sortes de représentations. Pour Pierre Boulez qui

évoque les “trouvailles de Joyce ; dans *Finnegans Wake* lorsqu’il tord le vocabulaire, lorsqu’il croise les mots les uns avec les autres, lorsque plusieurs langues se rencontrent ...”, les idées musicales “n’existent que réalisées dans l’écriture. (...) L’improvisation est une écriture, bien qu’immatérielle. Il n’y a pas dans notre civilisation, de phénomène qui soit délié de l’écrit. Et même dans les civilisations de tradition orale, il existe des *grammaires*, comme celles des langages tambourinés africains. L’écriture préside à tout, qu’elle soit réelle ou virtuelle” (Pierre Boulez *Texte et pré-texte* revue Genesis, 4, 1993). L’utilisation de l’ordinateur pour la musique représente l’aboutissement de ce que Hugues Dufourt a désigné comme l’artifice de l’écriture dans la culture occidentale (cf. : *L’artifice de l’écriture dans la musique occidentale*, Critique numéro 359, 1981). En effet , une utilisation conséquente de l’ordinateur peut amener le compositeur à un processus de formalisation sans précédent dans l’histoire de la musique, à travers un contrôle explicite de l’ensemble des phénomènes musicaux, dont l’enjeu est de les faire passer dans l’espace du calcul et de la manipulation symbolique, ce qui les rend alors accessibles à la composition. Cette conception se rapproche de l’idée de l’hypertexte littéraire de George P. Landow, s’inspirant du concept de *texte idéal* de Roland Barthes, texte étendu à d’autres textes et abolissant la distance entre auteur, texte et lecteur.

### 3.1. PLAO et hypertexte :

Ce type de texte offre l’extension la plus grande possible, sans que cela entraîne pour autant un manque de lisibilité. Les éditeurs d’une telle édition sont en mesure de donner toutes les explications qui leur semblent nécessaires et d’indiquer toutes les correspondances qui leur apparaissent. On peut consulter à ce sujet l’ouvrage de George P. Landow : *Hypertext. The Convergence of Contemporary Critical Theory and Technology* , The John Hopkins University Press, Baltimore and London, 1992, p.96-99, qui commente entre autres réalisations, *The Dickens Web* et *In Memoriam* du projet *Intermedia* de Brown University. The Dickens web est un ensemble de documents multimedias intégré dans le projet d’hypertexte Intermedia. Cet ensemble contient 245 documents et 680 liens, et rassemble une collection de documents en

relation avec Charles Dickens et son roman *Great expectations*, replacé dans son contexte : le règne de la Reine Victoria, les problèmes de la pauvreté et de la santé publique dans l'Angleterre du XIX<sup>ème</sup> siècle, ainsi que divers aspects religieux et culturels de la vie quotidienne. George P. Landow décrit son travail en référence aux *lexies* de Roland Barthes, à l'idée exprimée dans *S/Z* que le texte est sans cesse et de part en part traversé par des codes, mais il n'est pas l'accomplissement d'un code (par exemple le code narratif), il n'est pas la "parole" d'une "langue" (narrative). Ce changement de perspective, fait entrer la lecture dans le monde des connotations, et dans le chapitre liminaire de *S/Z*, Roland Barthes invoque l'instrument de la connotation comme étant celui par lequel il convient d'approcher les textes classiques. Pour George P. Landow la notion de connotation rejoint les techniques de l'hypertexte (liens, noeuds, réseau, tissu et chemins), puisque dans le "texte idéal" de Barthes "les réseaux sont nombreux et en interaction constante, sans que l'un surpasse le reste ; le texte est une galaxie de signifiants, non une structure de signifiés ; il n'a pas de commencement ; il est réversible" (Roland Barthes *S/Z*, Paris : Éditions de Seuil, 1970). Les codes repérés par Roland Barthes, notamment lors de la lecture de *Sarrazine* d'Honoré de Balzac, (code proairétique ou code des actions, sémique : code des sèmes psychologiques, culturel : code de savoir, herméneutique : code de recherche, de quête d'une vérité, d'une solution), peuvent être utilisés pour créer des liens d'hypertexte, comme ceux décrits par George P. Landow autour du poème d'Alfred Tennyson *In Memoriam*, avec le logiciel Intermedia.

A cette utilisation nouvelle de l'hypertexte, on peut ajouter des fonctions de bureau électronique et de bibliothèque personnelle. Comme l'indique Bernard Stiegler "ce poste de lecture est conçu comme un bureau électronique en même temps que comme une bibliothèque personnelle. Une fois constitué son corpus personnel, le chercheur doit pouvoir effectuer un traitement numérique du texte, des analyses et des exploitations de contenus profitant des ressources propres aux supports magnétiques et enrichissant de la sorte l'accès du lecteur au texte- ce qui ne signifie en aucune manière que la lecture électronique se substitue à la lecture traditionnelle : elle vient l'amplifier.

en aucune manière que la lecture électronique se substitue à la lecture traditionnelle : elle vient l'amplifier.

Un aspect essentiel de la lecture est la corrélation. Lire, c'est alors rassembler sous des mots clés, structurés en index, en thesaurii, sur des fiches, dans des dossiers, par des notes, citations, remarques, commentaires. Tout ce qui est recueilli au cours des opérations de constitution de corpus, d'annotation, d'extraction, de documentation et de commentaire, caractéristiques de la lecture du chercheur, qui se mue toujours immédiatement en écriture, tout cela, pour former un véritable appareil de lecture-écriture, doit être mis en réseau, relié, corrélé. Cela se manifeste par des activités effectuées à divers rythmes. Les fiches sont une manière de créer de tels liens. Mais c'est dès le moment de l'annotation que cette opération est effectuée par anticipation : l'application d'un même mot clé à deux passages, où le mot n'est pas nécessairement présent dans les textes commentés eux-mêmes, est la création d'un lien conceptuel, d'une corrélation, l'élaboration d'une "grille de lecture". La mise en relation s'effectue par comparaisons, juxtapositions, réductions, assimilations, etc. La lecture est cette alchimie. Ce n'est pas une chimie parce qu'aucune technique de lecture, aussi efficace qu'elle puisse apparaître, n'éliminera jamais le travail de taupe qui, comme "dans le dos de la conscience", guide profondément la juste lecture.

*Corréler*, c'est accomplir la lecture dans une écriture qui est sériation. Le problème est qu'en très grande part, ces corrélations, lorsqu'elles sont effectuées selon les méthodes traditionnelles, sont seulement entrevues au cours des lectures et ne se réeffectuent pas lorsque cela devient nécessaire. C'est encore la mémoire psychologique qui effectue le tri dans son essentielle inconscience : refermés, empilés sur le bureau, alignés dans la bibliothèque, les livres sont autant de cryptes sans autre commune mesure que les faibles souvenirs du lecteur, chemins aléatoires traversant ces arcanes selon la mouvante géologie de sa mémoire. C'est cette situation que vient bouleverser la lecture assistée par ordinateur. Le principal avantage de la numérisation est, outre le stockage sur un même support d'énormes quantités de textes (et des images et des sons pouvant accompagner ces textes), la possibilité de lire à la vitesse de la lumière, par exemple en adressant au fonds numérique une

l'intégralité du texte numérisé - en quelques fraction de seconde. C'est là un des bénéfices les plus immédiatement appréhendables de la délégation du savoir lire aux machines cybernétiques. Mais lire avec la machine, c'est aussi et surtout écrire - c'est-à-dire corréler, sous forme électronique - à partir de cette lecture : c'est organiser, au sein de fichiers hypertextuels, de bases de données personnelles, de systèmes d'archivage électronique, des séries corrélatives, sans perte de mémoire; c'est pouvoir en afficher en permanence les structures, intervenir sur les séries et réduire les structures de manière à limiter le bruit qu'engendre cette puissance de mise en échos et miroirs, et finalement accéder à de nouvelles séries et métastructures. En ce sens, la machine apporte une dimension absolument originale à la lecture" (cf : Bernard Stiegler : *Machines à lire*).

La notion de transduction est au coeur de cette question : on appelle transducteur un dispositif assurant la transformation d'une grandeur physique en une autre (par exemple la pression d'une onde sonore ou d'une vibration en signal électrique comme dans un haut-parleur, un microphone). Un scanner qui transforme les lettres d'un texte en code numérique utilisable par l'ordinateur répond à la définition première du transducteur. Bernard Stiegler, s'inspirant de la conception de la transduction du philosophe Gilbert Simondon, analyse la lecture comme un phénomène transductif :

"...la relation entre le texte et son lecteur (et son auteur comme premier lecteur) est transductive , au sens que Simondon 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.

Le point de départ de la question simondonienne de l'individuation est la critique du schème hylémorphique et du substantialisme :

Il existe deux voies selon lesquelles la réalité de l'être comme individu peut être abordée : une voie substantialiste [...]; une voie hylémorphique [...] : toutes deux supposent qu'il existe un principe d'individuation antérieur à l'individuation elle-même, susceptible de l'expliquer, de la produire, de la conduire (Gilbert Simondon, *L'Individu et sa genèse physico-biologique*, PUF, 1964).

Ces deux voies se donnent l'individu déjà constitué pour rendre compte de l'individuation. Or, il s'agit de ne se tenir ni dans l'unité, ni dans la dualité, mais dans un processus : il s'agit de "connaître l'individu à travers l'individuation plutôt que l'individuation à partir de l'individu". L'individu reflète le processus, qui l'englobe et le dépasse, le traverse et le tend, il n'en est pas l'origine, mais un moment ou une phase :

L'individu serait alors saisi comme une réalité relative, une certaine phase de l'être qui suppose avant elle une réalité préindividuelle, et qui, même après l'individuation, n'existe pas toute seule, car l'individuation n'épuise pas d'un seul coup les potentiels de la réalité préindividuelle, et d'autre part, ce que l'individuation fait apparaître n'est pas seulement l'individu mais le couple individu-milieu. L'individu est ainsi relatif en deux sens : parce qu'il n'est pas tout l'être, et parce qu'il résulte d'un état de l'être en lequel il n'existait ni comme individu ni comme principe d'individuation" (ibid p.4).

Lire est toujours écrire et inversement précisément dans cette mesure où le texte n'existe que dans sa manifestation, comme le dit Gadamer (*Vérité et méthode*), ou la lecture est alors un processus d'individuation du texte lu autant que de son lecteur, manifestation individuante du texte qui coïncide avec l'individuation de la textualité du lecteur, pleinement accomplie lorsqu'il inscrit les effets de sa lecture quelque part, que leur forme en soit textuelle ou non. Le texte est alors pour le lecteur un milieu préindividuel dont la textualité joue en effet comme rapport transductif entre lecteur et texte constitué par leur relation. Celle-ci est à son tour déterminée par les spécificités du support mis en jeu. Le texte est donc un objet qui ne se constitue que dans les conditions d'accès qu'y ménage son support matériel, conditionnant la nature de la relation du lecteur à ce qu'il lit par une mise en forme matérielle dont la performativité précède le partage hylémorphique. En oubliant la matérialité de cette relation, dont le support dynamique n'est qu'un cas, c'est le lecteur lui-même qu'oublie une théorie du texte. Ces conditions d'accès sont l'effectivité de la relation transductive, elles surdéterminent ce qu'elles mettent en relation.

### 3.2. Condition de production de l'hypertexte

Il s'agit d'un travail de recherche et d'étude mené par une équipe de spécialistes. Toutes les parties consacrées au commentaire sont développées, on ne se contente plus d'établir des relations entre des passages du texte et des commentaires, mais les commentaires concernant des passages isolés sont rapprochés les uns des autres et ceux-ci, à leur tour, sont mis en rapport avec un commentaire global. Il peut s'agir de tisser un réseau à partir de mots isolés ou de paragraphes entiers, les possibilités de connexion sont illimitées. Les aspects contradictoires à propos d'un passage d'un texte peuvent être mis en évidence. Le fil conducteur qui devra guider les choix reste l'objectif principal que s'est fixé l'édition informatique, à savoir assurer l'accès facile -mais non simpliste - au texte. Il ne s'agit nullement de faire naître l'impression d'une univocité et d'une sécurité qui n'existent pas. Cela reviendrait à nier l'autonomie de l'oeuvre d'art et à la réduire au statut d'énigme à déchiffrer - retour en force du message caché de l'auteur ! Il s'agit de présenter le commentaire comme une suggestion dûment fondée, que le lecteur peut accepter ou refuser. La prétention à la vérité y perd son aura, mais non pas son attrait, qui est même rehaussé par la possibilité d'appeler dans un second temps, à côté de la remarque primaire (qui apparaît sur demande dans une fenêtre d'information et fournit les autres indications sous forme de mot-clef), des éléments qui la complètent, ou la contredisent, et permettent de comparer les deux. Ce genre de mise en réseau peut être poursuivi *ad infinitum* et contribue ainsi à favoriser une attitude de lecture active et à ouvrir à de nouvelles synthèses. C'est précisément en raison de la richesse de ses annotations qu'une telle édition demeure fidèle à l'idéal de l'ouverture du texte : "une oeuvre classique ne doit jamais pouvoir être totalement comprise", écrit Fr. Schlegel ( cité par Dierk Hoffman *Édition-rhizome : à propos d'une édition historico-critique fondée sur le concept d'hypertexte d'hypermedia* revue Genesis n°4, 1994). La confrontation avec une oeuvre demeure selon Peter Szondi, "une reconnaissance en perpétuel devenir ". Sur cette incalculable puissance du message artistique on peut également rappeler ce qu'écrit Gilles Deleuze à propos de Marcel Proust : " Proust ne veut pas dire autre chose en nous conseillant non pas de

lire son oeuvre, mais de nous en servir pour lire en nous" (Gilles Deleuze *Proust et les signes*, PUF, 2<sup>ème</sup> éd. 1970). Dans l'échelle des apprentissages proustiens l'accès à un langage transparent (analytique, logique), ne constitue pas la performance la plus haute. "La première, Françoise me donna l'exemple (que je ne devais comprendre que plus tard) que la vérité n' a pas besoin d'être dite pour être manifestée " (Marcel Proust *Le Côté de Guermantes*, Gallimard, Bibliothèque de la Pléiade, tome II p. 66). Pour Jacques Derrida on construit sur une valeur qu'on n'interroge pas, la *présence*. La métaphysique de la présence pense selon deux temps (logiques et souvent historiques) : présence, d'abord, du monde à un regard, d'une conscience à elle-même, d'un sens à l'esprit, de la vie en soi, d'une sensation à un corps; absence ensuite - le monde voilé, la conscience égarée, le non-sens, la débauche, le langage, le sevrage. En pensant le deuxième temps comme dérivé par rapport au premier, on ramène, ne fut-ce qu'en pensée, le complexe au simple, le secondaire au primaire, le contingent au nécessaire. C'est l'ordre même de la raison et du sens, du *logos*, et on y échappe pas aussi facilement que peuvent le croire ceux qui nomment rapidement l'inconscient ou la matière ou même l'autre. Une pensée qui ne s'ordonne plus suivant ces schémas conduit à l'idée développée dans *De la grammatologie* qu'il n'y a pas de hors-texte (voir également Geoffrey Bennington *Jacques Derrida*, éd. du Seuil, 1991), et qu'on "peut sans trop de peine, présenter *La pharmacie de Platon* comme une sorte de lecture indirecte, peut-être distraite, de *Finnegans Wake* qui mime, entre Shem et Shaun, entre le *penman* et le *postman*, jusque dans le détail le plus fin, le plus finement ironisé, la scène du *pharmakos*, du *pharmakon* et les diverses fonctions de Toth, *th'other*, etc. Je ne pourrais pas reconstituer ici même de loin, l'extrême complexité du réseau", écrit Jacques Derrida. "J'avais dû me contenter de jouer d'une seule note (*La dissémination*, Paris, Le Seuil, 1972, p.99, note 17) à rappeler que bien entendu, comme "on l'aura vite compris", l'ensemble de *La pharmacie de Platon* n'était qu'une lecture de *Finnegans Wake* ". Une lecture possible entre autre" (*Ulysse Gramophone : deux mots pour Joyce*, Jacques Derrida, Editions Galilée, 1987).

Partant de cette idée qu'il n'y a pas de hors-texte, certains chercheurs envisagent l'hypertexte comme "une révolution potentielle dans les études littéraires".

### **3.2.1. Le modèle hypertexte propose un mode particulier d'accès au texte et à l'information.**

En outre les applications multimédias demandent simultanément l'utilisation de plusieurs médias qui seront transportés par les réseaux de télécommunications de façon non intégrée dans un premier temps et de plus en plus intégrée dans le futur. La recherche d'information dans les réseaux n'est pas statique mais dynamique. L'information sous la forme électronique évolue rapidement, l'édition sur support optique est florissante et le CD-ROM permet de constituer des collections de plus en plus diversifiées. Beaucoup d'auteurs estiment que nous nous trouvons dans une période de changements culturels et technologiques fondamentaux analogues à la révolution de Gutenberg. Toutefois nous connaissons les leçons du passé : les travaux d'auteurs comme Roger Chartier, Elisabeth Eisenstein, Alvin Kerman qui ont étudié la période complexe de transition du manuscrit à la culture imprimée suggère trois idées principales :

- la première est que ce type de changement peut se dérouler sur des périodes longues. Les études sur l'influence de la technologie sur les modes de lecture et d'écriture mettent en lumière plusieurs centaines d'années de changements graduels durant lesquels les pratiques de lecture, les modes de publication, furent lentement modifiés. Selon Kerman, c'est seulement vers 1700, que la technologie de l'imprimerie fit passer les pays les plus avancés d'Europe d'une culture orale à une culture de l'imprimé (Alvin Kerman, *Printing Technology, Letters on Samuel Johnson* Princeton, Princeton University Press, 1987).

-le second point est qu'il est difficile de déduire tous les changements apportés par les relations des nouvelles techniques d'inscription et de la culture écrite.

-le second point est qu'il est difficile de déduire tous les changements apportés par les relations des nouvelles techniques d'inscription et de la culture écrite. Comprendre la logique d'une technologie ne permet pas de faire des prédictions, parce qu'en relation avec des conditions sociales, historiques, et politiques conditions multiples, la même nouvelle méthode de travail peut produire des effets contradictoires. Marie-Elizabeth Ducreux et Roger Chartier que le texte imprimé et le manuscrit fonctionnaient comme instrument d'acculturation religieuse contrôlée, par les autorités, mais dans certaines circonstances, ils étaient le support de la résistance pour des croyances persécutées, l'ultime recours contre les conversions forcées (cf. Marie-Elizabeth Ducreux : *Lire à en mourir. Livres et lecteurs en Bohême au XVIII ème siècle* . dans l'ouvrage sous la direction de Roger Chartier : *Les usages de l'imprimé*).

- Le troisième point est que la relation entre les pratiques de lecture, l'information et les technologies, la culture, se situent dans un contexte politique, et ont des implications concernant l'organisation sociale, économiques, les modes de circulation et de répartition des biens, les choix stratégiques intervenant dans la prise de décision économique (cf : Fritz Machlup dans : Roberta Fausti, *Recherche sur l'économie de l'information, aperçu de la littérature américaine*, ENSSIB, 1993 )

Les critiques littéraires américains et anglo-saxons attachent une grande importance à la convergence des théories critiques sémiologiques contemporaines (Barthes, Derrida, Foucault, Lyotard), et de la technologie symbolisée par l'idée d'hypertexte conçue par Teodor Nelson dans les années soixante, dernier maillon de tout un courant de pensée que l'on fait remonter à Raymond Lulle (1235-1315), Pierre de la Ramée (1515-1572) et au XX ème siècle Paul Otlet (1868 - 1944) auteur d'un projet de bibliographie universelle et Vannevar Bush ( conseiller scientifique du président Roosevelt ), Douglas Engelbart de l'Université de Stanford , auteur d'un article retentissant : "Un cadre conceptuel pour l'accroissement de l'intellect humain " et inventeur de la célèbre "souris", périphérique de pointage .

Les notions de compilateur, balisage de texte, de formatage, de scanner et de parser, d'interface sont la partie cachée ( un ensemble de *métaconnaissances* au sens donné par Jacques Pitrat à ce terme "des connaissances pour manipuler les

casse liée à celle des mathématiques. Et plus spécialement des statistiques qui ne sont apparues qu'à la fin du quatorzième siècle (...) En effet il ne suffisait pas de fondre un à un tous les caractères, chiffres, lettres, ponctuations, ligatures et abréviations, nécessaires pour la composition d'un texte. Il fallait aussi tenir compte de la fréquence de telle ou telle lettre dans telle ou telle langue. Il fallait plus de *t* et de *h* en anglais qu'en français par exemple. Il fallait donc constituer des ensembles ou polices, différents pour différentes langues : polices françaises, allemandes, anglaises, etc. Et autant de casses. Car les caractères n'y étaient pas répartis selon l'ordre alphabétique mais selon l'ordre statistique. Chaque casse ne contenait qu'un seul caractère dans un seul corps. Rares étaient les imprimeurs qui avaient les moyens de se payer plusieurs caractères dans tous les corps et qui pouvaient les rentabiliser par un usage courant". (Fernand Baudin, *L'effet Gutenberg*, Editions du Cercle de la Librairie, 1994). Dès l'origine l'imprimerie impliquait une notion de méta-connaissances sur le texte, qui annonçait les travaux de la bibliométrie. C'est également avec Gutenberg qu'apparaît en Occident une manière de lire, en silence et par les yeux, qui fait rupture avec la lecture nécessairement oralisée longtemps universelle (ou presque). Les avancées de cette compétence nouvelle à l'âge du manuscrit, progressivement conquise par les copistes monastiques, puis les milieux scolastiques et universitaires, enfin les aristocraties laïques, sont d'une "décisive importance" (Roger Chartier). Avec la lecture silencieuse une nouvelle relation à l'écrit est instaurée, plus secrète, plus libre, tout intérieure. Le même texte peut dès lors être utilisé de diverses manières lu en silence pour soi, dans le retrait du cabinet ou de la bibliothèque, ou bien à haute voix pour d'autres, et dans un même livre (par exemple le livre d'heures) peuvent être différenciées les parties destinées à l'usage rituel, donc à la déclaration commune, et celles qui doivent nourrir une dévotion personnelle appuyée sur une lecture faite bouche close. (Roger Chartier : *Les Usages de l'imprimé*).

Pour l'ordinateur machine traitant des signaux postifs ou négatifs, constitués d'un flux d'électrons faire l'analyse d'une chaîne de caractères, c'est reconnaître que les mots qu'elle contient sont organisés selon la structure voulue pour le langage qu'on

Pour l'ordinateur machine traitant des signaux postifs ou négatifs, constitués d'un flux d'électrons faire l'analyse d'une chaîne de caractères, c'est reconnaître que les mots qu'elle contient sont organisés selon la structure voulue pour le langage qu'on étudie. Les lettres d'un texte sont codés en code normalisé (le plus souvent code ASCII). Ce code peut être encadré délimité par des balises, des signaux qui indiquent sa structure : c'est dans le domaine de l'imprimerie où il était nécessaire d'établir des équivalences entre le fichier texte issu du micro-ordinateur dédié au traitement de texte et le fichier d'entrée de la photocomposeuse . A partir de ce dispositif de conservation et de reproduction des textes on peut concevoir toutes sortes d'activités humaines : traitement de texte, édition à la demande de textes dont l'édition originale est épuisée, création de liens hypertexte, annotations, transfert par un réseau par le protocole FTP (File Transfert Protocol : transmission de documents numérisés à distance par le réseau informatique) , mise en rapport avec d'autres textes. S' il est vrai que "la télévision partage en effet avec d'autres inventions de cette fin de millénaire, comme l'avion, l'automobile, l'ordinateur..., la faculté de briser spectaculairement nos limites physiques, motrices, sensorielles ou intellectuelles" (Jean-Michel Salaün *A qui appartient la télévision ?*, Aubier 1989), l'hypertexte s'inscrit dans l'histoire complexe des rapports au texte, aux pratiques de la création intellectuelle et retrouve la trace d'autres ruptures spectaculaires passage du *volumen* au *codex*, du manuscrit à l'imprimé, du récit plein de commentaires, de digressions, multipliant les relatives et les répétitions , au récit articulé, linéaire, objectif , puis aux nouvelles pratiques littéraires et philosophiques du XX ème siècle ( roman de Marcel Proust, James Joyce, déconstruction et différance de Jacques Derrida, lexies de Roland Barthes, archéologie du savoir et épistémé de Michel Foucault).

Pour George P. Landow dans un chapitre intitulé *Hypertextual Derrida, Poststructuralist Nelson*, " quand les créateurs de logiciels informatiques examinent une page de *Glas* ou de *De la grammatologie*, ils se trouvent confrontés à un Derrida hypertextuel, digitalisé ; et quand les théoriciens de la critique littéraire examinent *Literary Machines*, ils rencontrent un Nelson déconstructionniste ou

convergé". ( George P. Landow, Brown University : " HYPERTEXT : the convergence of contemporary critical theory and technology", the John Hopkins University Press, 1992).

La notion d'étude critique de fragments de textes littéraires universels et dispersés remonte aux humanistes qui s'employaient à reconstituer les textes de l'Antiquité en partie perdus lors des incendies de la bibliothèque d'Alexandrie . Un pas décisif fut l'invention de l'imprimerie qui remplaçant les anciens *scriptorium* comme celui de Fleury ou de Marmoutier qui parvenaient à grand'peine à satisfaire les commandes royales et conventuelles, fournit aux Universités un relais aux manuscrits enchaînés aux lutrins, sous forme de *codex* que chacun devait reproduire pour soi et répéter au *Magister*. Elisabeth Eisenstein étudiant la révolution copernicienne (*The printing press as an agent of Change*, 1979 Cambridge University Press, Cambridge) remarque que les conséquences "radicales" qui suivirent le travail "modeste et non révolutionnaire" de Copernic "sembleraient moins étranges si les pouvoirs nouveaux de la presse à imprimerie étaient pris en considération". Avant l'imprimerie une version complète de l' *Almageste* de Ptolémée se trouvait rarement disponible dans une bibliothèque. Copernic cherche à rendre systématique le puzzle épars des textes adultérés de l' *Almageste*. En cinquante ans, entre les mains de Copernic, les textes de Ptolémée deviennent un système et, pour les mêmes raisons s'effondrent (cf l'article de Bruno Latour *Vues de l'esprit* dans le numéro 14 de la revue *Culture technique* ). La création littéraire et la recherche scientifique voient leurs conditions d'existence modifiées par l'apparition de nouvelles techniques d'inscriptions. Comme le remarque Jack Goody, l'écriture décontextualise le savoir : l'énoncé coupé de son énonciation peut voyager loin de sa source, faire l'objet d'un stockage, d'un traitement à part . Le champ cognitif de chacun peut s'étendre et rendre l'esprit plus indépendant. Les notions d'hypertexte et de réseau rendent les esprits plus ouverts et plus indépendants alors que les rapports de conditionnement qu'un nombre considérable d'individus entretiennent avec l'écran de télévision relève d'une névrose artificielle se surimposant et masquant d'autres facteurs de trouble. Cette aliénation relayée au Japon par la pratique

intensive des jeux électroniques et la lecture des bandes dessinées, a donné lieu à des descriptions proprement psychopathologiques (chez les jeunes : le "clan des mués"-*otaku zoku*- et chez les adultes : les "maris murés"-*otaku teishu* (Nippon Kenzai Shimbun traduit dans le numéro 35 du 4 juillet 1991 de *Courrier international*). Il est du devoir des institutions dédiées à la culture et au savoir de réintroduire la lecture, l'écriture, la tradition philosophique, la démarche de l'esprit scientifique, dans l'univers électronique. La lutte pour l'indépendance de l'esprit et pour la réflexion critique nécessaire à la survie de l'humanité est à ce prix. Dans ce combat l'hypertexte et les formes normalisées de balisage des textes électroniques sont des armes précieuses mises au point par l'ensemble de la communauté scientifique.

### 3.3. PLAO et études historiques :

On peut déjà remarquer comment le travail des historiens a été modifié par les nouvelles technologies. Dans les années 1970-80 un historien pour exploiter des informations chiffrées devait se rendre dans un Centre de Calcul et écrire en collaboration avec un informaticien un programme en langage FORTRAN, puis inscrire ses données sur des fiches perforées (cf. : Revue les *Annales, Économie, Civilisation, Société*, juillet-octobre 1972 *Un programme d'étude démographique sur ordinateur* par Y. Daubeze et J.-C. Perrot ). Aujourd'hui avec les micro-ordinateurs et les logiciels qui sont accessibles, son souci principal sera la mise en machine et l'exploitation des résultats. Le passage par un langage de programmation tâche lourde et demandant un investissement personnel en temps et en patience n'est plus indispensable. Il faut remarquer que les historiens ont mis au point des méthodes de calcul adaptées à leurs recherches par exemple le programme *Analyses Factorielles*. On pourra consulter, sur ce point, plusieurs études d'historiens publiées aux Presses Universitaires de Rennes (Histoire et Informatique), notamment l'article : *Édition structurée et indexation hypertextuelle d'actes médiévaux*, de Jacques André et Hélène Richy dans le cadre du projet Opéra de l'Irisa qui explique l'utilisation du logiciel GRIF prototype de l'INRIA (cf. : Hélène Richy, *Grif et les index électroniques* , Rapport de Recherche n°176, Inria, 1992 et du même auteur *A Hypertext electronic*

based on the Grif structured document editor, Électronique Publishing - Organisation, Dissemination and Design, 7, p.1-14, mars 1994), système de production de documents structurés, distinguant la description logique du document de sa structure physique (graphique); ses concepts sont proches de ceux de SGML (qui sont exposés au paragraphe 4.3). Pour GRIF, un document est d'abord une structure logique qui assemble des éléments tels que titres, chapitres, sections, paragraphes, notes, apparats critiques, etc. Ces éléments et leurs possibilités de composition sont définis dans des schémas propres à une classe de documents (par exemple les *Que sais-je ?*) à l'aide d'un formalisme proche de celui des grammaires de Chomsky. A chaque type d'élément défini dans un schéma de structure est associé un ensemble de règles de présentation qui définissent l'aspect graphique de ce type d'élément. La structure logique est principalement arborescente (hiérarchique). Mais on y a introduit la notion de liens pour donner aux documents une dimension hypertexte (cf : Vincent Quint et Irène Vatton, "Combining hypertext and structured documents in Grif", ECHT'92 : Proceedings of the ACM Conference on Hypertexts. D. Lucarella, J. Nanard, M. Nanard et P.Paolini, éd. p. 23-32, ACM-Press, USA, 1992).

La notion de structuration permet toute une série de calculs (numérotation des sections en bas de page et de l'apparat critique, etc.), de traitements linguistiques, de vérification de cohérence, de validité, etc. mais aussi de prévoir le travail coopératif de plusieurs personnes sur un même document faisant de GRIF un véritable outil professionnel.

La méthode traditionnelle (à la main avec fiches, gommes et crayons) de construction d'index a fait l'objet de diverses publications tant en français (par exemple le célèbre *Tapuscrit*) qu'en anglais (*Chicago manual of style*). GRIF a une philosophie différente les index sont de vrais liens hypertextuels, mais assujettis à la notion de type, ce qui permet des vérifications de cohérence, de validité.

En pratique, on distingue la mise en page des tables d'index, de leur contenu :

-les tables d'index sont des éléments du schéma de la structure du document, ce qui permet notamment d'indexer une table d'index ;

-les références peuvent représenter soit un renvoi (à la note, à une figure, à une section), soit une inclusion d'un élément du document ou d'un autre document (permettant ainsi d'indexer plusieurs volumes);

- les tables d'index sont tenues à jour en permanence;

- la présentation des index (emploi des caractères gras, numérotation vers les pages ou les sections) est guidée par l'utilisateur et non imposée par les systèmes.

Dans le cadre des Postes de Lecture Assistée par Ordinateur l'interêt serait de pouvoir présenter le texte étudié (ici le *Cartulaire de Saint-Laurent* découvert par Marc Bloch, parmi les papiers de l'abbaye de Saint-Magloire, étudié par Anne Ternoire et édité par Lucie Fossier), sous une forme numérisée comprenant :

- l'intégralité du manuscrit original

- une forme transcrite avec apparat critique

- des résumés

l'ensemble étant accompagné des outils de consultations utiles concernant l'annotation, l'indexation (GRIF comprend un système de commentaires électroniques que l'on peut rattacher à un endroit du document, comparable à l'annotation dynamique des Postes de Lecture Assistée par Ordinateur).

#### **4. L'insertion des Postes de Lecture Assistée par Ordinateur dans l'univers des réseaux internationaux**

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 dra-com) sur un poste connecté à Internet pour se retrouver en liaison avec le système d'interrogation propre à cette

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 électroniques 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 multimedia. (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 stocké 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*) , nommage 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

#### 4.1. Document électronique

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 *Glacis* 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 suspend 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). L'informatique, "physique des signes ininterprétés", selon l'expression de Maurice Nivat, rencontre ce nouvel âge du savoir. De là on peut observer l'Intelligence artificielle. Comme l'écrit Bruno Bachimont : "La question qui se pose alors est : tout problème peut-il être résolu informatiquement ? Tout problème peut-il être codé selon un alphabet de signes ininterprétés ? Autrement dit, tout problème peut-il être modélisé informatiquement ? (...) Alors que l'informatique résout des problèmes en les réduisant à une manipulation de signes ininterprétés, l'IA cherchera à manipuler informatiquement des signes porteurs de signification, pour utiliser les connaissances disponibles sur les problèmes. L'IA veut réintroduire une interprétation que la théorie de l'information avait chassée. La question précédente se comprend comme l'interrogation sur la capacité de l'informatique à aborder la question du sens et de la connaissance. Mais pour y répondre, il faut préciser la notion de connaissance, comme l'informatique l'a fait pour la notion d'information, et définir dans ce cadre ce qu'est une modélisation des connaissances. Ce que fit le cognitivisme computationnel" (Bruno Bachimont "Le contrôle dans les systèmes à base de connaissances : contribution à l'épistémologie de l'intelligence artificielle" Paris : Hermès, 1992). L'auteur définit une connaissance comme " un état mental sémantiquement interprété"(...) "La science étant incomplète, elle ne détermine pas les significations que nous trouvons pertinentes pour expliquer la situation : de mêmes situations physiques possèdent des significations différentes et des explications distinctes en termes de connaissance". Il s'agit de prendre en charge l'écart entre les connaissances non modélisables formellement et leur représentations symboliques, écart qui fonde l'interprétation. Le problème du contrôle comme choix, à un moment du processus de résolution, d'une représentation et de la manière de l'utiliser, s'avère décisif. Il s'agit d'une nouvelle conception de l'organisation de l'information, qui permet à la fois de repenser la bibliothèque, mais aussi qui doit

hériter du savoir-faire et des réflexions des bibliothécaires pour ne pas entraîner le lecteur vers une perte de sens par surabondance et désorganisation de l'information. Ce qui nous amène à nous replacer dans la perspective de l'histoire de l'hypertexte.

#### 4.1.1. Hypertexte l' aspect historique

"Selon notre façon de penser" : "As we may Think " est la titre de l'article de Vannevar Bush parut en 1945 dans la revue scientifique *Atlantic monthly* et exposant la nécessité d'organiser autrement le travail intellectuel, et l'accès aux connaissances. L'information doit être organisée selon notre façon de penser, et non pas l'inverse : notre pensée ne devrait pas être contrariée par le manque de mise en relation, de possibilité d'information. Vannevar Bush était "littéralement obsédé par le problème de la perte de l'information utile, décisive, "stratégique" dirions-nous aujourd'hui, prenant souvent l'exemple des découvertes de Mendel sur la génétique "perdue le temps d'une génération..." (cf : Alexandre Serres : "Hypertexte une histoire à revisiter " Documentaliste-Sciences de l'information, 1995, vol 32, n°2).

Parmi les précurseurs des postes de lecture assistée par ordinateur on peut également citer Paul Otlet et Henri Lafontaine connus pour leur projet de Répertoire bibliographique universel et leurs travaux sur la Classification Décimale Universelle. Ils sont mentionés comme fondateurs des sciences de l'information, partisans de la démocratisation des sciences et de la culture soutenue par l'utilisation de technologies nouvelles : microfiches, téléphone, téléscripneur, écran à tube cathodique de ce qui allait devenir la télévision ("la télévision, en tant que mode de transmission immatériel des informations est un substitut du livre", "la télescription, c'est-à-dire la possibilité a) d'écrire facilement à distance; b) d'ajouter à distance des inscriptions à des textes existants ..."), qui constituent l'amorce d'une messagerie intelligente et d'un réseau hypertextuel. (cf : Jean-Michel Salaün, *Les sciences de l'information en question : le point de vue du lecteur*, Réseaux n°58 CNET-1993, et Alexandre Serres, opus cité, l'ouvrage de Paul Otlet *Traité de documentation : le livre sur*

le livre, paru en 1934 a été réédité en 1989 par le Centre de lecture publique de la communauté française de Belgique à Liège).

#### 4.1.2. Les principaux points de repères historiques sont :

1945 La publication par Vannevar Bush de l'article intitulé *As we may think* dans le numéro de juillet de la revue *Atlantic monthly*. Bush remarquait que les progrès de la recherche étaient freinés par les difficultés d'accès aux documents, et proposait une machine appelée *memex* (memory extender) utilisant des micro-fiches portant des numéros de code, accessibles à tout moment dans un répertoire de codes, et des bandes magnétiques. Les documents apparaissaient sur un petit écran translucide, leur défilement était accéléré ou ralenti au moyen de leviers. Des liens pouvaient être créés entre les documents, l'utilisateur pouvait annoter son parcours de lecture et conserver la trace de son cheminement. Il s'agit "d'un système de codage par points rattachés aux mots ou aux documents que l'on souhaite lier".

1963 - Douglas Engelbart publie un article *Augmenting Human Intellect : A Conceptual Framework* (Un cadre conceptuel pour l'accroissement de l'intellect humain). *Vistas in Information Handling*. Ed. P.D. Howerton and D.C. Weeks Spartan Books, Washington D.C. 1: 1-29. Le projet ARGUMENT le conduit à proposer NLS on Line System, système en ligne, qui transforme la machine à calcul qu'était l'ordinateur en outil universel de manipulation cognitive. Il est également l'inventeur de la souris, périphérique de pointage qui modifie la conception classique en informatique des périphériques d'entrée et de sortie, du premier système de courrier électronique et de téléconférence en réseau. AUGMENT fut développé au sein de la firme McDonnell Douglas puis à l'Université de Stanford en Californie où ont été expérimentés le multifenêtrage d'écran, les liens associatifs entre les données, les graphes dynamiques pour représenter les idées (processeur d'idées), l'aide en ligne intégrée au logiciel. Les chercheurs du Xerox parc de la firme Rank Xerox de Palo Alto (projet Dynabook d'Alan Kay auteur du langage orienté objet Smalltalk) se sont inspirés

de ses idées qui furent concrétisées par les micro-ordinateurs Liza puis Macintosh d'Apple.

1965 - Theodor Home Nelson crée le terme hypertexte .

L'idée le poursuivait depuis quelques années. Il s'en explique ainsi : "Elle m'est venue en octobre-novembre 1960 alors que je suivais un cours d'initiation à l'informatique qui au début devait m'aider à écrire mes livres de philosophie. Je cherchais un moyen de créer sans contrainte un document à partir d'un vaste ensemble d'idées de tous types, non structurées, non séquentielles, exprimées sur des supports aussi divers qu'un film une bande magnétique ou un morceau de papier . Par exemple, je voulais pouvoir écrire un paragraphe présentant des portes derrière chacune desquelles un lecteur puisse découvrir encore beaucoup d'informations qui n'apparaissent pas immédiatement à la lecture de ce paragraphe".( cf. A. Baritault, *Xanadu, Sciences et Vie Micro*, novembre 1990). Le mot "Xanadu" fait allusion à un célèbre fragment poétique de S.T. Coleridge, où s'exprime un exotisme visionnaire romantique, dont la reprise par Nelson ne va pas sans humour. Orson Welles dans "Citizen Kane "baptise le domaine crée par Kane, mélange de chateau hanté, de jungle tropicale et de musée europeen, Xanadu. Depuis trente ans Ted Nelson ( suivant l'appellation familière des auteurs américains) a défendu sa conception de l'hypertexte. En 1988 la société Autodesk , Inc. a investi dans Xanadu et formé the Xanadu Operating corporation.

1968 - Un groupe de chercheurs de Brown University dirigé par Andries Van Dam (Van Dam, Andries (1988). Hypertext '87 keynote address (transcript) Communication of the ACM : July v31 n7 887-896), réalise le premier système véritable d'hypertexte Hypertext Editing System puis File Retrieval and Editing System (1982). Actuellement Brown University développe Intermedia, utilisé pour enseigner la biologie et la littérature. Ces nouvelles méthodes pédagogiques ont été décrites par George P. Landow professeur de littérature a Brown University dans plusieurs ouvrages importants : *Hypertext, the*

*Convergence of Contemporary Critical Theory and Technology*, The John Hopkins University Press, 1992, et *The Digital Word Text-based computing in the humanities*, 1993, MIT-press, Cambridge Massachusetts.

1983-1986 Les années quatre-vingts sont marquées par le succès commercial foudroyant de la micro-informatique : IBM-PC et Apple Macintosh. En 1982, Peter Brown invente le premier système commercialisé de création d'hypertexte sur micro-ordinateur, qui reçoit le nom de *Guide* en 1985, quand le premier prototype *Office Workstations Limited* (OWL) est adapté au Macintosh d'Apple puis à l'IBM-PC en 1987. Il rendait accessible au public les concepts d'interface graphique, la création de noeuds et de liens.

En 1986 Xerox Parc réalise NoteCard, qui comme Intermedia de Brown University supporte des noeuds et des liens avec des éléments graphiques et des séquences d'animation, ainsi qu'un survol graphique de la structure de l'hyperdocument pour aider la navigation dans l'hypertexte. Chaque noeud est représenté à l'écran par une carte. Les cartes sont classées par type. Certains types de cartes sont prédéfinis et d'autres peuvent être créés par l'utilisateur.

1987 - "HyperCard n'est pas du tout de l'hypertexte" écrivit Ted Nelson quand ce logiciel fut édité (Nelson Theodor H. (1987) *Literary Machines*. Vers. 87.1 The Distributors, South Bend, IN). HyperCard ne permet pas les liens entre les textes reliés par des noeuds. Toutefois comme il était livré avec chaque micro-ordinateur, il a sans doute permis de populariser certains concepts comme la navigation entre des données reliées (métaphore de la pile et des boutons reliant les cartes). De nombreuses piles ont été produites par des auteurs avec Hypercard. Toutefois la tendance actuelle de l'informatique est d'utiliser les codes standards (ASCII), et les systèmes normalisés (ISO), pour éviter d'être liée à la production d'une entreprise particulière qui n'est jamais à l'abri des manoeuvres financières.

Également en 1987 fut commercialisée la version IBM-PC d'*Hyperties*, qui était apparue en 1983 sous le nom d' *Interactive Encyclopedia System* (TIES) à

l'Université du Maryland (Human-Computer Interaction Laboratory), et se présente suivant la métaphore du livre électronique ou de l'encyclodédie.

Les années 1990 sont marquées par le développement de l'hypertexte en réseau au niveau mondial via Internet et les adresses de type http et WWW et par l'importance des programmes de numérisation de textes en dehors de celui de la Bibliothèque Nationale de France. En 1991, l'Université de Georgetown (Center of Text and Technology Academic Computer Center), publiait un premier inventaire de 329 organismes qui développent des programmes de bases textuelles informatiques.

#### **4.2. Implémentation des Postes de Lecture Assistée par ordinateur :**

Le projet de Poste de Lecture Assistée par Ordinateur s'est concrétisé, sous la direction d'Alain Giffard directeur du département des nouvelles technologies par des études portant sur des stations de lecture expérimentales à partir de 1990. Dans un premier temps différentes machines ont été expérimentées d'abord :

un Macintosh II Ci avec 8 Mo de RAM, 80 Mo de disque dur, un écran de PAO affichant deux pages simultanément, un scanner Apple, une unité de sauvegarde du type boîte de Bernoulli de 44 Mo, une imprimante laser, un modem et un lecteur de CD-ROM. Les logiciels utilisés étaient : Word 4.0, MacDraw, Phrasea, Mark-Up, Marco-Polo, Hypercard et Candide. Un corpus de 20 000 pages digitalisées en mode alphanumérique et bit-map permettait de donner à l'expérimentation un aspect concret.

De juin à mai 1992 de nouvelles machines furent expérimentées : deux modèles ont été étudiés concurremment , l'un sur NeXT avec un écran de 22 pouces, et un autre sur Sun Sparcstation 2, moniteur de 19 pouces, une unité centrale dotée de 32 Mo de RAM, et 660 Mo pour la capacité du disque dur. Trois cents Postes de Lecture Assistée par Ordinateur sont prévus sur le site même de la Bibliothèque Nationale de France et les pôles associés pourront choisir leur matériel en fonction des facilités offertes par l'architecture client/serveur de type Internet et l'utilisation des formats normalisés d'échange.

Quatre logiciels ont été expérimentés à domicile par des chercheurs de différentes spécialités (germanistes, anglicistes, archéologues, historiens, écrivains, mathématiciens, philosophes), disposant de corpus de textes numérisés dans leurs disciplines. (cf : Dominique Chouchan, La Recherche 228 janvier 1991 volume 22, *The Digital word, Text-based computing in the humanities* éd. par George P. Landow et Paul Delany, MIT-Press, 1993, ainsi que AFCET *International Conference on Document and Analysis Recognition*. Berlin: Springer-Verlag 1991). 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 comme composants de cette base. Ces concepts constituent le contexte d'apparition des langages d'annotation, de la lecture et de l'écriture éditoriales assistés par ordinateur, et plus généralement, de l'hypertraitement de texte. Les premiers logiciels testés furent :

-*Phrasea* qui a pour fonction de générer des bases de données personnelles à partir de fichiers en texte intégral selon les caractéristiques définies par l'utilisateur (structure de l'unité documentaire, famille de synonymes etc). Il permet la constitution d'un dictionnaire et d'un index de mots du texte et une interrogation par opérations de type booléen : présence, absence, co-occurrences de mots ou d'expressions que l'on souhaite inventorier avec le contexte de chaque occurrence (par exemple les vingt lignes précédentes et suivantes).

-*Mark up* destiné à tout le travail d'annotation, de commentaires, de marques diverses dont le lecteur professionnel est coutumier. Ces traces de lecture ainsi que les résultats d'analyse par *Phrasea* ou d'autres documents saisis par scanner peuvent être ensuite classés par l'utilisateur avec le logiciel *Marco Polo*. 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. Avec le

logiciel *Candide* qui utilise la méthode Leximappe, méthode des mots associés ; l'analyse des mots associés est une méthode statistique développée par le Centre de sociologie de l'innovation (CSI) de l'École des mines de Paris et le Centre de documentation scientifique et technique (CDST) du CNRS ; 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. (cf : Histoire et informatique : bases de données, recherche documentaire, multimédia, textes réunis par Martine Cocaud, Presses Universitaires de Rennes, 1995 article : Une application de Candide, logiciel d'analyse textuelle pour une histoire de la la traduction littéraire M.-L. Malingre URFIST de Rennes- Université Rennes 2).

Le logiciel *Candide* a été intégré à la maquette LECAO (Lecture et Ecriture Critique Assistées par Ordinateur) maquette simulant l'interface d'un poste de lecture assistée par ordinateur. Le système définitivement retenu devant être expérimenté entre 1993 et 1995.

#### 4.3. Postes de Lecture Assistée par Ordinateur et édition électronique

A ces recherches sur le mode de transmission des textes s'ajoute une question médiologique : le problème de la presse scientifique en crise depuis plusieurs années du fait de l'augmentation des coûts des abonnements, de la multiplication des titres, de la diminution du nombre des abonnés, des délais importants de parution pour les articles dans les revues à comité de lecture (de 8 à 36 mois ). Ces délais sont "intolérables pour des chercheurs dont la vie scientifique rythmée par les nouvelles technologies est confrontée à une culture de l'urgence" (Communication de Françoise Renzetti à la journée d'étude organisée par l'Association des conservateurs de bibliothèque, le 6 avril 1995 :

"l'arrivée d'Internet dans les bibliothèques et centres de documentation : conséquences et pratiques"). Dans le domaine des mathématiques, auquel il existe une vingtaine de revues électroniques, un grand nombre étant en préparation. Des études prévoient que d'ici la fin de la décennie, la totalité des revues scientifiques seront électroniques". Le système d'évaluation des articles risque dévoluer vers un fonctionnement collectif et médiatisé "Des procédures permettraient de transformer le travail des comités de lecture en téléconférences dont les transactions respecteraient l'anonymat et au cours desquelles chaque message serait évalué ". Ce système fut mis en place par les chercheurs eux-mêmes : dans un article de la revue *Science*, B.R. Schatz et J.B. Hardin de l'Université de l'Illinois rappellent que WWW est né d'une attente des chercheurs du CERN, le Laboratoire Européen de Physique des particules à Genève . Un groupe de chercheurs conduit par Tim Berners-Lee du CERN proposa un système d' hypertexte en réseau destiné à être utilisé par les chercheurs de l'institut. Un premier document exposant ce projet est daté d'octobre 1990 (T. Berners-Lee, Internal CERN document proposing WWW (october 1990)). Les documents de référence, dont ceux définissant la norme HTML 2.0, sont disponibles sur le site du CERN, où le concept du World-Wide Web a été crée (pour accéder à ces documents l'adresse est <http://www.cern.ch>). Web combine deux idées qui seules sont utiles et combinées ont prouvé leur extraordinaire puissance : l'information en réseau et l'hypertexte. Le serveur WAIS permet d'interroger plus de deux cents bases bibliographiques. Il semble évident qu'un certain nombre de concepts élaborés dans le monde de la recherche vont s'appliquer à l'édition et aux bibliothèques :

a) Les normes d'échange de documents et les périodiques électroniques

La gestion électronique de documents entraîne un changement dans la diffusion de l'écrit. Il ne suffit plus que celle-ci soit une simple diffusion, à sens unique, d'informations lisibles. Les utilisateurs veulent désormais, non seulement lire le document, mais encore réutiliser l'information qu'il contient,

c'est-à-dire reprendre des parties dans un nouveau document, ou encore stocker des références.

La possibilité de réutiliser les documents ajoute une dimension à l'infrastructure nécessaire à leur diffusion. Un document réutilisable doit être plus qu'un simple ensemble de textes et d'images, présenté pour qu'on puisse le lire; il doit inclure des informations supplémentaires sur la structure sur la présentation, peut-être sur l'objet ainsi que d'autres caractéristiques du document.

b) Les normes sur le traitement et l'échange des documents :

Les normes sur les documents visent à définir des formats pour promouvoir l'échange et le traitement du document, quant à :

- son contenu : la partie du document qui se lit,
- sa structure logique : la construction logique du document,
- sa mise en page et sa feuille de style : la présentation du document.

Au fur et à mesure que notre civilisation croît en complexité, les besoins relatifs à une communication souple se font plus pressants. Les normes du document électronique répondent à la nécessité d'échanger des documents sous forme révisable.

Beaucoup de documents sont utiles autant parce que nous pouvons en réutiliser l'information, que par l'information que nous retirons de leur lecture. Par exemple un concepteur produira des documents pendant le processus de conception, processus qui sera probablement porteur d'informations utiles pour le manuel d'utilisation, la maintenance, la formation et l'activité commerciale. Pour réutiliser ces informations dans d'autres documents, il doit être possible d'introduire les documents dans d'autres systèmes de traitement de texte, d'où le besoin de disposer d'un format commun d'échange de documents.

Avec les documents que nous produisons nous-mêmes, il est généralement possible de réutiliser l'information. Pour les documents que nous recevons d'autres sources, c'est généralement très difficile. Par exemple, la plupart des documents que nous recevons sont imprimés, et les avoir sur un système de

traitement de texte signifie qu'on doit les resaisir au clavier, ou les digitaliser au moyen d'un numériseur (scanner transformant les signaux optiques en code numérique). Même s'il s'agit de documents reçus sous forme lisible par la machine, l'information n'est souvent accessible qu'en fichier "à plat" (c'est-à-dire que la seule information disponible est le contenu textuel). Pour être vraiment exploitable, les informations physiques et structurelles doivent accompagner le contenu brut d'un document.

c) L'échange des informations physiques et le formatage

Formater et mettre en page un document consomment du temps. Comme peu de systèmes ont des fonctionnalités permettant d'envoyer une instruction de formatage, ce travail est souvent perdu quand on transfère un document. Le destinataire du document doit alors le reformater s'il souhaite le réutiliser. Les instructions de formatage utilisées dans les différents systèmes sont en fait rarement compatibles. S'il était possible d'envoyer l'information physique (mise en page) avec le document, cette procédure de reformatage pourrait être évitée. Le modèle de document SGML (Standard Generalized Markup Language) permet l'échange de documents électroniques structurés. C'est une norme pour la création de langages de balisage d'applications et la création de documents échangeables et révisables. C'est un métalangage, un langage pour définir d'autres langages, à savoir des langages de balisage.

SGML est relié au multimédia par le langage HyTime : Hypermedia /Time based Structuring Language c'est-à-dire Langage de Structuration Temporel/Hypermedia. HyTime définit un langage et un modèle associé pour une représentation des hyperdocuments qui établissent des liaisons et synchronisent l'information statique et dynamique contenue dans un document multimédia. Elle permet de représenter le temps telle une écriture musicale-une partition-et d'utiliser les unités de temps réel, définies par l'utilisateur.

#### 4.3.1. La sémantique de présentation de documents et le langage de spécification

##### a) Le parser comme interprétant

DSSSL (Document Style Semantics and Specification Language) ou sémantique de présentation de documents et langage de spécification vise à définir et à déterminer les commandes pour la mise en page des documents révisables. DSSSL permet la définition du style ou de l'apparence d'un document composé, indépendamment de tout processus de mise en page ou de tout type de programme de mise en page. La spécification des attributs doit être indépendante du matériel de mise en page. Ce modèle conceptuel est centré sur les spécifications de composition. Il consiste à développer un ensemble normalisé de concepts fonctionnels pour décrire la mise en page et les feuilles de style pour toute application relative à l'impression.

##### b) L'indexation en texte intégral :

L'indexation en texte intégral est certainement la méthode d'indexation automatique la plus répandue . Elle s'applique dans de nombreuses banques de données diffusées dans les systèmes commerciaux. Les dépêches des agences de presse, les versions électroniques des journaux et des revues, les lettres d'information, les textes de loi et la jurisprudence .

L' indexation en texte intégral considère que tous les mots non grammaticaux d'un texte sont les meilleurs descripteurs de ce texte. Les descripteurs du document sont alors tous les mots lexicaux (noms, verbes, adjectifs ) présents dans le texte. Pour indexer en texte intégral, le système repère d'abord chaque mot du texte. Après élimination des signes de ponctuation, un mot est la chaîne de caractère comprise entre deux espaces . Un anti-dictionnaire permet d'éliminer les termes ayant peu de valeur documentaire, notamment les mots grammaticaux.

La recherche en texte intégral prend une importance grandissante avec l'apparition de support permettant le stockage et l'exploitation de grande quantité de texte. Son essor est favorisé par l'utilisation accrue de la PAO dans

les entreprises. Celle-ci génère une documentation précieuse mais non structurée et difficile à gérer à l'aide des bases de données classiques.

La recherche en texte intégral ne doit pas être confondue avec l'indexation automatique, ou opposée de manière abrupte à l'indexation sélective des textes.

En effet si la recherche et la définition d'une fonction de correspondance entre une requête d'un utilisateur et un corpus de documents constitue un sujet de recherche connu depuis plusieurs années sous le nom de recherche d'informations ou recherche documentaire en français, information retrieval ou text retrieval en anglais, les notions de balisage de texte, de liens hypertextes, et de lecture en texte intégral sont plus récentes (normes SGML, langage HyTime et HTML). Les fonctions des systèmes de recherche d'informations sont la gestion et l'accès à des bases documentaires de manière à permettre à l'utilisateur d'obtenir des réponses à des questions.

Un poste de lecture assistée par ordinateur doit permettre la fonction de recherche documentaire, mais aussi le téléchargement dans la station de travail des ouvrages du fonds, indentifiés comme pertinents pour le travail du chercheur et d'augmenter ce corpus par des fonds numérisés personnels déjà possédés par le chercheur, permettre des opérations de lecture, annotation, commentaire, constitution de bases de données personnelles, de comparaison, de structuration, d'analyse, de traitement et de publication de textes.

Il y a donc deux types de fonctions, deux ensembles distincts de méthodes de manipulation de documents permettant tout d'abord de les représenter, de les stocker et de les organiser. La première méthode porte sur les attributs externes des documents, comme leur titre, leur auteur, date d'édition, langue, pays d'édition, format, l'autre porte sur le contenu des documents. La réponse à une requête se compose d'un ensemble de références vers les documents sélectionnés, ou bien du contenu même des documents, voire de certaines parties du contenu.

La fonction de correspondance requête / document implique l'utilisation des attributs externes et des attributs de contenu, séparément ou

simultanément. En ce qui concerne les attributs externes, la recherche des documents pertinents est déterministe, et la correspondance requête /document exacte. En ce qui concerne les attributs de contenu, la correspondance ne peut être exacte dans le cas général. Il faut en effet, à ce niveau, établir une comparaison sémantique (et non une égalité) entre des concepts figurant dans un document et d'autres énoncés dans une requête. Une telle comparaison aboutit rarement à des équivalences strictes, du fait des nuances de langage et d'expression des connaissances. On se heurte ici à la question soulevée par le nominalisme : le concept n'est qu'un nom et n'existent effectivement que les individus auxquels renvoient les noms. Dans cette perspective, toute la réalité, du moins pour l'esprit de l'homme, est donnée dans des expériences singulières. Or le langage introduit, comme moyen de connaissance des entités conceptuelles, qui ne peuvent être alors que des représentations de formes, et des instruments de classification pour les entités réelles. Et la question est : comment ces formes symboliques renvoient-elles cependant au réel. Cette formulation, sans doute trop simpliste, a cependant le mérite de montrer la parenté du nominalisme et de l'empirisme, la tâche du premier se présentant comme inverse de celle du second. Des nominalistes médiévaux (Abélard, Guillaume d'Ockham), aux empiristes logiques néo-positivistes et aux travaux de Bertrand Russell et Wittgenstein, l'interrogation sur la manière dont la connaissance abstraite se développe par la médiation du langage est un sujet de réflexion essentiel. Aussi les recherches d'informations en texte intégral ne peuvent elles être à l'image des systèmes de gestion de bases de données, fondées sur des stratégies de correspondance exacte requête/base de données, parce qu'elles mettent en oeuvre deux types de connaissances : l'intuition qui est saisie immédiate de substances et d'accidents, entités singulières données comme existantes et contingentes, et la connaissance abstraite. Celle-ci est un *habitus* proprement humain qui permet de concevoir un objet non présent, sans autoriser pourtant, à en affirmer une propriété contingente, qui serait du ressort de la connaissance intuitive, mais

exigerait l'existence présente de l'objet. En parcourant un texte de politique française nous savons intuitivement ou par habitus que l'expression "l'hôtel Matignon" est équivalente à celle de "premier ministre" ou de "chef du gouvernement", mais un système de recherche fondé sur une correspondance exacte requête/document retournerait des réponses vides. L'intérêt du livre électronique est d'être une base de données spécifique munie de procédures de navigation et d'appropriation par le lecteur utilisateur qui donne lui même sens à sa lecture. Alors que l'indexation en texte intégral fait porter sur l'utilisateur le fardeau de la description documentaire en l'obligeant s'il veut obtenir un bon taux de couverture pour sa recherche à imaginer toutes les formes possibles que l'expression qu'il recherche pourrait prendre dans le texte, l'hypertexte reste dans le cadre de la lecture, en apportant des possibilités d'interrogations, et la lecture assistée par ordinateur permet de garder des traces du cheminement de la pensée par des annotations, des sauvegardes des stratégies de recherche, des surlignages. La distance sémantique entre la requête (pointer sur un mot interrogeable du texte) et la réponse est évaluée immédiatement par l'utilisateur qui peut alors corriger sa recherche en fonction des autres unitermes présents dans le texte . Le lecteur doit pouvoir utiliser le résultat d'une première recherche documentaire pour pointer des mots apparaissant à l'écran . Cette opération relance la recherche en utilisant un terme repéré et jugé pertinent (concept subjectif), par l'utilisateur. Il est souhaitable d'apporter à l'utilisateur des instruments pour juger de la qualité du système de recherche d'information c'est à dire de sa capacité à fournir des réponses ayant une bonne mesure de pertinence. Les critères classiques , qui permettent de mesurer le résultat global de la requête (ensemble de réponses obtenues) et de les comparer, font intervenir cette notion de pertinence (adéquation d'un document à la requête) au travers de deux paramètres qui sont le rappel et la précision :

- les mesures de *rappel* donnent la proportion d'informations pertinentes retrouvées par rapport au nombre total de réponses pertinentes;

- la *précision* mesure la proportion d'informations pertinentes retrouvées par rapport au nombre total de réponses données.

On peut également mesurer :

- le *silence* le complémentaire du rappel, c'est-à-dire le nombre de documents pertinents non retrouvés par rapport à tous les documents pertinents ;

- le *bruit* le complémentaire de la précision, c'est-à-dire la proportion de documents non pertinents donnés en réponse par rapport au total de documents donnés en réponse.

Dans les systèmes utilisant les jugements de pertinence, un ensemble de caractéristiques calculé sur plusieurs documents, est utilisé pour relancer la recherche. Ici le lecteur reconnaît et sélectionne les termes au cours de sa lecture, ce qui lui évite d'essayer de deviner ce qui est contenu dans la base de données, mais seuls les termes pointés sont pris en compte . La recherche devient alors beaucoup plus sensible aux limites de l'indexation, et la dispersion due aux ambiguïtés des termes employés est plus forte. Il faut que le poste de lecture permette à la fois la recherche documentaire et la lecture active, avec la possibilité pour un utilisateur d'annoter les références qu'il sélectionne (pourquoi est elle conservée, rangement dans un dossier particulier) et la conservation d'un historique du chemin parcouru, et plus encore d'un historique des noeuds considérés comme importants pour son travail par le lecteur-auteur.

Cette conception est rendue possible par l'apparition de postes de travail personnels dotés d'un écran de visualisation graphique constitué d'une grille de points et d'un dispositif de désignation d'un point de cet écran appelé souris. Cette interaction avec l'utilisateur revient sur une distinction fondamentale de l'informatique classique : la séparation entre les dispositifs d'entrée des données et les dispositifs de sortie. Dans le cas d'un bouton dessiné à l'écran sur lequel l'utilisateur va pointer avec sa souris pour déclencher une action, conceptuellement le dispositif de sortie (écran) devient un dispositif d'entrée :

d'où un mode particulier d'accès à l'information : la navigation. On peut distinguer plusieurs types de recherches d'information liés à ce mode d'accès particulier :

-la *lecture séquentielle*, qui correspond au texte écrit, à l'image animée, au son, et qui s'écarte de la *lecture linéaire* étudiée par Jacques Derrida (*De la grammatologie*, La dissémination, concept de "différance").

-la *formulation de requêtes*, utilisée dans les banques de données, liée à la notion d'indexation

-le *browsing* traduit par "*butinage*", qui fonctionne par association d'idées ou par approfondissement autour d'un point focal. Ce type de recherche est comparable à l'utilisation d'un dictionnaire ou d'une encyclopédie : à partir d'un point d'entrée suivre les divers corrélats. C'est également la méthode utilisée quand on part de l'index d'un ouvrage pour rechercher les pages concernant un sujet : sous cet aspect il s'agit de pratiques du travail intellectuel qui se rattachent aux recherches Pierre de La Ramée (1515-1572), aux variations dans le découpage des textes bibliques, en particulier des éditions de Robert Estienne et leurs versets numérotés, ou aux dispositifs propres au livre imprimé (titre et page de titre, découpage en chapitres, regroupement des notes et glossaires), c'est également la méthode utilisée par les clients d'une librairie ou par les usagers d'une bibliothèque en libre accès . (Dans tout ces exemples le *browsing* est tributaire du travail de préparation effectué : qualité de l'index des noms cités, sérieux de la pagination, méthode adoptée par le libraire pour présenter ses ouvrages, cotation et rangement des ouvrages en accès libre ). D'où le danger "d'une violence faite au texte" qui a été évoquée par Roger Chartier ("du codex à l'écran", in *Pour une nouvelle économie du savoir*, Groupe Interuniversitaire de Recherche en Science de l'Information et de la communication, dossier n°1, Presses Universitaires de Rennes, 1994). Le travail d'annotation et de renvoi nécessite des compétences scientifiques et littéraires qui font encore parfois défaut, et la mise en place d'une nouvelle

grammaire qui pourrait s'appuyer sur les travaux sur le signe, de Charles S. Pierce, Roland Barthes ou Jacques Derrida.

Ces modèles d'information ne sont pas opposés : l'hypertexte doit intégrer ces trois modes de recherche à un système de lecture et d'écriture personnelle modifiant la conception classique en informatique d'entrée et de sortie des données, pour s'insérer dans un Poste de Lecture Assistée par Ordinateur.

Dans un réseau hypertexte, l'information est décomposée en blocs élémentaires (les *noeuds*), qui sont reliés entre eux par des liens qui autorisent le passage direct d'un noeud à l'autre. Cette notion de lien est l'essence même des systèmes hypertextes. Elle existait déjà dans d'autres formes d'organisation des informations. Par exemple, il existe des liens explicites dans les textes scientifiques (les citations), dans les textes juridiques (renvoi à des textes de loi, de décret ou de jurisprudence), dans les encyclopédies (renvois à d'autres articles) et des liens implicites dans certains types de documents qui renvoient à des illustrations à partir du texte (livre d'art, programmes de télévision...). Pour être qualifié d'hypertexte, le système doit fonctionner avec des liens à action immédiate. Invoquer un lien provoque un déplacement dans un champ du savoir : ce type de mise en relation peut prendre l'aspect d'un labyrinthe angoissant ou rejoindre "le bonheur extravagant" de Borges. Il suffit parfois d'une erreur minime dans la conception d'une interface, pour que l'utilisateur se trouve aux prises avec une contradiction logique insurmontable et perde les fruits de son travail en étant contraint de réinitialiser sa machine ( cf. la première version du langage HyperTalk qui entraînait parfois l'apparition d'une bombe sur l'écran et la réinitialisation).

L'hypertexte se rattache à la capacité de l'être humain de traduire pour tous les sens, l'expérience d'un seul et d'en faire pour l'esprit une image unifiée; on peut regrouper la littérature sur un domaine, circuler entre les textes par des liens associatifs, qui autoriseraient des annotations par les différents "lecteurs" (lecture, écriture, communication orale ou écrite), mettre

en perspective un document dans le contexte de tous les autres et opérer le dévoilement des rapports explicites (la citation), ou implicite (analogie) avec un certain nombre d'entre eux. Il est désormais possible de programmer des états de la machine qui prolongent les besoins de notre conscience autour des thèmes de la lecture active (annotations, liens analogiques), du travail collectif autour d'un document ou d'un ensemble de documents (cf : l'article de Madame Ghislaine Chartron "Nouvelles technologies et organisation de travail coopératifs : quelques repères" *Pour une nouvelle économie du savoir* (opus cité)), du cheminement personnel dont on peut éventuellement garder la trace, du découpage du texte en éléments reliés.

##### 5. Postes de Lecture Assistée par Ordinateur et culture post-moderne

En fait comme le rappelle Alexandre Serres ("Hypertexte une histoire à revisiter", *Documentaliste-Sciences de l'information*, 1995, vol. 32 n°2), l'essor considérable de l'hypertexte depuis une dizaine d'années consacre l'accomplissement d'un projet plus que cinquantenaire : ce n'est pas en 1945, date de la publication du texte fondateur de Vannevar Bush : "As we may think" (*Atlantic Monthly*, juillet 1945), qu'est apparue cette notion, mais dès le début des années trente (travaux d'Emmanuel Goldberg sur le microfilm et son utilisation, de Paul Otlet sur la documentation. Des siècles d'hégémonie de l'écriture linéaire, nous ont dissimulé d'autres modes de présentation du savoir, d'autres agencements cognitifs, d'autres modèles de pensée. Jacques Derrida ("De la grammatologie", Paris Editions de Minuit, 1967) reprenant l'inspiration d'André Leroi-Gourhan ("Le geste et la parole. Tome 2 : la mémoire et les rythmes", Paris, 1965), évoque "la guerre qui s'est installée, et un refoulement de tout ce qui résistait à la linéarisation" : "l'écriture au sens

étroit - et surtout l'écriture phonétique- sont enracinés dans un passé d'écriture non linéaire. Il a fallu le vaincre". Et Derrida en appelle au concept de "linéarisation" pour relire toute l'histoire des écritures, des pictogrammes aux écritures alphabétiques et linéaires. Roland Barthes dans *Éléments de sociologie*, commentant le *Cours de linguistique générale* où Saussure insiste sur la langue comme le domaine des articulations, le sens étant un ordre, mais cet ordre étant essentiellement division, pensée du discontinu, du combinatoire. En dépit des progrès apportés par la technologie de l'imprimerie celle-ci figeait les virtualités du texte, de manière linéaire, bien que toute lecture soit une lecture comparative, une mise en rapport du livre avec d'autres livres. Lire, ce serait faire émerger la *bibliothèque vécue c'est-à-dire la mémoire des lectures antérieures et des données culturelles*. Le genre du livre, le lieu d'édition, les critiques, le savoir scolaire nous placent en position d'écoute, en état de réception. La culture est ce qui nous prédispose à une réception particulière des textes. L'École de Constance a créé le concept *d'horizon d'attente* de Jauss : chaque époque construit ses modèles et ses codes narratifs. La première théorie de la narrativité naît avec Aristote et concerne la tragédie. Depuis lors toutes les théories du texte narratif se sont référées à ce modèle. Aristote parle de l'imitation d'une action, c'est-à-dire d'une séquence d'événements, réalisée en construisant une *fabula*, c'est-à-dire une histoire, une séquence d'actions. Umberto Eco définit ainsi la construction d'un récit linéaire de ce type : " la recette aristotélicienne est simple : prenez un personnage auquel le lecteur puisse s'identifier, ni franchement mauvais ni trop parfait, et faites lui vivre des aventures qui l'amènent à passer du bonheur au malheur et vice-versa, à travers maintes péripéties et reconnaissances. Tendez au maximum l'arc narratif, afin que lecteurs et spectateurs éprouvent pitié et terreur à la fois. Quand la tension aura atteint son maximum, faites entrer en action un élément qui vienne démêler le noeud inextricable des faits et des passions en résultant. Que ce soit un prodige, une intervention divine, une révélation ou un châtement inopiné, il faudra que cela donne lieu à une catharsis - dont on ne sait

clairement si, chez Aristote, il s'agit d'une purification de l'auditoire, soulagé du poids que la série insoutenable des événements avait imposé, ou une purification de l'histoire elle-même, laquelle trouve enfin une solution acceptable, cohérente avec l'idée que nous avons de l'ordre logique (ou fatal) des événements humains. En élaborant cette recette, Aristote savait pertinemment que le paramètre de l'acceptabilité ou de l'inacceptabilité d'une histoire ne réside pas dans l'histoire elle-même, mais dans le système d'opinions régissant la vie sociale. Pour être acceptable, l'histoire doit donc paraître vraisemblable, le vraisemblable n'étant autre que l'adhésion à un système d'expectatives habituellement partagé par l'auditoire. Quant à la pitié et la terreur, curieusement, il les définit non dans la *Poétique*, qui traite des structures narratives, mais dans la *Rhétorique*, qui traite de l'opinion publique et de la façon de l'utiliser pour susciter des effets de consensus. (De Superman au surhomme, Grasset, 1993). Chaque époque construit ses modèles et ses codes narratifs et à l'intérieur de chaque moment il existe des codes divers selon les groupes. Cervantès se moque des romans de chevalerie qui sont à son époque un héritage du passé, et du public qui en accepte les effets de rédibilité et les codes narratifs. Au XX<sup>ème</sup> siècle Marcel Proust, James Joyce, Bertold Brecht, modifient la manière de structurer une oeuvre, tandis que le récit linéaire forme l'ossature du *best-seller* et du récit filmé (cinéma, télévision). George P. Landow remarque que les nouvelles formes de la littérature (*Ulysses*, *Finnegans Wake*, Michel Butor, Marc Saporta, Jorge Luis Borgès, les romans Lation-américains), ne nécessitent pas l'usage de l'hypertexte pour être pleinement compris, mais que ce procédé montre de nouveaux principes d'organisation au cours de la lecture, faisant apparaître des éléments qui n'avaient pas été notés précédemment. Les stratégies d'écriture que l'on trouve chez ces auteurs, force le lecteur à trouver le thème non pas dans le récit, mais en filigrane à travers tout le texte ; ce sont des auteurs *difficiles*, parce que leurs stratégies d'écriture choquent le lecteur, l'obligent à construire un point de vue qui ne sera jamais explicitement décrit.

La notion de réseau mise en avant par la critique littéraire moderne, comme par l'informatique moderne, est essentielle pour l'hypertexte et pour la théorie structuraliste et post-structuraliste de l'écriture. Roland Barthes, Bakhtin, Derrida, Foucault, ont conçu le texte en terme de réseau. Michel Foucault liait ce modèle au projet d'une analyse archéologique du savoir, conçue comme une réinterprétation des savoirs, dévoilement des *épistémè*, recherche des généalogies, des diversités des approches "de jugement". En ce sens, il ne peut être fait abstraction d'une herméneutique ou d'une phénoménologie, en tout cas des divers niveaux de complexité où est interrogé un ensemble de données. L'ordre pour Foucault est pour une part, "la loi intérieure, le réseau caché", et le réseau est le phénomène qui relie ensemble une large étendue de classifications contradictoires, d'observations, d'interprétations, de catégories, de lois. Jacques Derrida utilise les termes de liens, trames, réseaux, matrices et entrelacements, tandis que Jean-François Lyotard écrit : "le soi est peu, mais il n'est pas isolé, il est pris dans une texture de relations plus complexe et plus mobile que jamais. Il est toujours, jeune ou vieux, homme ou femme, riche ou pauvre, placé sur des <<noeuds>> de circuits de communication, seraient-ils infimes. Il est préférable de dire : placé à des postes par lesquels passent des messages de nature diverse. Et il n'est jamais, même le plus défavorisé, dénué de pouvoir sur ces messages qui le traversent en le positionnant, que ce soit au poste de destinataire, ou de destinataire, ou de référent" (*La condition postmoderne : rapport sur le savoir*).

### 5.1. Rationalisation des liens et lecture assistée par ordinateur normes SGML et langage HTML

Quand l'usage du clavier pour entrer des informations et communiquer avec l'ordinateur, se répandit, tous les éléments étaient en place pour l'étude électronique de corpus de textes, mais les résistances étaient d'ordre psychologique : Jacques Vallée a raconté comment en 1967, alors qu'il travaillait pour une compagnie pétrolière française il suggéra à son patron qu'il serait

intéressant de pouvoir retrouver des informations sur les stations service et leurs gérants reliées entre elles. On lui répondit fermement que l'ordinateur était destiné à calculer, et non pas à des opérations "de type linguistique" (*The Network Revolution : Confession of a Computer Scientist*. Berkeley- AND/Or P, 1982). Le secteur particulier de l' Intelligence artificielle (ce terme date de 1956, le langage LISP date de 1959) était réservé aux milieux universitaires et militaires et n'avait pas de relation avec le monde économique. C'est seulement dans les années soixante-dix que la notion de traitement de texte popularisée par les micro-ordinateurs s'est répandue, tandis que l'idée de gestion de bases de données relationnelles étaient codifiée à partir de 1970 (CODASYL puis le model relationnel de E.F. Codd, mathématicien du centre d'IBM-San José), et que l'idée de traitement automatique des langues (langage PROLOG d'Alain Colmerauer au Canada puis à l'Université d'Aix-Marseille 1970) activité reliée au secteur de l' Intelligence artificielle, devenait un sujet d'études important . Le développement croissant de l'informatique a renforcé les recherches autour du traitement par ordinateur des langues naturelles. Les travaux de Noam Chomsky ont mis en évidence les propriétés des langues formelles et ont permis une grande avancée dans la compilation des langages de programmation d'abord, puis de l'analyse syntaxique des langues naturelles. La production ou la reconnaissance d'une phrase, sera alors commodément décrite au moyen d'un modèle de ré-écriture c'est-à-dire la paire <grammaire, stratégie>. Cependant reconnaître qu'une expression appartient au langage produit par une grammaire ne signifie pas qu'on a compris cette expression, lorsqu'elle est une phrase du langage naturel.

SGML est un langage normalisé de balisage de documents. Il ne s'agit plus de comprendre la totalité des inférences en langue naturelle d'un texte (traitement automatique du langage naturel ), mais de travailler sur la structure du document, sur la description des points significatifs. Cette tâche plus

modeste permet d'apporter une aide active à l'utilisateur en réintroduisant les concepts logiques dans le foisonnement infini du texte.

Comme l'indique son nom Standard Generalized Markup Language, il s'agit d'un métalangage, un langage pour définir d'autres langages, à savoir des langages de balisage, descriptifs qui se conforment à une norme de description des types de données et objets constitutifs d'un document, indépendamment des types d'ordinateurs, des systèmes d'exploitation et des logiciels de mise en page (cf : SGML-ODA, présentation des concepts et comparaison fonctionnelle : Danish Standards Association, Afnor technique, 1989 édition Danoise, 1991 AFNOR pour l' édition française, Victor Sandoval : SGML outil pour la gestion électronique de documents, Hermès, Paris, 1994).

SGML fournit un nouveau moyen pour traiter l'information et crée la possibilité de mettre toute l'information dans une base de données qui peut avoir des contenus divers (textes, graphiques, tableaux, video...) .

Un langage formel est par définition un langage ne possédant qu'une syntaxe et pas de sémantique. Un tel langage est très différent des langages naturels puisqu'il ne comprend qu'une grammaire et que le sens des mots n'intervient pas : tel est le langage de base de l'ordinateur muni d'un "automate de reconnaissance" permettant de reconnaître si une phrase appartient à un langage déterminé. Il s'agit d'un dispositif ayant un nombre fini d'états et muni d'une mémoire finie. C'est en partant non de la linguistique, mais de l'activité linguistique des individus, et en cherchant à résoudre un double problème , à savoir celui de l'individu parlant - capable en apprenant un nombre fini de mots et de phrases, de former ensuite un nombre infini de phrases syntaxiquement correctes - et celui de l'individu écoutant - capable de reconnaître la correction syntaxique d'une phrase qu'il n'a jamais entendue au préalable -, que Noam Chomsky aboutit à la théorie des grammaires génératives et des automates de reconnaissance . La théorie du langage chez Chomsky a été suivie d'un travail important sur la formalisation des langages artificiels, qui a consisté à étudier leurs propriétés, puis à définir leurs classes de hiérarchie.

suivie d'un travail important sur la formalisation des langages artificiels, qui a consisté à étudier leurs propriétés, puis à définir leurs classes de hiérarchie.

Un système de balisage de document étant établi, on peut utiliser des clauses de logique du premier ordre pour établir des inférences : c'est la destination du langage PROLOG ( acronyme de programmation en logique).

Une approche a été proposée à l'Université Joseph Fourier - Grenoble I : il s'agissait de l'étude et la réalisation de l'indexation du système d'information RIME système d'indexation d'un corpus de comptes rendus médicaux utilisant le langage PROLOG (Thèse de Catherine Berrut : *Une méthode d'indexation fondée sur l'analyse sémantique de documents spécialisés. Le prototype RIME et son application à un corpus médical*, 1988). Plus proche de la stratégie de l'hypertexte, Thomas Knopik et Sigrid Ryser dans un article intitulé : *A.I. methods for structuring hypertext information*.

SGML ayant pour but l'échange de la structure logique de documents, et PROLOG permettant de constituer des analyseurs de la structure d'une phrase ou d'un texte, on peut être tenté d'hybrider les deux techniques pour arriver à une structure d'assistance à la constitution de liens hypertextes.

### 5.1.1 Méthodes littéraires et traitement automatique du texte

Dans le premier recensement des métiers des bibliothèques publié dans le numéro 6 ( juillet 1995 ), du bulletin d'information de la Direction de l'information scientifique et technique et des bibliothèques, sous-direction des bibliothèques, apparaît " l'imminence de nouvelles transformations liées, notamment à l'informatique (diminution des coûts, développement des réseaux et du S.G.M.L., numérisation, multimédia, etc.)".

L'utilisation des normes de descriptions des structures logiques de documents, conduit à l'élaboration de projets en ingénierie bibliothéconomique qui reposent sur le traitement automatique du texte avant son édition électronique, et sur les possibilités inédites que ces traitements permettent. Le texte numérisé n'est plus un livre ordinaire digitalisé, mais un outil de lecture

tout à fait nouveau. Chaque texte peut être traité par un analyseur automatique qui a deux fonctions : il réalise une description exhaustive de tous les niveaux syntaxiques et morphologiques et constitue également un premier niveau d'analyse sémantique, ce qui permet la création d'index, qui peuvent être analysés par des réseaux sémantiques. Toutes ces analyses constituent des réseaux de relations qui peuvent être enregistrées dans l'index de recherche mis à la disposition de l'utilisateur. C'est pourquoi il importe de confronter les nouvelles possibilités offertes par la numérisation avec les méthodes littéraires modernes :

a) Mise en abyme

C'est André Gide qui le premier, emploie l'expression ; elle figure dans son *Journal*, en 1893, à propos de *La tentative amoureuse*. La "mise en abyme" consiste à réintroduire au sein même de la fiction un récit qui en reproduit les données. Le procédé est très fréquemment utilisé par les nouveaux romanciers mais il est porteur, chez chacun d'entre eux, d'une signification différente, voire opposée.

b) Déconstruction

La déconstruction rejette dès le départ la philosophie classique du concept. En traitant nos concepts comme s'ils étaient univoques, parfaitement distincts les uns des autres et modelés d'après la réalité, nous sommes victimes d'une illusion, d'après Jacques Derrida. Chaque concept garde une trace de tous les autres en particulier du concept opposé. *Hymen* signifie à la fois le mariage et ce qui lui résiste, *pharmakon*, le poison et le remède. En généralisant cette observation, Derrida trouve la source du sens dans un flux infini de traces et différences transcendantes. Depuis toujours, la philosophie de la tradition, la théologie et la science, en un mot la "métaphysique occidentale" se sont, cependant efforcées de réprimer la fluidité sémiotique transcendante, pour fixer la pensée en oppositions conceptuelles rigides : corps/esprit, sujet/objet, mâle/femelle, centres/marges. Déconstruire ces oppositions, en démolissant du même coup l'édifice de la métaphysique occidentale, est, dès lors, la tâche

centrale de la Philosophie. Ceci explique la place de la philosophie au centre du système des sciences : elle étudie les conditions de formation des concepts (ce qui semble justifier le choix de classement des collections à la Bibliothèque Nationale de France : rattacher la physique à la philosophie et lier l'art aux sciences (cf : Bulletin des Bibliothèques de France n° 4, 1995).

Aux États-Unis et dans les pays anglo-saxons, (Jacques Derrida est docteur Honoris causa de l'Université de Cambridge, bien que certains opposants à son élection aient affirmé que "la lecture de son oeuvre pouvait détériorer définitivement le cerveau"), cette thèse est largement diffusée dans les départements de littérature comme le montrent les travaux de George P. Landow mais se heurte, auprès des philosophes au mélange d'indifférence et d'hostilité que la tradition analytique réserve aux spéculations transcendantales. En 1977 un échange de vues a eu lieu entre Derrida et John Searle à propos de la philosophie des actes de parole. Cette polémique a été publiée aux Editions Galilée en 1988, sous le titre *Limited inc*, et montre le statut des oppositions conceptuelles. Searle est d'avis qu'entre des couples de notions comme sérieux/fictif ou central /marginal l'opposition est graduelle. Un énoncé peut être plus ou moins sérieux, plus ou moins fictif ou central/marginal l'opposition est graduelle. Derrida pense au contraire, qu'il n'y a pas d'oppositions sémantiques graduelles. Les véritables concepts philosophiques ne peuvent être que précis et discriminatoires. Le télos, l'idéal même du concept implique nécessairement la pure plénitude du sens. Idéalement, les concepts philosophiques excluent l'approximation et l'accomplissement partiel du sens est donc un trait essentiel des concepts, et pour en rendre compte, Derrida postule, à un niveau plus profond, l'existence d'une propriété transcendantale des concepts, qu'il appelle *itérabilité*.

Pour Derrida l'itérabilité garantit à la fois la plénitude du sens, en tant que norme du concept, et la possibilité de son échec, tandis que dans les années trente Gödel et Church énonçaient que tout système symbolique complexe contient des propositions dont on ne peut savoir si elles sont vraies ou fausses.

le  $\lambda$  - calcul (Barendregt H. *The lambda Calculus, its syntax and semantics* Amsterdam, Elsevier Science Publishers, 1981), a été introduit par Alonzo Church dans les années trente pour déterminer les fondements de la logique et des mathématiques. D'une certaine manière il s'agit d'une déconstruction de la métaphysique occidentale au même titre que l'oeuvre de Derrida. Celui-ci insiste sur le fait que l'itérabilité et sa logique opèrent non pas au niveau du discours ordinaire, mais à un méta-niveau plus profond qui sous-tend le discours ordinaire, la différence entre sa position et celle de Searle est d'ordre spéculatif. Alors que pour Searle le flou et la flexibilité sont des propriétés intrinsèques de certains concepts. Derrida soutient que tous les concepts sont essentiellement univoques, mais qu'à un niveau plus profond ils sont tous régis par une sorte de flexibilité transcendantale. Derrida s'efforce de saisir les régularités de l'émergence historique du concept. En posant un niveau idéal où les oppositions de sens fonctionnent à fond, et un méta-niveau où elles surgissent du magma des différences, Derrida fournit un puissant instrument à l'archéologie conceptuelle. Dans le domaine qui nous intéresse cette archéologie conceptuelle bute sur la question de la définition d'un ensemble et de la notion de fonction (relation entre des ensembles). Si l'information arrive rapidement de l'ensemble des points du monde, il importe de savoir choisir parmi la masse informationnelle la part, qui intéressera un utilisateur précis, à un moment précis. Pour chaque utilisateur la masse de documents pertinents est nécessairement limitée. Si la bibliothèque est l'une des clefs qui ouvre l'accès aux autoroutes de l'information, la recherche que fait le lecteur au sein d'une masse presque illimitée de données les plus diverses directement accessibles, demande une fonction d'assistance intelligente. La navigation dans l'hypertexte ne peut être utile que si les liens sont orientés de manière pertinente, si un travail de "base arrière" a été effectué en dehors de la présence du public.

A travers l'exemple des Postes de Lecture Assistée par Ordinateur on peut s'interroger sur l'évolution nécessaire de la théorie de la communication :

celle-ci, “ théorie mathématique sur la circulation de l’information montée à partir du modèle du téléphone ou, plus largement des télécommunications” ( Jean-Michel Salaün, Revue Réseaux N°58, mars-avril 1993), théorie statistique sur des signaux binaires ne rend pas compte des nouvelles possibilités offertes par la digitalisation et la dématérialisation de corpus de textes pouvant être consultés à distance ou transférés. George P. Landow remarque que lors de la lecture d’un texte littéraire digitalisé, la notion de “bruit” ne s’applique pas, puisqu’il ne s’agit pas de la réception d’un message, mais de la construction autonome indépendante d’une signification, d’un sens qui n’est pas forcément strictement égal à celui que l’auteur avait imaginé. On est plus proche avec l’hypertexte du point de vue de Roland Barthes (S/Z), que de la théorie de la communication de Shannon et Weaver (George P. Landow *Hypertext* opus cité p.73). Pour Monsieur Hervé Le Crosnier “c’est à une véritable refondation de notre profession que nous devrions assister dans les années qui viennent” (...) ” Nos bibliothèques ont pris l’habitude de traiter des documents clos (livres et même documents électroniques unitaires : un texte connaît des “versions”, que l’on peut gérer du point de vue bibliothéconomique comme autant d’éditions). Or les hypertextes en réseau sont des documents évolutifs”.

#### 6. État actuel du projet :

Depuis le mois de mars 1990, un banc d’essai réunissant neuf “grands lecteurs” a comme nous l’avons vu (paragraphe 4.2. p. 59) évalué les hypothèses sous-jacentes à la conception de la station en utilisant un environnement matériel et logiciel existant sur le marché et s’approchant du cahier des charges de la future station.

Deux projets ont été sélectionnés : l’un présenté par la société AIS, filiale de Berger-Levrault, l’autre par Cap Gémini Innovation. Les maquettes furent ensuite installées respectivement sur des stations SUN et NeXT.

Actuellement sur le site de l’établissement public à Ivry, ces applications facilitent la poursuite du travail d’expertise dans deux directions. D’une part,

les expérimentations portent sur la gestion des interfaces avec d'autres logiciels, et sur les outils complémentaires qui peuvent accompagner les fonctions centrales du PLAO, et d'autre part sur l'élargissement du public concerné. Des laboratoires du CNRS et d'autres équipes ont été invités à s'impliquer dans l'évaluation. Leur apport permettra de tester, sur des corpus qui peuvent être très spécifiques, la capacité de la machine à répondre à tous les besoins, et de faire l'inventaire des potentialités génériques à développer.

### 6.1 La coopération avec les institutions et les éditeurs :

Des conventions de partenariat ont été signées avec l'Institut des textes et manuscrits modernes (ITEM), les Archives Husserl (École normale supérieure), le Centre Louis Gernet (École des hautes études en sciences sociales), l'unité de recherche associée, Histoire des théories linguistiques (Université Paris VII), le Groupe d'études quantitatives interdisciplinaires sur la langue du XVI<sup>ème</sup> siècle (Université Blaise Pascal, Clermont-Ferrand), l'Institut Mémoire de l'édition contemporaine (IMEC), l'Université de technologie de Compiègne et l'Institut de recherche en informatique de l'Université Paul Sabatier de Toulouse (cf.: Michel Richard, *Le programme de numérisation de la Bibliothèque de France*, B.B.F t. 38 n°3, 1993) .

Le partenariat avec d'autres institutions spécialisées doit permettre de compléter les collections de référence. La coopération avec les Archives de l'Académie des sciences, service qui développe un projet de numérisation, facilite l'accès à des ouvrages difficilement communicables, les *Comptes-rendus des séances* ou disponibles en peu d'exemplaires, les *Mémoires des savants*. Des échanges sont également en cours avec la bibliothèque de l'École polytechnique. De même, des conventions avec les organismes disposant de bases textuelles informatiques ont pour but d'abonder le fonds initial de la bibliothèque. L'accord de partenariat avec l' Institut national de la langue française en est le meilleur exemple : il conduit à la mise à disposition de l'utilisateur du PLAO, non seulement toute les oeuvres rassemblées dans *Frantext*

mais également du *Trésor de la langue française*. L'existence d'un tel corpus, alimenté éventuellement par d'autres corpus linguistiques émanant de laboratoires associés, constitue un fonds unique au monde sur la langue française.

Enfin, l'une des ressources dont pourra disposer le fonds électronique de la bibliothèque viendra des prêts ou dépôts des bandes de photocomposition des éditeurs. En effet la norme ISO 12083, publiée en 1994, application de la norme SGML fixant le codage de la structure logique générique des monographies, des publications en série, des articles dans les publications en série et des formules mathématiques, sert de base pour toute la chaîne de l'édition, et sert de base de travail à toute personne désireuse de construire une DTD (Définition de type de document) pour mettre en oeuvre une application de SGML (cf. : Catherine Lupovici : *Le périodique électronique*, B.B.F. t.40, n°2, 1995). Dans ce domaine les obstacles ne sont plus techniques, mais portent sur la question des droits de reproduction des oeuvres.

## 7. Cartographier et analyser les textes numérisés :

### Bibliométrie, Scientométrie, Infométrie et Postes de Lecture Assistée par Ordinateur.

La recherche infométrique a porté dans un premier temps sur des données propres aux centres de documentation et à la recherche scientifique. La Bibliométrie, ensemble des méthodes et techniques quantitatives -de type mathématiques/statistiques - s'est développée par l'introduction de l'analyse quantitative dans le domaine de la bibliothéconomie (cf. : Monsieur Thierry Lafouge : *Un outil de mesure des prêts à la bibliothèque de l'ENSSIB*, Documentalistes - Sciences de l'information, 1994, vol.31, n°6). La numérisation de grandes quantités de textes, jointe à la généralisation de la norme SGML conduit à envisager des recherches en texte intégral dans différents domaines du savoir. Tel est le sens des études présentées dans *Les sciences de l'information : bibliométrie, scientométrie, infométrie* sous la direction de

Monsieur Jean-Max Noyer (Presses Universitaires de Rennes, 1995), qui sont consacrées aux nouveaux outils dans les sciences sociales et humaines. Les outils statistiques utilisés par la Bibliométrie visent à élaborer des indicateurs concernant les *outputs* (publications) des diverses pratiques de recherche. Ce qui est visé : la classification, les fréquences et les types de distributions... bref tout ce qui permet d'aider à définir de nouvelles stratégies en matière d'acquisition, de mise à jour, de gestion des bibliothèques ou des bases de données. Pour mettre à jour les dynamiques, les flux, les relations à l'oeuvre dans le champ scientifique, des outils, des méthodes sont en cours d'étude. Tels sont par exemple les outils scientométriques qui aident à faire émerger les frontières d'un champ de recherche, les limites d'une discipline. A l'origine, les techniques fondées sur l'analyse des co-citations (d'auteurs et / ou de revues cf.: E. Garfield, et le *Science Citation Index*), puis des techniques proches de celles mises au point par le Centre de Sociologie de l'Innovation de l'École des Mines, se sont attachées à faire apparaître les acteurs-réseaux hétérogènes, à cartographier ces derniers, et ce afin de tenter d'approcher ce que Gilles Deleuze et Félix Guattari appellent : "Agencements Collectifs d'Énonciation" et "Équipements Collectifs de Subjectivation". La méthode des mots associés, considère les mots clés comme des indicateurs de connaissances (contenu des documents indexés) et se base sur leurs co-occurrences pour mettre en évidence la structure de leurs relations (clusters). L'idée de co-occurrence est essentielle. En effet, si l'on considère que deux documents sont proches parcequ'ils sont indexés par des mots clés similaires, alors deux mots clés figurant ensemble dans un grand nombre de documents seront considérés comme proches. Cependant la co-occurrence ne permet pas à elle seule de mesurer la force des associations entre deux mots clés (leur proximité), car elle avantage les mots clés de haute fréquence par rapport à ceux de basse fréquence. L'emploi d'un indice statistique approprié permet de normaliser la mesure de l'association entre deux mots clés. Dans une étude récente (parue dans *Les sciences de l'information ...*, opus cité), la nature textuelle des données

à analyser, la diversité de leurs structures à amené les chercheurs à adopter la norme SGML pour la description de la structure logique des documents, et des outils UNIX, tels LEX ou Awk pour extraire l'information sur un flot de données structurées.

La numérisation de grandes quantités de documents rend possible l'utilisation de logiciels de ce type dans le cadre des Postes de Lecture Assistée par Ordinateur. Monsieur Bernard Stiegler a décrit les logiciels Candide et LECAO, permettant la génération automatique de graphes "de cartographie sémantique". Le graphe d'un texte permet une navigation d'un nouveau type, qui est possible parce qu'un texte, ou un corpus de textes forme un ensemble de tissus, une réalité plane et cependant non dénuée de profondeurs : syntaxique, sémantique, narrative, métrique. Le processus de numérisation des textes et la plasticité qui en découle ouvre de nouveaux horizons à la Bibliométrie, à la Scientométrie, et l'Infométrie.

## Conclusion et propositions de recherches :

Les bibliothèques sont confrontées de manière centrale, vitale pour leur existence, à l'apparition de nouveaux supports à acquérir, classer, conserver, présenter et offrir au public. Leurs caractéristiques sont différentes de celles de l'imprimé et imposent de nouveaux choix, différents de ceux du secteur commercial. L'information écrite, orale, digitalisée se vend bien et en grandes quantités : c'est l'élément moteur de l'économie nouvelle, comme le remarquait Jean-François Lyotard dans *La condition postmoderne* dès 1979. L'actualisation récente du manifeste de l'Unesco sur les bibliothèques publiques est un témoin de cette évolution. Alors que le développement rapide des produits informationnels, leur donne une place centrale dans le domaine des échanges marchands, les bibliothèques intègrent la révolution technologique dans un ensemble formant réseau. Elles réaffirment leur rôle culturel spécifique, face à ceux qui tenteraient une vision purement distributrice de leurs ressources.

Avec la numérisation de textes et des images, la Bibliothèque Nationale de France ouvre la voie de l'avenir. La numérisation permet de diminuer la pression sur les collections et à ce titre intéresse les bibliothèques : la communication d'un document numérisé évite d'avoir à prêter l'original ; le texte numérisé est communicable simultanément à plusieurs lecteurs, et peut même être accessible à distance. Pour le lecteur, la numérisation facilite et élargit l'accès aux collections. Elle lui apporte de nouvelles possibilités de lecture savante assistée par ordinateur. Il peut se constituer sur son ordinateur un corpus personnel de textes numérisés. Il peut pour éclairer le texte, faire appel à des documents s'y rapportant (dictionnaire, éditions critiques, biographies, articles...). Il travaille son texte directement sur écran, structure ses documents, effectue ses recherches comme sur n'importe quel ordinateur, ce qui représente un gain de temps considérable. Il sera possible de télécharger sur son micro-ordinateur tout ou partie d'un texte numérisé et accéder aux collections de diverses bibliothèques.

Les Postes de Lecture Assistée par Ordinateur et la numérisation sont un lieu stratégique pour observer cette évolution : si la numérisation permet de protéger le patrimoine tout en le communiquant, l'hypertexte, les logiciels d'indexation et d'annotation dynamique (LECAO), la normalisation (SGML, HTML, HyTime), la connexion à Internet sont les outils, avec l'activité liée aux bases de données (BN-OPALE, PANCATALOGUE, CCN), d'une assistance active au travail intellectuel de tous niveaux, d'une approche raisonnée de la connaissance, version contemporaine de l'ambition des encyclopédistes : "exposer autant qu'il est possible, l'ordre et l'enchaînement des connaissances humaines".

C'est pourquoi des recherches portant sur la pratique des Postes de Lecture Assistée par Ordinateur pourraient être d'une grande utilité.

### Propositions de recherches :

#### 1°) Les Postes de Lecture Assistée par Ordinateur :

il s'agirait d'étudier :

- comment sont intégrées les fonctions de constitution d'un corpus personnel
- la définition de la lisibilité : choix des écrans, vitesse de défilement, création de fenêtre
- l'aide à la rédaction, l'édition, la publication, traitement de texte, logiciels de Publication Assistée par ordinateur, la préparation d'exposés, cours, conférences (transparents, diapositives, tableurs, gestionnaires d'idées), impression, échange, transferts
- l'environnement de lecture : dictionnaire de langue, traduction de synonymes d'étymologie, mono et multilingues, encyclopédies, manuels de langue, grammaires, thesaurii spécialisés, bases de données bibliographiques
- la communication : entre la bibliothèque et les autres centres serveurs.

A ces fonctions contextuelles de base s'ajoutent les fonctions centrales des Postes

de Lecture Assistée par ordinateur :

- la structuration du corpus personnel par le balisage des concepts, des éléments structurant des textes
- l'analyse des textes, et la mise en lumière des structures créés à partir de la lexicographie, des mots-clefs, concepts, en utilisant des diagrammes, tableaux histogrammes, graphes
- l'annotation dynamique : inscription autour des textes de symboles, de commentaires associés à des passages, créations de liens entre passages, typage de tels liens, classement, indexation des passages annotés et des commentaires.
- archivage dans des dossiers structurés des textes et des commentaires, création de thesaurii personnels, bases de données textuelles
- interface entre la station et l'utilisateur

2°) La politique de numérisation à la Bibliothèque Nationale de France :

Le projet prévoit de numériser 100 000 textes et 300 000 images. L'étude pourrait porter sur :

- le choix des ouvrages, constitution d'un fonds et notion de collection numérisée
- la notion de collection numérisée entraîne-t-elle une modification de l'espace de la bibliothèque ?
- choix des procédés : mode image (bit-map) pour les textes anciens et illustrés mode texte favorisant l'utilisation de l'hypertexte
- les accords avec les éditeurs et les auteurs ce qui pose le problème textes numérisés et de la norme SGML. Peut-on utiliser directement les textes produits par les imprimeurs, qui emploient ces procédés pour la photocomposition, et pourraient les fournir au Dépôt légal ?

Bibliométrie, Infométrie, Scientométrie et Postes de Lecture Assistée par Ordinateur

Un dernier groupe de questions pourrait porter sur le marketing des Postes de Lecture Assistée par Ordinateurs, leur insertion dans les réseaux et enjeux économiques.

## Bibliographie :

Bachimont, Bruno

Le contrôle dans les systèmes à base de connaissance / Bruno Bachimont.- Paris : Hermès, 1992.

Balpe, Jean-Pierre

Hyperdocuments, hypertextes, hypermedia / Jean-Pierre Balpe.- Paris : Eyrolles, 1990.

Balpe, Jean-Pierre

Macro-structures et micro-univers dans la génération automatique de textes à orientations littéraires, colloque de Cerisy "la production automatique de textes littéraires, PUF, Paris, 1990

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

Barthes, Roland

S/Z / Roland Barthes.- Paris : Éditions du Seuil, 1970.

Barthes, Roland

L'empire des signes / Roland Barthes.- Genève : Skira, 1970.

Beaudiquez, Marcelle

"Le programme de numérisation est en bonne voie de réalisation"  
in Actualités de la Bibliothèque Nationale de France, n°2, Octobre-Décembre 1995.

Bemmington, Geoffrey

Jacques Derrida / Geoffrey Bennington.- Paris : Éditions du Seuil, 1991.

Berrut, Catherine

Une méthode d'indexation fondée sur l'analyse sémantique du document spécialisé / Catherine Berrut.- Grenoble : Grenoble I, Université Joseph Fourier, 1988.

Bibliothèque de France : 1992 l'année du socle / Établissement public de la Bibliothèque de France.- Paris : EPBF, 1992.

- Blasco, Philippe  
"La chaîne du savoir" in revue la Recherche, numéro spécial sur la mémoire, n°267 volume 25 juillet-août 1994
- Borges, Jorge Luis  
Fictions / Jorge Luis Borges.- Paris : Gallimard, 1983.
- Bush, Vannevar  
"As we may think" in Atlantic Monthly, n°176, Juillet 1945.
- Chartier, Roger  
L'ordre des livres / Roger Chartier.- Aix-en-Provence : Alinéa,1992.
- Chartier, Roger  
Les usages de l'imprimé / Roger Chartier.- Paris : Fayard, 1987.
- Chouchan, Dominique  
"Du livre à l'ordinateur" in la Recherche n°228, vol.22, janvier 1991.
- Les catalogues en ligne : enquête à la Médiathèque de la Cité des sciences et de l'industrie / coordination Mohammed Hassoun et Danielle Roger.- Villerbanne : Enssib, 1994.
- Conseil supérieur des bibliothèques  
Rapport du Président pour 1994 / Michel Melot.- Paris : Association du Conseil supérieur des bibliothèques, 1995.
- Derrida, Jacques  
L'écriture et la différence / Jacques Derrida.- Paris : Éditions du Seuil, 1967.
- Derrida, Jacques  
La Grammatologie / Jacques Derrida.- Paris : Éditions de Minuit, 1967.
- Derrida, Jacques  
La dissémination / Jacques Derrida. - Éd. du Seuil, 1972.
- Derrida, Jacques  
Ulysse gramophone / Jacques Derrida.- Paris : Galilée, 1987.
- Eisenstein, Elisabeth  
The printing press as an agent of change Cambridge University Press, 1979.

Engelbart, Douglas C.

A conceptual framework for Augmentation of Man's Intellect in Vistas in Information Handling, vol. 1, Spartan Books, Washington, 1963.

Goldfarb, Charles H.

The SGML Handbook / Charles H. Goldfarb.- Oxford : Clarendon Press, 1990.

Histoire et informatique : bases de données, recherche documentaire, multimédia / Martine Cocaud, ed.- Rennes : Université Rennes2, 1995.

Husserl, Edmund

L'origine de la géométrie/ Edmund Husserl, trad. et préf. de Jacques Derrida.- Paris : PUF, 1962.

Hypermedia / Hypertext and Object-oriented Databases, edited by Heather Brown. Unicom applied information technology 8. - London : Chapman & Hall, 1991.- 322 P. ; 27 cm.  
ISBN 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 )  
ISBN 0-07-0166622-6

Landow, George P.

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

Landow, George P.

Hypertext : the convergence of contemporary critical theory and technology / George P. Landow.- Baltimore, London : The John Hopkins University Press, 1992.- 242 p. ; 28 cm.

Latour, Bruno

"Vues de l'esprit" in Culture technique n°14, juin 1985.

Le Crosnier, Hervé

Systèmes d'accès à des ressources documentaires / Hervé Le Crosnier.- Aix-Marseille : Université Aix-Marseille 3, 1990.

Le Crosnier, Hervé

"Une introduction à l'hypertexte" in B.B.F. t. 36, n°4, 1991

Le Crosnier, Hervé

"L'hypertexte en réseau" in B.B.F. t.40, n°2, 1995.

Lyotard, Jean-François

La condition post-moderne : rapport sur le savoir / Jean-François Lyotard.- Paris :

Maignien, Yannick

"La bibliothèque virtuelle ou de l'Ars memoria à Xanadu" in B.B.F. t.40, n°2, 1995.

Nelson, Theodor N.

Literary Machines / Theodor N. Nelson.- Self published, 1981.

Nelson, Theodor N.

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

Peirce, Charles Sanders

Écrits sur le signe / Charles Sanders Pierce.- Paris : Éditions du Seuil, 1978.

"Pôles associées à la Bibliothèque Nationale de France" in B.B.F. t. 40, n°4, 1995

Propositions pour une grande bibliothèque : rapport au Premier Ministre / Patrice Cahart, Michel Melot.- Paris : la Documentation française, 1989.

Rada, Roy

"Ranking documents with a thesaurus" in Journal of the American society of information science, 40, 1989, 5, p.304-310.

Richard, Michel

"Le programme de numérisation de la bibliothèque de France" in B.B.F. t.38,n°3, 1993

Richy, Hélène

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

Roger, Danielle

"Catalogues en ligne accessible par le public, recherche exploratoire" in B.B.F. t.39, 2, 1994.

Salaün, Jean-Michel

A qui appartient la télévision ? / Jean-Michel Salaün.- Paris : Aubier, 1989.- 239 p. ; 28 cm.- (Res Babel).-ISBN 2-7007-1855-0

Salaün, Jean-Michel

Marketing des bibliothèques et des centres de documentation / Jean-Michel Salaün.- Paris : Éditions du Cercle de la Librairie, 1992. - 133 p. ; 31 cm. (Collection Bibliothèques)

"Les sciences de l'information en question : le point de vue de l'utilisateur", Jean-Michel Salaün, pp. 9 à 25, in Réseaux n°58, communication technologie société / CNET-CNRS, sous la dir. de Paul Beaud, Patrice Flichy. -Issy-les-Moulinaux, 1993.

Serres, Alexandre

"L'hypertexte une histoire à revisiter" in Documentaliste Sciences de l'Information, mars/avril 1995, vol.32, n°2.

Simondon, Gilbert

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

Simondon, Gilbert

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

Stendhal

Oeuvres complètes/ dir. V. del Litto & E. Abravanel. - Paris : Éditions Champion, 1967-1973.

Stiegler, Bernard

La technique et le temps .1. La faute d'Épiméthée / Bernard Stiegler.- Paris : Galilée, Cité des Sciences et de l'industrie, 1994.- 279 p. ; 30 cm.- ISBN : 2-7186-0440-9

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

Syrien, André-Pierre

”La bibliothèque localement universelle” in BBF t.40,n°3 1995

Van Herwijnen, Éric

Practical SGML / Éric Van Herwijnen.- Dordrecht : Kluwer Academic Publishers, 1990.

Virbel, Jacques

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

Using HTML / Tom Savola.- Indianapolis : Que Corporation, 1995.

Young, Douglas A.

X Windows : programmation avec les Xt Intrinsics / Douglas A. Young.- Paris, London : Masson, Prentice Hall, 1993.