

# La lecture assistée par ordinateur

## ou l'informatique au service de la lecture

par Josiane Coyac  
*Ingénieur Documentaliste*  
*Ecole polytechnique*

*Parmi les éléments les plus novateurs de la Bibliothèque de France, il convient de mettre l'accent sur les postes de lecture assistée par ordinateur (PLAO), outils totalement nouveaux au service du lecteur-chercheur, qui le mettront en communication avec l'ensemble du système d'information de la Bibliothèque de France et avec les ressources externes. Le but de cet article, tiré d'un travail de D.E.A réalisé à la Bibliothèque de France de janvier à mai 1992, est de clarifier les caractéristiques de l'outil et de faire un point sur les ressources externes disponibles pour les chercheurs et des aides qui en favorisent l'accès.*

### **Les caractéristiques essentielles du PLAO**

La lecture assistée par ordinateur portera sur un fonds important de documents sous forme électronique, constitué par la Bibliothèque de France dans le cadre de sa politique de numérisation et d'acquisition. Ce fonds comportera à l'ouverture de la bibliothèque en 1995, environ 100 000 ouvrages en mode image. A partir de ce fonds et, ultérieurement, de bases externes, le poste de PLAO pourra :

- mettre à disposition des chercheurs dans une sorte de bureau individuel

ou "carrel" un outil informatique de type ordinateur personnel relié au système d'information de la Bibliothèque de France par un réseau local et constituant un nouvel instrument de lecture ;

- à partir du poste de travail, accéder au catalogue et survoler rapidement le fonds électronique de la Bibliothèque de France, y sélectionner des ouvrages ou dans une phase ultérieure des documents d'autre nature - images fixes, séquences sonores en particulier ;

- Télécharger dans le poste de travail les ouvrages du fonds identifiés comme

pertinents pour le travail du chercheur et constituer ainsi un corpus de recherche personnel accessible localement ;

- enrichir éventuellement ce corpus par des fonds numériques personnels apportés par le chercheur, ou d'autres sources externes ;

- à partir du poste de travail, effectuer sur un corpus individuel de recherche des opérations de lecture active et d'appropriation du texte, de constitution de bases de données personnelles, de comparaison, de structuration, d'annotation, de recherche et d'analyse textuelles. Il s'agit ainsi pour les chercheurs de retrouver et d'enrichir, dans un cadre informatique, les fonctions courantes du travail en bibliothèque, où lecture et écriture se combinent étroitement.

Ce PLAO comportera en périphérie des moyens de restitution (une imprimante), des moyens de saisie (un scanner avec un logiciel OCR - reconnaissance optique des caractères). Environ 300 PLAO seront, à terme, répartis dans tous les départements thématiques de la bibliothèque, d'autres pourraient être installés à l'extérieur de la BdF dans les bibliothèques associées.

## ■ Les fonctions du PLAO

Les fonctions attendues pour le PLAO <sup>(1)</sup> en 1995, toutes rassemblées sous une interface permettant de passer d'une application à une autre, sont les suivantes.

### **Accès**

Ce sont les fonctions permettant l'identification et la demande des documents dans le fonds de la BdF, leur parcours rapide, leur feuilletage. Ces fonctions d'accès sont également disponibles en cours de lecture dans la mesure où l'interface avec l'utilisateur permet de les appeler à l'écran à côté des textes affichés.

### **Saisie - acquisition**

Il s'agit de systèmes permettant d'introduire des documents et des données dans le poste de travail grâce à un dispositif général de téléchargement. Seront ainsi

déchargées les données accessibles par le réseau de la BdF ou par les systèmes de numérisation des textes (scanner, OCR...), les images fixes (scanner et/ou caméra de digitalisation), voire animées (magnétoscopes de digitalisation), les sons, ainsi que les systèmes de lecture de produits éditoriaux électroniques. Il sera possible de convertir les données déjà saisies sous un autre format dans le format utilisé par le poste de travail.

### **Lecture**

Ce sont les fonctions des dispositifs matériels et logiciels assurant de bonnes conditions de lecture des textes numérisés à l'écran, qu'ils soient en mode image ou caractère. Elles portent sur la taille de l'écran, sa définition, son inclinaison, son éclairage, la vitesse de défilement des textes, la possibilité de zoomer, l'agencement des parties présentées et sur tout autre élément de confort pour la lecture.

### **Organisation dynamique des données textuelles**

Cette fonction permet l'organisation de textes "en vrac", en mode image ou caractère, sous forme de base sur laquelle des opérations de recherche peuvent être effectuées. Cette opération peut prendre la forme d'une extraction automatique d'un index, d'une reconnais-

ce de structure de page ou d'une adjonction de repères bien identifiés permettant la définition et le repérage automatique des unités documentaires constituant la base, en particulier en ce qui concerne le mode image.

Elle permet de transformer un ouvrage ou un ensemble d'ouvrages et de documents numérisés en une véritable banque de données, et d'y effectuer très rapidement des opérations de recherche (recherches lexicographiques, calculs d'occurrences, localisation de concepts, recherches croisées) sur un corpus constitué selon les besoins de l'utilisateur.

### **Cartographie sémantique**

Cette fonction offre une représentation de la façon dont des éléments de textes préalablement organisés sont liés : il s'agit, à partir d'un travail séquentiel d'indexation, de construction du thésaurus d'un texte, ou plus généralement d'organisation des données textuelles, de montrer les résultats des choix effectués automatiquement ou par l'utilisateur sous forme de graphique. Cette fonction fournit ainsi une cartographie sémantique de l'ensemble textuel étudié.

### **Annotation - Glosier**

Il s'agit de la possibilité d'inscrire à côté du texte, en mode image ou caractère,

(1) - extraits du cahier des charges du PLAO

des signes propres au lecteur, des notes, qui en constituent une sorte de commentaire, et qui peuvent prendre des formes très variées telles que traits, croix, soulèvements, entourage, mots, phrases, références bibliographiques. Cette fonction doit permettre l'intervention directe sur le texte commenté, avec des possibilités graphiques et éventuellement des opérations de classement automatique des passages annotés.

Par ailleurs, le lecteur peut également effectuer sur les annotations, dont le

volume en cours de travail devient vite important, des opérations de manipulation et de gestion électronique comme sur les textes eux-mêmes ; autrement dit, les annotations peuvent à leur tour être traitées comme un texte, être éventuellement interprétées et donner lieu à des exécutions.

#### **Connexion - Comparaison**

Cette fonction permet de comparer différentes versions d'un texte (éditions, traductions) - en mode image et en mode caractère, d'opérer des rattachements entre plusieurs textes ou textes/ images/sons, de constituer de véritables dossiers sur un thème rassemblant des documents issus de fichiers différents.

Les liens établis entre documents doivent être typés et commentables.

Certains liens appellent une taxonomie et donc une standardisation, et posent le problème de la structuration homogène et a priori des fichiers liés entre eux (ainsi de liens entre traductions diverses d'un même texte, qui supposent le repérage automatique des concordances, par des balises du type pagination originale, structure des chapitres et sous-chapitres, etc...) - système très apprécié des traducteurs, mais également des lecteurs en général pour la comparaison entre différentes versions d'un même texte.

#### **Classement et thésaurus automatiques**

Il s'agit de récolter le travail effectué, de le ranger avec un classement, d'avoir un contrôle de l'ensemble des opérations utilisées en cours de lecture. C'est l'équivalent des dossiers que l'on constitue au fur et à mesure d'un travail de recherche, où l'on range des textes, des articles, des annotations, des fiches et sur lesquels on colle des étiquettes avec des noms.

Il y a en fait construction d'un thésaurus au fur et à mesure de l'expression des choix effectués par le lecteur, cette fonction permet de contrôler en lui fournissant *l'historique de tous ses choix antérieurs*, sous quelque logiciel que ce soit, en sorte que le système lui signale, lorsqu'il réemploie un item, qu'il est déjà en usage, et dans quels contextes.

#### **Copie**

Il s'agit d'une fonction élémentaire de duplication de documents ou de fragments de documents sur divers types de supports externes (disques ou bandes, impression laser, etc...). Des problèmes juridiques sont évidemment posés par l'existence de cette fonction. Elle doit relever ce qui est copié, de façon à pouvoir fournir au système la trace des opérations de sortie effectuées sur les documents. Ces statistiques seront le cas échéant utilisées pour la facturation ou l'évaluation des droits.

#### **Edition**

Ce sont les fonctions des logiciels de traitement de texte et d'édition, y compris la PAO. La gamme des fonctions s'étend des logiciels élémentaires de traitement de texte à des produits d'édition plus sophistiqués intégrant par exemple des fonctions d'aide linguistique ou de traitement de langues alphabétiques ou non. Ces produits pourraient être complétés de logiciels d'aide à la préparation d'exposés, cours et conférences, de constitution et d'édition de bibliographies, voire même de maquettage pour la mise en page en vue d'édition.

#### **Environnement interne**

Ce groupe de fonctions englobe des aides à la traduction automatiques, des dictionnaires électroniques et des encyclopédies, des instruments du type "plan d'idées", ou tout autre outil d'aide à la lecture disponible à partir du poste de travail.

#### **Group-ware**

Il s'agit des fonctions permettant des travaux sur des fichiers communs à partir de plusieurs postes de travail reliés entre eux, la conférence assistée par ordinateur en étant la forme la plus courante.

#### **Communication locale et externe**

Ce sont les fonctions permettant, à partir du poste de travail, de passer des commandes de documents divers aux services de la bibliothèque, et les dispositifs de communication avec d'autres postes de travail internes ou externes à la Bibliothèque de France permettant principalement de transférer des fichiers. L'accès à des messageries électroniques fait aussi partie de ce groupe de fonctions.

### **Intendance**

Ce sont les services de réservation des salles de travail ou de conférence, de repas ainsi que les divers services communs de ce type offerts par la Bibliothèque, essentiellement sur des messageries prévues à cet effet.

### **Scénarii types**

Ce service consiste à donner une visibilité sur l'ensemble des fonctions qui sont disponibles sur le poste de lecture : il peut prendre la forme de la composition de plusieurs scénarii standards d'utilisation des autres services, complémentaires les uns des autres et déterminés en fonction d'ensembles cohérents de traitements sur un corpus de travail. Un exemple de scénario d'utilisation consiste à commencer par une annotation d'un texte puis à le transformer en une base de données et enfin à effectuer une interrogation sur le texte ainsi organisé.

### **Les ressources externes**

Les caractéristiques du PLAO doivent lui permettre d'utiliser les ressources externes offertes par les bases de données dont la multiplicité, la diversité et la richesse constituent une source précieuse d'informations à la disposition des utilisateurs. On estime aujourd'hui qu'il existe plus de 4000 banques de données disponibles dans le monde et couvrant tous les secteurs d'activité et de la connaissance humaine. La télématique a favorisé cet essor en permettant à tout utilisateur équipé d'accéder à ces fonds documentaires distants.

Présenté à ses débuts comme le concurrent direct des banques de données en ligne, le CD-ROM (Compact Disk-Read Only Memory) en devient un instrument complémentaire, souvent l'étape préalable, pour qui le peut, à une interrogation de banques de données en ligne. Il offre une *consultation améliorée* grâce à la convivialité proposée par les interfaces, le chercheur peut tester les concepts choisis pour l'interrogation sans contrainte de temps passé en ligne.

## **Les aides à la recherche d'information**

Les systèmes en ligne ont des langages de commandes différents les uns des autres, les banques de données sur un

même serveur ont quelquefois des structures différentes. Donc, pour l'utilisateur occasionnel, une grande partie de l'effort consenti à la recherche d'information n'est pas productif. Cet effort est principalement consacré à la localisation de la source d'information, à l'accès à cette source, à la lecture des manuels d'instructions et, quand enfin l'information est obtenue, au tri et à l'analyse des réponses et, dernière étape, à l'acquisition des documents correspondants. En fait, au cours d'un processus de recherche seule une faible part du travail effectif est consacré à l'extraction de l'information, le reste du temps étant sacrifié à l'appropriation momentanée de l'outil. C'est pourquoi se développent des interfaces d'aide à la recherche, certaines sont proposées par les serveurs eux-mêmes, d'autres sont développées par des sociétés qui ne sont ni producteurs ni serveurs. Ne sont indiqués ci-dessous que les développements les plus significatifs.

### **Développements proposés par les serveurs installés sur leur site**

Il s'agit soit d'offre de menus, soit de commandes particulières à chacun des serveurs permettant la plupart du temps de naviguer entre les diverses bases abritées par le serveur et ou de proposer des orientations nouvelles à partir de la question posée.

### **Anté-serveurs ou "Gateways"**

Un gateway ou antéserveur est un serveur intermédiaire placé entre le terminal de l'utilisateur et le serveur de banques de données documentaires. L'utilisateur se connecte au gateway et interagit avec lui, par contre les connexions entre le gateway et les serveurs lui sont totalement transparentes. L'antéserveur doit être un point d'entrée unique pour l'accès à une large variété de banques de données, ce qui simplifie les aspects contractuels et comptables en ayant un seul interlocuteur, et la procédure de connexion toujours la même. Il doit proposer un grand répertoire de serveurs, ce qui devrait élargir le nombre de banques de données offertes et offrir des interfaces intelligentes et conviviales pour faciliter le choix des banques et l'interrogation.

Il existe trois gateways opérationnels (EASYPNET, Ii, DGIS), qui possèdent plus

ou moins les fonctionnalités décrites ci-dessus. Mais il faut signaler que toutes les bases logées sur un serveur ne sont pas systématiquement disponibles à travers l'antéserveur.

### **Front-ends ou logiciels frontaux**

L'expression front-end est employée pour signifier des logiciels placés généralement sur le micro-ordinateur de l'utilisateur et gérant l'accès aux bases de données. Dans ces interfaces, l'aide aux chercheurs a continuellement augmenté si bien que certaines peuvent être appelées intelligentes.

Les plus simples sont les logiciels de communication, ils furent même les premiers à exister et sont nombreux sur le marché. Certains sont fournis d'office avec la carte modem. Parmi les plus courants, on peut citer *Pc-Talk*, *Dialolink*, *Infolog*, *Com2400*...

D'autres front-ends plus évolués mais encore relativement simples et mécaniques offrent en plus un traitement élémentaire du langage de l'utilisateur et l'aident à respecter les règles syntaxiques et les règles de transcodification nécessaire pour aboutir aux langages d'interrogation spécifiques. *IT*, *Sci-mate Searcher*, *Search Helper*, *Iani* appartiennent à cette catégorie.

Des front-ends encore plus ambitieux utilisent largement l'intelligence artificielle et prennent alors en charge la conduite de l'ensemble de la stratégie de recherche. Appartiennent à cette famille des produits comme *Cansearch*, fondé sur le thesaurus *Mesh*, *Cite* utilisé à la National Library of Medicine, *Intellect* utilisé à la Bibliothèque du Congrès, *Tome Searcher*, réalisé par l'université de Londres et appuyé sur le thesaurus *Inspect*, *Méribel*, véritable station intelligente d'interrogation en cours de développement à la Mission Information Scientifique et Technique du Commissariat à l'énergie atomique.

## **■ Applications pour le PLAO**

Sur les PLAO, même si les chercheurs ne maîtrisent aucune technique de recherche "on-line", ils peuvent être considérés comme des spécialistes de leur domaine. Ceci leur donne l'avantage de trouver les termes appropriés à leur demande et de déceler facilement et

rapidement la pertinence d'un document. Il faut donc suivre avec attention tous les logiciels "front-end", en particulier tous ceux qui développent des aides linguistiques, fort appréciées par les utilisateurs novices ou des utilisateurs à la limite de leur domaine de compétence. L'idéal serait de pouvoir consulter thésaurus, dictionnaire encyclopédique et multilingue du domaine correspondant à la recherche d'information.

Les deux systèmes "front-end" et "gateway" ne sont pas exclusifs, ils pourraient être utilisés l'un comme l'autre en fonction des connaissances des utilisateurs et du confort qu'ils leur apportent.

### Exploitation des résultats d'une recherche sur des banques de données

Pour exploiter au mieux les possibilités de la recherche en ligne et du PLAO, celui-ci devrait avoir, outre les fonctions décrites plus haut, quelques fonctions supplémentaires telles que le reformatage des références et des données téléchargées. Des logiciels tels que *Infotrans* et *Bibliolinks* offrent des possibilités intéressantes dans ce domaine.

Les fonctions de dédoublement seront aussi indispensables pour repérer les informations fournies en double par les divers serveurs utilisés. Certains serveurs (Dialog, BRS, DataStar) ont développé cette possibilité dans leur système, par exemple en utilisant la commande RD (Remove Duplicate), ce traitement économise le nombre de références et le temps de téléchargement. Il existe aussi des logiciels travaillant en local intégrant cette fonctionnalité. *Infotrans* par exemple repère avec une bonne performance les doublons, qu'ils proviennent d'un seul fichier ou de fichiers différents.

### Calendrier prévu

Dans une phase intermédiaire, des maquettes du PLAO de deux types différents ont été réalisées. Elles servent à expérimenter les fonctions exposées ci-dessus sous des formes simplifiées. Ces maquettes ont été distribuées auprès d'organismes universitaires pour être testées, en particulier chez les "grands

lecteurs" de la BdF jusqu'en Décembre 1992. Suivra une série de prototypes dont l'insertion dans l'ensemble du système d'information de la BdF sera expérimentée jusqu'en 1995.

### Conclusion

Après une période de formation à la technique de recherche bibliographique en ligne, le chercheur sur son PLAO pourra naviguer de banques de données en banques de données à la poursuite de l'information désirée. Des aides évoluées et bientôt conviviales l'assisteront dans ses requêtes. Au besoin une coopération plus personnalisée lui sera proposée par le Service de recherche bibliographique de la BdF.

Il lui appartiendra aussi d'aller lire le contenu de sa boîte aux lettres électronique, où attendront parmi les divers

messages de collègues, l'article commandé en ligne, la liste des dernières acquisitions de la BdF, les dernières actualités sur les banques de données ou le sommaire de sa revue préférée. Des aides pour un tri du courrier électronique sont en cours d'étude.

Malgré les apparences, il ne s'agit pas là d'un projet de science-fiction, mais d'un produit en cours d'étude et qui fait appel à des technologies connues. Une grande part de l'originalité du concept tient à cette mise en œuvre, dans un seul outil, des apports du traitement de texte, de la numérisation, de la recherche en ligne et de la télématique. Il reste à espérer qu'au-delà de la Bibliothèque de France, d'autres bibliothèques publiques et d'étude et de recherche puissent à terme disposer de Postes de Lecture Assistée par Ordinateur dans des conditions économiquement supportables.

### BIBLIOGRAPHIE

Barthes, C., Glize, P., Carpeutat, B. *Un système expert en recherche documentaire multibase et multiserveur*. RIAO85, Grenoble, 1985.

Bassano, J.C., Archambault, D., Desroques, G., Mekaouche, A. *Utilisation des systèmes experts dans des interfaces pour la recherche documentaire*. Agard CPP-505, 1991, p 3/1-3/14.

BIBLIOTHEQUE DE FRANCE. *Cahier des charges des PLAO*, 1991.

BIBLIOTHEQUE DE FRANCE. *Rapports des groupes de travail*, 1990-1991.

Fluhr, Christian, Machard, Claudine. *Gateways "intelligents"*. Agard CPP-505, 1991, p 2/1-2/8.

Kinyon, William R. *Current awareness: are front-end systems the answer?* National online Meeting. New-York, 1987, p 245-251.

Landau, Herbert B. *A data-base producer's view of online vendor gateways*. Information services & use, 1987, vol. 7(4/5), p 163-165.

Lardy, Jean-Pierre, Bador, Pascal. *Utilisation de références téléchargées et importation dans un SGBD*. Documentaliste, 1992, vol. 29-1, p 35-39.

Larrieu, Françoise. *Systèmes intelligents en recherche d'informations. Le projet AQUEDUC : les mécanismes de consultation*. Thèse de doctorat Toulouse III, 1989.

Larsen, G. *Searching the intelligent gateway Easy Net : the end-user's point of view*, Electronic library, 1987, vol. 5(3), p 146-151.

Machard, Claudine. *Station "intelligente" d'interrogation de bases de données en ligne : choix d'une base de données et formation de la question*. Mémoire d'ingénieur CNAM en informatique, 1991.

Still, Julie. *Using Easy Net in libraries*, Online, 1991, vol. 15(5), p 34-37.

Trautman, Rodes. *An expert system for microcomputers to aid selection of online databases*, Reference librarian, 1989, vol. 23, p 207-238.

Vickery, Alina. *Intelligent interfaces for online searching*, Aslib information, 1989, vol. 17-11/12, p 271-274.