

OPACITÉ ET TRANSPARENCE DES CATALOGUES INFORMATISÉS POUR L'USAGER

par Pierre Le Loarer

consultant chez GSI-ERLI

«... Il m'offre de me conduire à la salle des catalogues et de m'y laisser seul, bien que ce soit en principe interdit, les bibliothécaires seuls ayant le droit d'y travailler. Ainsi, je me trouvai réellement dans le saint des saints de la bibliothèque. J'avais l'impression, je t'assure, d'être entré à l'intérieur d'un crâne. Il n'y avait rien autour de moi que des rayons avec leurs cellules de livres, partout des échelles pour monter, et sur les tables et les pupitres rien que des catalogues et des bibliographies, toute la quintessence du savoir, nulle part un livre sensé, lisible, rien que des livres sur des livres : ça sentait diablement la matière grise,... »

Robert Musil

L'homme sans qualités, t. II.

Seuil, 1973, p. 225 (Folio n° 489).

« Il y a un vertige taxonomique. Je l'éprouve chaque fois que mes yeux tombent sur un indice de la Classification Décimale Universelle (CDU). Par quelles successions de miracles en est-on venu, pratiquement dans le monde entier, à convenir que :

668.184.2.099

désignerait la finition du savon de toilette et

629.1.018-465

les avertisseurs pour véhicules sanitaires, cependant que :

621.3.027.23

621.436:382

616.24-002.5-084

796.54

913.15

désignaient respectivement : les tensions ne dépassant pas 50 volts, le commerce extérieur des moteurs Diesel, la prophylaxie de la tuberculose, le camping et la géographie ancienne de la Chine et du Japon ! »

Georges Perec

Penser/Classer

Hachette, 1985, p. 162.

*
**

Les bibliothèques représentent un puits de connaissances potentiellement accessible au plus grand nombre. Les bibliothèques françaises évoluent et ne méritent plus, à quelques exceptions près, la remarque sévère¹ que Jean Paulhan leur adressait dans son carnet de réflexions journalières alors qu'il avait vingt ans.

L'ouverture réelle des bibliothèques au public se traduit de différentes manières : d'abord par des modalités d'accès et/ou d'inscription, ensuite par la qualité des services offerts à l'utilisateur. Parmi ces services, tout ce qui facilite l'utilisation des ressources documentaires proposées et donc, au premier chef, la connaissance de celles-ci, joue un rôle fondamental.

1. « Il y a la Bibliothèque Victor Cousin. Elle présente cette particularité remarquable qu'il est possible à ses seuls conservateurs de la fréquenter. Il n'y a qu'en France que les bibliothèques soient faites pour les bibliothécaires ». In : « Carnet du jeune homme », *Le nouveau commerce*, 1977, p. 32. Dans son contexte, l'expression « faites pour » semble simplement signifier que les bibliothèques sont « réservées aux » bibliothécaires. L'expression « faites pour » peut aussi être interprétée comme « conçues pour », ou plus globalement comme « conçues pour les bibliothécaires et réservées à ceux-ci ».

La connaissance des fonds documentaires d'une bibliothèque provient essentiellement de trois démarches : le contact direct avec les collections; la consultation du catalogue² et de documents complémentaires au catalogue (listes d'acquisition, thématiques, descriptions de la nature des fonds, etc.); enfin, le renseignement demandé au professionnel (bibliothécaire).

Ces différents moyens de prendre connaissance des richesses offertes ne s'excluent évidemment pas, mais la situation particulière de chaque bibliothèque peut favoriser, voire rendre déterminant, un moyen par rapport aux autres : une bibliothèque dont la majeure partie des fonds est en magasin³ limite de fait la connaissance des fonds aux deux dernières possibilités indiquées précédemment; l'existence, l'insuffisance⁴, voire l'absence de service de renseignements⁵ pour les lecteurs peuvent selon les cas, indirectement, favoriser ou, au contraire, réduire, voire rendre difficile⁶ l'accès aux fonds documentaires.

Pour l'usager, la recherche à travers le catalogue reste souvent un moyen privilégié d'accès aux collections. Dans le cas de bibliothèques « en accès direct », le moyen privilégié est d'abord l'accès aux rayons qui permet le contact direct avec les collections. « Déambuler » ou « vagabonder » à travers les collections favorise tous types de démarches⁷ et peut

provoquer de réelles stimulations intellectuelles, mais ne résout pas toutes les difficultés. L'immédiateté apparente que suggère cet accès s'oppose à la médiation obligée qu'implique le catalogue. Celui-ci reste-t-il un « mal nécessaire » : le seul moyen que l'on ait trouvé de présenter en réduction un ensemble de documents en s'affranchissant du mode de classement physique des documents sur les rayons ? En ce sens, le catalogue permet également, à condition de savoir l'utiliser, une grande richesse d'exploration, rapide et efficace, facilitant ou orientant la consultation directe des documents. Le volume des collections d'une bibliothèque et, par contre-coup, le volume du catalogue constituent également un paramètre à prendre en compte dans l'utilisation d'un catalogue.

DES CATALOGUES AU SERVICE DE QUI ?

Qu'est-ce qu'un catalogue pour l'usager ? Cette question semble anodine. L'expérience de chaque bibliothécaire est ponctuée de constats concernant la méconnaissance des catalogues, les difficultés rencontrées par les usagers⁸ à en comprendre la logique, les contresens quant à la signification des mentions indiquées, les recherches infructueuses, etc.

Un catalogue est d'abord un instrument qui « condense » la représentation de collections physiques à travers une description élaborée selon des règles plus ou moins arbitraires et permet de retrouver (localiser, pour reprendre le terme bibliothéconomique) les documents correspondant à un critère de recherche donné (principalement : auteur, titre, sujet) ou à une combinaison de critères (auteur-titre, par exemple). Cette « condensation » concerne à la fois des éléments relatifs à la description bibliographique (la fiche d'identité du document) et à son contenu (de quoi parle tel document ?). Chacun de ces éléments obéit à des règles d'écriture et à des règles de classement conçues initialement dans une logique linéaire de présentation et de classement pour les catalogues manuels⁹.

2. A ce stade de notre réflexion, nous utilisons le terme « catalogue » de manière générique pour désigner l'ensemble des catalogues d'une bibliothèque sans faire de distinction selon leur nature (auteurs, matières,...), leur objet (couvrant un fonds spécialisé ou l'ensemble du fonds d'une bibliothèque), leur support (fiches bristol, microfiches, ou informatisé). De même, nous employons le terme « bibliothèque » de manière générique pour désigner différents types de bibliothèques et centres de documentation.

3. C'est malheureusement le cas d'un bon nombre de bibliothèques universitaires parisiennes, par exemple, pour des raisons qui tiennent à l'histoire, à l'organisation des locaux, etc.

4. Qui peut être mesurée de diverses manières en fonction de l'emplacement dans la bibliothèque, la disponibilité du personnel affecté à cette fonction, de sa qualification, etc.

5. Toutefois, il ne suffit pas de disposer de bureaux d'information pour le public pour que ceux-ci soient utilisés. La demande de renseignement constitue un acte volontaire qui exprime un certain rapport au savoir et à ceux qui sont censés le détenir... La qualité de l'accueil, les panneaux d'information explicatifs, et tout autre moyen (plan-guide de la bibliothèque, visites organisées, etc.) peuvent faciliter la démarche des usagers.

6. Nous renvoyons le lecteur au « modèle négatif en vingt et un points de la mauvaise bibliothèque » présenté par Umberto ECO lors d'une conférence prononcée le 10 mars 1981 pour célébrer le 25^e anniversaire de l'installation de la Bibliothèque communale de Milan dans le Palais Sormani, texte édité sous le titre *De bibliotheca*, en 1986, par l'Echoppe, Caen.

7. Cf. les études menées par le Service des études et de la recherche de la Bibliothèque publique d'information du Centre Pompidou sur le comportement des usagers, et notamment *Publics à l'œuvre : pratiques culturelles à la Bibliothèque publique d'information*, par Jean-François BARBIER-BOUVET et Martine POULAIN, Paris, la Documentation française, 1986.

8. Ce constat ne concerne pas seulement une catégorie de bibliothèques ou de lecteurs. Il peut être vérifié, à des degrés divers, aussi bien parmi un public d'étudiants, de chercheurs, d'enseignants dans une bibliothèque universitaire qu'auprès d'un public plus large en bibliothèque municipale.

9. Ce sont aussi bien les catalogues sur fiches que les catalogues imprimés sous forme d'ouvrages.

Si la remarque sévère de Paulhan est aujourd'hui abusive, ne pourrait-on affirmer, en revanche, qu'elle s'applique encore, non aux bibliothèques françaises, mais à leurs catalogues ? La formule « *Il n'y a qu'en France que les catalogues soient faits pour les bibliothécaires* », a-t-elle un sens aujourd'hui ? Les catalogues, dans la mesure où ils sont conçus et réalisés par des bibliothécaires, sont de fait, sinon réservés à ceux-ci, du moins bien adaptés à ceux-ci. Qui est mieux placé qu'eux pour les utiliser efficacement ? Il reste donc à concevoir et implanter dans les catalogues des dispositifs suffisants pour que les bibliothécaires n'en deviennent pas les seuls utilisateurs bénéficiaires. Seule une sérieuse étude de bibliothéconomie comparée nous permettrait de décider si cette situation est générale aux bibliothèques ou s'applique plus particulièrement à la France. Notre connaissance de quelques catalogues étrangers nous incline à penser que la situation des bibliothèques étrangères¹⁰ est sur ce point souvent proche de la situation française, même si certains dispositifs absents de la plupart des catalogues français sont davantage orientés vers les utilisateurs (par exemple, principe du catalogue dictionnaire, ouvrages sur un auteur présentés à la suite des ouvrages de cet auteur).

La formule « Il n'y a qu'en France que les catalogues sont faits pour les bibliothécaires » a-t-elle un sens aujourd'hui ?

Cette interrogation sur le destinataire réel du catalogue en suscite aussitôt une autre : comment rendre plus facile l'appropriation du catalogue par l'usager, c'est-à-dire par une personne qui n'a pas présidé à sa conception et ne dispose pas de connaissances préalables sur le contenu du catalogue et ses logiques d'accès ? Les panneaux d'information, les guides du lecteur, les modes d'emploi de toutes sortes constituent le vade-mecum minimal pour

faciliter l'usage de la bibliothèque et de son catalogue. L'apprentissage réel de la bibliothèque et de son(ses) catalogue(s) s'effectue d'abord à travers leurs utilisations concrètes.

CATALOGUES MANUELS ET INFORMATISÉS

Toutefois, la situation actuelle évolue : nous vivons — commençons à vivre — le passage du catalogue manuel vers le catalogue informatisé, le fameux « OPAC »¹¹. Cette évolution met en lumière les difficultés qui existaient précédemment (connaissance des conventions concernant la description, l'indexation, le classement à travers l'usage d'une classification, etc.), mais étaient moins visibles.

Linéarité ou rupture dans le cheminement

On peut toujours feuilleter un catalogue imprimé ou bien ouvrir des tiroirs et lire des fiches en observant une progression physique correspondant à un mode de classement donné (alphabétique auteurs, matières, systématique à travers la succession d'indices de classification).

Cette linéarité n'est que très imparfaitement reproduite par l'ordinateur aujourd'hui. Quel que soit l'OPAC choisi, la logique de cheminement suit une succession d'étapes : affichage de la liste de termes correspondant au point d'accès choisi (auteurs, titres, sujets) puis/ou affichage d'une liste de références en affichage réduit correspondant au résultat de la recherche, puis possibilité de sélection d'une référence dans la liste pour un affichage complet des informations (« pavé ISBD », cote, éventuellement disponibilité du document). La « gestion intellectuelle » des allers-retours d'un niveau à un autre, en fonction des réponses du système, la perception du point où on en est, des possibilités offertes ou non à un moment donné, la mémorisation des informations affichées précédemment (et pas toujours rappelées) sont autant de difficultés, voire de pièges, que doit surmonter l'usager.

10. Les différences entre pays résident à la fois dans l'offre en bibliothèques, dans leur richesse, dans leur conception, dans leur usage plus ou moins fréquent et plus ou moins intégré dans les pratiques socio-culturelles.

11. L'abréviation OPAC (Online public access catalogue) est utilisée ici en l'absence d'abréviation française pour désigner les Catalogues en Ligne Accessibles par le Public (« CLAP » n'est pas homologué...). La notion d'OPAC s'étend aujourd'hui à toutes les formes de catalogues informatisés (y compris les catalogues sur CD-ROM) et ne se limite donc plus aux seuls catalogues accessibles « en ligne ». Ajoutons qu'OPAC est particulièrement euphonique et qu'un « hasard linguistique » souligne, en français comme en anglais, une de leurs faiblesses actuelles.

De ce point de vue, les nouvelles approches de type « hypertexte », bien que prometteuses, devront être utilisées avec beaucoup de circonspection pour faciliter la navigation de l'utilisateur dans le catalogue et non l'y noyer définitivement.

L'écran du Minitel, d'un ordinateur individuel ou d'un terminal offre des caractéristiques visuelles limitées : l'avantage que présente l'interactivité (et son corollaire, la « dynamique d'affichage »), est contrebalancé par la capacité limitée d'informations visualisables au même moment et par le caractère éphémère que prend toute information affichée. Certes, les écrans graphiques de format 19 ou 21 pouces utilisés avec les nouveaux postes de travail (ou « stations de travail ») vont changer ces limites, mais de tels écrans ne sont pas encore utilisés dans les OPACs actuels¹² (à part quelques rares expérimentations) et ne résolvent pas par eux-mêmes la véritable difficulté que constitue l'interface homme/machine dans un catalogue informatisé. La réalisation d'un OPAC satisfaisant ne se réduit évidemment pas à disposer d'un écran de grande dimension. La cinématique des écrans, c'est-à-dire leur enchaînement en fonction des actions de l'utilisateur, joue un rôle essentiel dans la facilité — ou la difficulté d'utilisation — d'un système interactif. Les stations de travail apportent, outre leurs écrans de grande dimension, le multifenêtrage, les menus déroulants, la technologie graphique souris et sont généralement multitâches (possibilité de lancer plusieurs traitements à la fois). Ces nouvelles possibilités pourront être intégrées dans les OPACs de demain, mais exigeront un travail de conception d'autant plus grand. Rappelons le : *plus un système semble facile et naturel à utiliser, plus il a été difficile à concevoir et réaliser.*

Le visible et le caché

La recherche dans un catalogue informatisé souligne et rend explicitement visibles les essais infructueux : avec le catalogue manuel, l'utilisateur vagabonde de page en page, de volume en volume ou bien de fiche en fiche, de tiroir en tiroir à partir d'un catalogue physiquement visible ; avec la plupart des catalogues informatisés existants, l'utilisateur est rapi-

Plus un système semble facile et naturel à utiliser, plus il a été difficile à concevoir et réaliser

dement confronté à une sorte d'univers caché, un iceberg¹³ dont il ne voit qu'une facette en fonction de sa question de départ.

Le système lui renvoie à l'écran sa méconnaissance du lieu avec : soit des messages tels que « aucun document relatif à cette clé », « recherche infructueuse », « ce sujet n'est pas dans la liste », « 0 réponse, autre question ? » ; soit un affichage d'index qui situe alphabétiquement — et non logiquement — sa requête.

Dans un cas, l'utilisateur est confronté au silence du système, dans l'autre cas, il a la satisfaction illusoire d'obtenir une réponse, satisfaction qui peut se transformer en frustration ou agacement quand il mesure la distance qui sépare sa requête de la réponse du système. Les constats les plus fréquents soulignent chez les utilisateurs une réelle ténacité — peu d'utilisateurs abandonnent après un essai infructueux¹⁴ — et une volonté de maîtrise du système. Ce constat est aggravé par le fait qu'en cas de non réponse le système n'explique pas à l'utilisateur les raisons de l'échec de la recherche. Et pour cause : dans la plupart des cas, s'il était en mesure d'expliquer la raison de l'échec, cela signifierait également qu'il est capable de l'analyser et donc de proposer une alternative à l'utilisateur, que l'échec soit dû à une faute de frappe, une faute d'orthographe ou une mauvaise utilisation de touche fonction.

13. L'image de l'iceberg n'est probablement pas la meilleure, car elle peut laisser supposer que l'utilisateur voit un dixième du catalogue, ce qui est loin d'être le cas. L'image du microscope électronique mériterait peut-être d'être creusée : on a l'œil sur une portion infime d'un tout, tout dont l'on ne peut reconnaître les différentes parties que si l'on a déjà une idée de sa nature et de son organisation, et si l'on sait à travers l'utilisation optimale du microscope, exploiter différentes visions grâce aux différents rapports d'agrandissement utilisables. L'observation du ciel étoilé à travers une lunette astronomique pourrait être également productive de métaphores intéressantes sur le catalogue informatisé.

14. A condition toutefois que l'utilisateur ait réussi à « entrer dans le système » : il semble qu'un fort taux d'abandon quasi immédiat concerne les utilisateurs qui n'arrivent pas à obtenir le premier dialogue (via une touche fonction ou la commande adéquate). Pour prendre une image simple, rappelons qu'avant même toutes les difficultés qui peuvent être liées à l'utilisation d'un service Télétel via un Minitel, si l'utilisateur ne sait pas utiliser la touche « CONNEXION FIN », il ne pourra jamais accéder au service qu'il recherche.

12. Pour trouver des références détaillées sur les OPACs, consulter l'imposante bibliographie réalisée par Nathalie Nadia MITTEV et Eftimis Nikolaos EFTIHLIADIS : *A classified bibliography on online public access catalogues*, London, 1987 (British library research paper, 23).

Enfin, l'absence de réponse peut induire l'usager à penser, à tort ou à raison, que le document qu'il recherche n'existe pas dans la bibliothèque — signalons à ce sujet que l'absence de réponse signifie l'absence de références, car la plupart des OPACs donnent d'abord et principalement comme résultat des références de documents. La mention de la cote est, selon les systèmes ou selon la place

Le catalogue informatisé introduit une dimension nouvelle par rapport au catalogue manuel

disponible sur l'écran, placée à la suite de la mention abrégée de chaque référence retrouvée (correspondant au premier écran de résultat) ou nécessite pour être obtenue un écran d'affichage supplémentaire. Pour les bibliothèques dont les collections sont en accès direct, ne devrait-il pas exister systématiquement une fonction¹⁵ permettant de localiser des documents — avec un plan de la bibliothèque ou des indications topographiques précises — dans un domaine donné sans être obligé de faire le détour par les références ? Ce point d'ergonomie a des conséquences quant à la représentation du contenu des catalogues dans le système. Il signifie qu'il faudrait concevoir des « ponts » entre la classification utilisée et les intitulés des indices de classification, voire avec les vedettes matières.

Outre les difficultés signalées précédemment, le catalogue informatisé introduit une dimension nouvelle par rapport au catalogue manuel, à savoir le comportement des usagers par rapport à la technique et plus précisément à l'informatique (attirance, désir d'appropriation, crainte, rejet). Ce comportement sera variable selon la nature des usagers (caractéristiques socioprofessionnelles, niveau d'études, sexe, âge,...), et nous laissons aux sociologues le soin de mener les études adéquates en la matière. L'emplacement des terminaux, la possibilité de consulter individuellement le catalogue sans se donner en spectacle, sans être « surveillé » par un bibliothécaire ou un autre usager influenceront

sur l'utilisation de l'OPAC et selon les cas faciliteront la démarche des uns et/ou rendront mal à l'aise les autres.

Un autre constat, fait lors d'une enquête du Service des études et de la recherche à la Bibliothèque publique d'information, concerne l'utilisation ludique du catalogue informatisé : on joue à utiliser l'OPAC comme d'autres — ou les mêmes ? — utilisent certains services télématiques ou des jeux d'arcade... Seule une analyse de sessions complètes d'interrogation, d'observations et d'enquêtes auprès d'usagers permet de rendre compte de ce phénomène, de son caractère massif ou exceptionnel. Dans le corpus que nous analysons plus loin, nous avons pu simplement noter quelques tentatives d'entrer sur des services Télétel disponibles sur le 3615 (recherche par sujet infructueuse telle que : 3615 CODE LA 5 [sic]). L'absence d'enquête ne nous permet pas de donner une interprétation précise de ce phénomène : confusion entre l'OPAC disponible à partir du Minitel et les services Télétel, tentative de reroutage¹⁶ vers un autre service télématique car cette fonction a été annoncée et sera, dans certaines conditions, bientôt disponible sur Télétel, affirmation d'une demande d'ordre ludique ou informationnelle sur des programmes télévisés, etc. ?

Ubiquité du catalogue informatisé

Et cependant nul ne peut nier les avantages qu'apporte l'informatisation des catalogues à l'utilisateur :

- disponibilité immédiate du catalogue en différents lieux¹⁷ de la bibliothèque (multiplication virtuelle du catalogue en fonction du nombre de terminaux disponibles);
- accès à l'ensemble du catalogue d'un réseau (bibliothèque municipale centrale et ses annexes, différentes sections d'une bibliothèque universitaire, etc.), possibilités de visions du catalogue en sous-ensembles « logiques » (fonds jeunesse, régional, discothèque,...);
- accès au catalogue de la bibliothèque sans se déplacer à partir de son bureau ou de son domicile si un accès télématique existe, et dans

15. Quand cette fonction existe, elle est rarement implémentée de manière à être utilisée facilement et semble plutôt rajoutée telle une verrue au système.

16. En recherche par sujet, on peut repasser par un code fonction à une autre fonction : recherche par titre, par auteur, accès à la liste des nouveautés, etc. Pourquoi ne pas pouvoir passer à un autre service télématique ?

17. Ceci est également possible dans le cas de catalogues informatisés qui sont imprimés ou reproduits sur microfiches et non disponibles via un OPAC. La mise à jour des informations dépendra de la fréquence de réédition du catalogue complet ou de suppléments.

certains cas possibilité de réserver à distance un document;

— multiplication des critères de recherche (variable selon les logiciels et selon la saisie ou la récupération des données effectuée pour le catalogue informatisé) par rapport au catalogue manuel (par exemple, accès par titre ou mots du titre, ISBN, année d'édition, type de document,...);

— possibilité de croisement de critères de recherche et donc de recherches plus ou moins complexes (dépendant des logiciels utilisés et de l'organisation des données);

— mise à jour rapide du catalogue et connaissance quasi immédiate de la disponibilité des documents si l'OPAC est lié dynamiquement à la fonction de gestion du prêt.

Sur la rapidité de mise à jour, une remarque s'impose. L'OPAC évite l'intercalation manuelle des fiches, permet l'affichage automatique des dernières acquisitions, etc. Cette quasi-instantanéité de mise à jour reste dépendante du travail des bibliothécaires (catalogage effectué ou récupéré sur des sources extérieures, ajout des informations propres à la bibliothèque).

Tous ces avantages fonctionnels indéniables ne doivent cependant pas laisser croire que toutes les difficultés sont résolues et que maintenant, grâce à la technique, l'immensité du savoir s'offre à tous : n'importe quel usager n'a plus qu'à taper quelques mots sur un terminal et la réponse — adéquate par rapport à la formulation de la question — apparaît... Il peut exister un fossé entre ce qui a été prévu pour l'utilisateur et ce qui est réellement compris par celui-ci.

ANALYSE D'UN CORPUS DE QUESTIONS

L'analyse de corpus de questions réelles nous paraît tout à fait fondamentale pour tirer des enseignements de la pratique des usagers, mesurer la distance qui sépare la conception et la réalisation du catalogue de son usage réel, sous le double aspect de son contenu et de son mode d'accès. Nous avons pu disposer de listes journalières des « *mots matières dont la recherche n'a pas abouti* », provenant de la bibliothèque municipale de Valence. Malgré les limites d'un tel corpus¹⁸, ces listes présen-

tent l'avantage d'être un échantillon concret issu d'un catalogue informatisé existant et utilisé. Elles ne peuvent avoir valeur de « test représentatif », mais ont le grand mérite d'exister. Nous formulons ici le souhait que les bibliothèques ouvrant des OPACs se donnent les moyens d'analyser l'utilisation de ceux-ci. A cet égard, le travail réalisé par la Bibliothèque publique d'information mérite d'être souligné¹⁹.

Le catalogue informatisé de la BM de Valence contient 80 000 ouvrages dont 40 000 environ ont fait l'objet d'une indexation matières. L'OPAC (logiciel OPSYS) est ouvert depuis fin 1987. Le public y a accès par 7 Minitels présents dans la bibliothèque et 2 terminaux professionnels situés dans deux annexes. Le personnel y a également accès *via* des terminaux professionnels. L'indexation des ouvrages a suivi pour l'essentiel le choix de vedettes matières publié par le Cercle de la librairie²⁰.

Remarques liminaires

Avant de présenter les enseignements que nous tirons du corpus analysé, signalons-en les limites principales :

— il s'agit d'une simple liste de termes ou d'expressions n'ayant pas abouti, c'est-à-dire ne correspondant à aucune clé présente dans l'index matières du système, *et non* d'enregistrements complets de sessions d'interrogation par les usagers;

— il serait nécessaire de disposer des enregistrements de sessions d'interrogation²¹ pour analyser les stratégies de recherche des usagers, des réactions de ceux-ci en fonction des réponses du système, etc.;

— les questions ayant abouti sont absentes du corpus, ce qui ne signifie pas que les réponses obtenues étaient satisfaisantes du point de vue de l'utilisateur, mais seule une enquête sur le terrain appuyée par une solide méthodologie aurait permis de vérifier ce point;

— les questions concernant les auteurs et les titres sont, théoriquement, absentes du corpus;

— en l'absence de connaissance exacte du fonds documentaire concerné, il est peu aisé d'indiquer précisément si les recherches non

19. Cf. Joëlle LE MAREC, *Dialogue ou labyrinthe ? la consultation des catalogues informatisés par les usagers*, Paris, Centre Georges Pompidou BPI, 1989.

20. Martine BLANC-MONTMAYEUR, Françoise DANSET, *Choix de vedettes matières à l'intention des bibliothèques*, Paris, éd. du Cercle de la librairie, 1986.

21. Ces enregistrements conservent l'anonymat des usagers, puisque ceux-ci n'ont pas à décliner leur identité lorsqu'ils consultent un OPAC.

18. Une quarantaine de pages correspondant à quelques jours d'interrogation situés entre mai et novembre 1988 sur le système de la bibliothèque municipale de Valence. Nous remercions vivement Martine Blanc-Montmayeur de nous avoir donné accès à ces informations.

abouties correspondent ou non à des documents existant dans la bibliothèque²²;

— de même, ne disposant que du choix de vedettes matières éditées par le Cercle de la librairie et non de la liste exacte des vedettes matières de la bibliothèque, il n'est pas toujours facile de déterminer si les recherches non abouties correspondent à l'absence des notions recherchées dans la liste ou à des formulations inadéquates²²;

— il faut signaler que les renvois matières n'ont pas encore été entrés dans le catalogue informatisé;

— de telles listes ressemblent à un « inventaire à la Prévert » qui serait principalement constitué de notions abstraites et non de listes d'objets matériels.

Nos remarques ne concernent donc pas, dans leur essence, un système particulier²³ dont nous présenterions une évaluation, mais plutôt une série de constats qui donnent à réfléchir pour l'amélioration de systèmes existants et la conception de systèmes futurs.

Phénomènes marquants

Voici quelques phénomènes marquants que nous avons pu repérer :

— la typographie utilisée était, à quelques très rares exceptions, la typographie pauvre en majuscules (non accentuées) correspondant à la frappe sur clavier Minitel²⁴;

— les fautes de frappe sont assez fréquentes dans le corpus et peuvent être dues à une faible habitude des claviers et/ou aux réactions du système (par exemple, léger délai d'affichage d'une lettre tapée qui peut amener certains usagers à les retaper) : répétition de lettres, omission d'une lettre ou d'un espace, lettre à la place d'une autre proche sur le clavier, frappe de deux lettres à la fois, par exemple : AANIMAL, COMMUNISRE, GENEAL-

POGIE, GYMNASYTIQUE, HISTIRE FRANCE, INFORMATIQHE, LLOOUIS 13, RACIME²⁵, REGENCR, SOCIOLOGIE, etc.;

— les fautes d'orthographe sont également fréquentes : dans certains cas, on peut hésiter avec la faute de frappe (par exemple, SUREALISME peut être provoqué par un second R répété trop faiblement à la frappe — même phénomène possible dans MAMIFERES, MOLUSQUES, LITERATURE); dans d'autres cas, la faute est beaucoup plus probable (par exemple, ACCOUSTIQUE, AMITIEE, AREDECHE, DACTILOGRAPHIE, ENIGME POLICIAIRE, ETHIMOLOGIE, ETHYMOLOGIE, EXISTENCIALISME, MARKETING, NASISME, ORQUESTRATION, RECIS SUR LA GUERRE DU VIETNAM, RADIESTESIE, SEXE SIMBOLE, VOAYAGES, VOILLIER);

— le niveau de langue est principalement celui de la langue courante sans incursions notoires dans le vocabulaire savant ou très spécialisé, ni dans le vocabulaire trivial;

— il faut souligner le caractère encyclopédique et la variété des questions posées par les usagers (il s'agit d'un corpus provenant d'une bibliothèque municipale²⁶, dont le fonds documentaire est par nature encyclopédique);

— les questions posées ne dépassent que très rarement trois mots (sans compter les prépositions et les articles) : ECOLOGIE URBAINE, ECONOMIE ITALIENNE, INSECTES, L'HUMANISME, TECHNIQUE DE GESTION, VIE DES ABEILLES, etc.

— au-delà d'un seul terme simple dans une question, les formulations respectant la syntaxe du français sont plus fréquentes que la succession de mots-clé; par exemple :

LES CADRES DANS L'ENTREPRISE, DROIT CONSTITUTIONNEL ET INSTITUTIONS POLITIQUES, EXPEDITIONS DANS LE DESERT²⁷, FLORE ET FAUNE EN ARDECHE²⁸, HIERARCHIE DANS

22. Nous avons pu vérifier partiellement ce point en interrogeant nous-même le service *via* un Minitel et en reposant sous des formes différentes des séries de questions. Un changement de formulation a le plus souvent permis de retrouver des documents.

23. Il a parfois été nécessaire d'indiquer une particularité du système pour permettre au lecteur de comprendre plus clairement certains phénomènes que l'on pourrait retrouver, sous des formes différentes, dans d'autres OPACs.

24. Rappelons-le : la plupart des terminaux implantés à la BM de Valence pour l'OPAC sont des Minitels. Pour rappeler la forme typographique de la frappe sur Minitel, les exemples des questions d'usagers sont donnés en majuscules non accentuées.

25. Un tel exemple montre la difficulté et les pièges éventuels des techniques de « redressage » (recherche automatique des formes les plus proches). En recherche par sujet, on a tendance à penser que la recherche portait sur RACISME, mais elle était peut-être le fait d'un lycéen cherchant une étude sur RACINE ou d'une personne intéressée par les RACINES VEGETALES, etc. En ce cas, un système de redressage automatique doit préciser la transformation qu'il a effectuée et la faire valider, et aussi permettre à l'utilisateur de reformuler sa question.

26. Nous pensons que, dans une bibliothèque universitaire, on retrouverait le même phénomène à un niveau de finesse différent quant à la demande (questions plus spécialisées en sciences humaines, en sciences exactes, en médecine, ou pluridisciplinaires et suffisamment diverses pour être encyclopédiques dans un ou plusieurs domaines du savoir). Ce point devrait être vérifié par des enquêtes dans les BU (Metz, par exemple) qui proposent ou proposeront un OPAC.

27. Il existe la vedette matières : RÈCIT DE VOYAGE:DESERT.

28. Il existe la vedette matières : FLORE:ARDECHE.

LA RUCHE, LES INSTITUTIONS CHEZ LES ROMAINS, MATHÉMATIQUES EN MATERNELLE²⁹, ORGANISATIONS DES CONSOMMATEURS, LES PHILOSOPHES DU DIX HUITIÈME SIÈCLE³⁰, PROGRES DE LA MÉDECINE AU 19ÈME SIÈCLE, PROTECTION DE LA FORÊT, RAPPORT ENTRE LA SCIENCE ET L'HOMME, LA SOCIÉTÉ PARISIENNE AU 19ÈME SIÈCLE, VILLES D'EUROPE³¹;

et dans la logique de type mot-clé en respectant l'ordre des termes tel qu'on le trouve dans les vedettes matières des catalogues manuels : AZTEQUE EMPIRE³², ÉLECTRONIQUE COMPOSANTS³³, FLORE MONTAGNE³⁴, FRANÇAIS PRONONCIATION, HISTOIRE MÉDECINE 19ÈME SIÈCLE;

— quelques questions relèvent de demandes qui s'apparentent au conseil plus ou moins personnalisé ou à la recherche d'informations précises : AUTEURS LATINS, AUTEURS TURCS, CONTE FACILE, UN AUTEUR, AUTEUR LITTÉRAIRE, LIVRES SPORTIFS, ROMAN D'AMOUR, ROMAN ENFANT RÉVOLUTION FRANÇAISE, ROMAN POLICIER;

Notons le caractère ambigu de ces formulations : que recherchent les usagers ? Une liste d'auteurs latins ou des ouvrages d'auteurs latins ? Un dictionnaire ou un traité sur la littérature turque, ou bien des ouvrages d'écrivains turcs ? Des ouvrages pour pratiquer des sports ou des ouvrages sur des vedettes sportives ? Un roman pour enfant sur la Révolution française, mettant en relief la vie d'un enfant, ou écrit par un enfant sur la Révolution française,... ? Des études sur le roman policier ou des suggestions de romans policiers à lire ?

Quant au CONTE FACILE (Andersen, contes africains, histoires enfantines,...), à l'AUTEUR LITTÉRAIRE (Hugo, Zola ou Blanchot) ou au ROMAN D'AMOUR (Belle du seigneur, un ouvrage de la collection Harlequin ou bien un roman érotique), ces notions introduisent des

critères qualitatifs qui renvoient à des univers extrêmement variés en fonction des usagers, de leur histoire, de leurs pratiques, etc. La qualité, la richesse d'un dialogue humain permettra peut-être de déceler la demande latente, l'ouverture sur l'imaginaire, le plaisir et — qui sait ? — d'y répondre, ou se transformera dans certains cas en dialogue de sourds (« *Vous ne savez pas ce que vous voulez !* »)...

On hésite à interpréter certaines requêtes comme l'expression d'un besoin documentaire au sens classique ou comme un appel lancé face à une situation critique, telles les séquences suivantes :

ENFANT INCOMPRIS DE LEURS PARENTS, ENFANTS VOLEURS A 15 ANS, ENFANT.,

SIDA, SYNDROME IMMUNO [sic], TOXICOMANIES ET REALITES, TOXICOMANIE ENGENDRANT LE SIDA,

TRAVAIL CHOMAGE, TRAVAIL PARLEZ [sic]³⁵;

— la confusion titre/sujet est assez fréquente à travers des requêtes telles que : L'ASSOMMOIRE [sic], DICTIONNAIRE DES OEUVRES, LES MISÉRABLES, UN SAC DE BILLES, LA VIE DEVANT SOI, YAN [sic] QUEFFÉLEC NOCES BARBARES³⁶;

— un nombre important de sujets exprimés correspondent également à des titres d'ouvrages; dans la collection « Que sais-je ? », par exemple : L'ALGERIE, LA GESTION DU PERSONNEL, LA PHILOSOPHIE, LE RACISME;

— quelques requêtes concernent des noms d'auteur (prénom ou nom patronymique en tête selon les cas) et sont soit des recherches aux œuvres de l'auteur cité, soit des recherches à ces auteurs pris en tant que sujets d'étude³⁷ : ALBERT CAMUS, CIORAN³⁸, VICTOR HUGO, MAUPASSANT, SOPHOCLE, TOURNIER MICHEL;

— certaines séquences laissent penser, bien que la nature du corpus ne nous permette pas de le démontrer — comment prouver qu'il s'agit du même usager ? —, que, face aux différentes formulations possibles d'une même requête, l'usager en tente successivement plusieurs et soit les dégrade, soit les complexifie devant l'absence de réponse :

29. Il existe la vedette matières : ÉCOLE MATERNELLE:PÉDAGOGIE, mais rien d'adapté à la question posée sous MATHÉMATIQUES.

30. Il existe les vedettes matières : PHILOSOPHIE:1700-1780 et PHILOSOPHIE:EUROPE:1700-1799.

31. Il existe la vedette matières : EUROPE:GUIDE TOURISTIQUE et des vedettes matières au nom de différentes villes européennes. Il existe également ARCHITECTURE URBAINE.

32. Il existe la vedette matières : AZTEQUE:PEUPLE.

33. Il existe la vedette matières : ÉLECTRONIQUE:CIRCUIT.

34. Exemple donné dans l'écran d'accueil de la recherche par sujet, mais avec les deux points (FLORE:MONTAGNE). Est-ce une démarche de l'usager pour s'assurer d'obtenir une réponse (en reprenant la question suggérée et en omettant par inattention les deux points) ou une tentative de l'usager pour vérifier ce qu'il advient quand on omet les deux points et comprendre le mode de fonctionnement du système ?

35. Rien ne permet d'affirmer qu'il s'agit du même usager dans chacune des requêtes même si elles sont datées du même jour.

36. Il devrait s'agir, selon la logique bibliothécaire, d'une recherche sur les études effectuées sur ce roman de Yann QUEFFÉLEC. Le catalogue informatisé de la BM de Valence n'offre pas de recherche mixte auteur-titre.

37. Seule une enquête auprès des usagers permettrait de le déterminer avec certitude.

38. À moins qu'un I ne soit glissé par erreur (I et O sont voisins sur les claviers) et que la recherche n'ait porté sur le CORAN.

AFFICHAGE, AFFICHE PUBLICITAIRE, AFFICHES

AGENT TECHNIQUE FORESTIER, GARDE FORESTIER, METIER DE LA FORET, METIER DE LA MONTAGNE

COMMENT DEVENIR ECRIVAIN, ECRIRE
CONTE DE FEE ET LEGENDES, CONTES CELTIQUES, CONTES MERVEILLEUX

DECLARATION DE L'HOMME ET DU CITOYEN, DECLARATION DES DROITS DE L'HOMME ET DU CITOYEN

MALADIES SEXUELLEMENT TRANSMISSIBLES, MALADIES SEXUELLES

FLEXIBILITE DES HEURES DE TRAVAIL, FLEXIBILITE DES HORAIRES

MATIERE PLASTIQUE, MATIERES PLASTIQUES
MOBILIER CONTEMPORAIN, MOBILIER MODERNE

PHOQUES ET MORSES, CETACES (PHOQUES ET MORSES)

POESIE GERMANIQUE, POETES ALLEMANDS
POP ART, POP-ART, POPULAR ART

POLYMERES, POLYMERISATION

QUANTA, QUANTAS [sic], THEORIE DES QUANTA

SHAMPOOING, SHAMPOOINGS

TRAVERSEE DE DESERT, TRAVERSEES, TRAVERSER

VIE DANS LES CAMPAGNES, VIE RURALE

— le même phénomène peut être constaté avec les sigles :

CEE, CEE, LA C.E.E., COM ECO EUR³⁹
CGT, CONFEDERATION GENERALE TRAVAILLEURS, SYNDICAT CGT, SYNDICAT TRAVAILLEURS

— et avec les dates :

19IEME SIECLE, 19EME SIECLE, DIX NEUVIEME SIECLE

— certaines recherches indiquent une difficulté de dialogue : l'utilisateur peut passer d'un mode de recherche à un autre en tapant le code fonction correspondant (AUTeur, TITre, SUJet, REVues, NOUveautés). Certains usagers tapent l'ensemble de l'intitulé du code (AUTEUR, SUJET) et non les seules trois premières lettres (qui sont mises en valeur par un fond différent des autres lettres du code) : dans ce cas le système ne reconnaît pas le code fonc-

tion et effectue une recherche qui n'aboutit pas. D'autres usagers anticipent et font directement suivre le code fonction, tapé correctement en trois lettres, de la chaîne de caractères recherchée (par exemple, AUT QUEFFELEC), ce qui correspond à une anticipation par rapport au système et représente une performance bien mal récompensée : la requête n'aboutit pas.

Il faut préciser que le système actuel ne tolère pas la présence d'articles ou de prépositions dans la recherche par sujet alors que celle-ci est tolérée dans la recherche par titres. En recherche par sujet, il est indiqué dès l'écran d'accueil qu'il faut séparer chaque sujet par le signe deux points (:). Cette contrainte peu naturelle introduit un nombre élevé d'erreurs : LA GESTION DU PERSONNEL ne donne aucune réponse alors que GESTION:PERSONNEL et PERSONNEL:GESTION ramènent des références. Il en est de même avec :

PROTECTION DE LA FORET → FORET:PROTECTION ou PROTECTION:FORET,

ORGANISATIONS DES CONSOMMATEURS → ORGANISATION:CONSOMMATEUR ou CONSOMMATEUR:ORGANISATION, à condition d'être mis au singulier.

La difficulté réside dans le fait que certains thèmes sont « lexicalisés », c'est-à-dire exprimés sous forme de mots composés, et que d'autres ne le sont pas, sans qu'il soit facile de déterminer *a priori* dans quelle situation on se trouve. Ainsi, une recherche par DESSIN INDUSTRIEL donne un résultat alors qu'une recherche par DESSIN n'affiche que les vedettes matières dans lesquelles le terme DESSIN est présent en tant que vedette ou sous-vedette (pour reprendre le jargon bibliothéconomique), mais jamais la vedette matière DESSIN INDUSTRIEL n'est présentée. De même une recherche sujet par DECLARATION DES DROITS DE L'HOMME ne donne aucun résultat alors que par DROITS DE L'HOMME (qui a été lexicalisé), le système affiche toutes les vedettes matières où cette notion apparaît complétée par des sous-vedettes des noms de pays et/ou de dates.

La logique informatique traditionnelle bute ici devant une difficulté de taille :

— soit toutes les chaînes de caractères entre deux blancs sont accessibles quelle que soit leur position dans le terme composé ou l'expression (vedette matières, auteur, titre). C'est la technique utilisée (recherche sur fichier inversé avec un dictionnaire de mots vides pour ne pas tenir compte des articles et prépositions) pour la recherche par titre dans le système que nous avons examiné ici. Celle-ci peut introduire beaucoup de bruit (réponses

39. Il s'agit probablement d'une tentative d'utiliser la troncature droite permise avec l'astérisque en recherche auteur, titre et sujet. Le message suivant indique cette possibilité : « Pour tous les mots dont l'orthographe est incertaine, indiquez au minimum les 3 premiers caractères suivis de * ».

erronées) notamment dans les catalogues de taille importante;

— soit les chaînes de caractères avec blanc(s) sont reconnues dans leur globalité comme formant une expression, un terme de langue en tant que tel (DESSIN INDUSTRIEL, DROITS DE L'HOMME) et, en ce cas, le système répond exactement à la question à condition que celle-ci soit exprimée exactement avec les mêmes chaînes de caractères dans le même ordre (DESSINS INDUSTRIELS ou INDUSTRIEL DESSIN ne seraient pas retrouvés). Une telle approche permet aussi de s'affranchir des articles et prépositions si un dictionnaire de « mots vides » a été créé, mais la difficulté réelle pour l'utilisateur est de connaître l'expression sous laquelle a été entré le sujet d'un document.

Ce phénomène, dû à l'organisation physique des index dans les systèmes informatiques, explique les deux logiques d'accès que l'on trouve, conjointement ou séparément, dans la plupart des OPACs : soit par sujet, par titre, par

auteur, ou alors par mot⁴⁰ du titre, mot du sujet, élément du nom de l'auteur. Dans les petits catalogues, la technique du fichier inversé systématique (appelé improprement « langage naturel ») est probablement préférable, malgré le parasitage que cela entraîne dans les réponses. Dans les catalogues volumineux, elle présente l'inconvénient déjà souligné de provoquer du bruit et de noyer les réponses pertinentes parmi un grand nombre d'autres : les expressions ne sont pas reconnues prioritairement.

Il existe quelques palliatifs, tels que la troncature (pour tenter de gérer les flexions morphologiques), le masque (pour récupérer des mots dont un caractère varie) ou l'opérateur d'adjacence⁴¹ (pour rechercher dans un ordre et à

40. Ici mot signifie « chaîne de caractères entre deux blancs »

41. Si la troncature implicite ou explicite (astérisque ou autre caractère spécial) se rencontre dans les OPACs, en revanche la fonction de masquage de caractère et l'opérateur d'adjacence (distance maximale entre deux chaînes de caractères) semblent réservés aux systèmes de recherche documentaire traditionnels. Seul un utilisateur professionnel averti les utilise à bon escient.

une distance donnés deux chaînes de caractères), qui permettent de surmonter partiellement ces difficultés⁴². Leur usage dans un contexte grand public s'avère inapproprié, et dans certains cas désastreux. La seule troncature droite (la plus commune dans les systèmes et la moins difficile à utiliser) provoque souvent des phénomènes de bruit documentaire bien connus. Ainsi, la troncature sur VILLE*, par exemple, ramène en recherche sur titre dans le catalogue examiné (de taille moyenne) 15 termes⁴³ et en recherche par sujet 41 vedettes matières⁴⁴ dans lesquelles le terme ville apparaît en tête ou dans la vedette matières.

Les OPACs ont généralement réussi, par rapport à la recherche documentaire informatisée traditionnelle, à affranchir l'utilisateur du langage de commande (commandes et opérateurs de recherche, de désignation des champs à interroger, d'affichage des résultats, de mémorisation des questions, etc.). Ce n'est pas pour aujourd'hui réintroduire cette logique et introduire des barrières supplémentaires dans leur utilisation⁴⁵. L'évolution va plutôt dans le sens inverse : les systèmes de recherche documentaire cherchent aujourd'hui à s'affranchir de cette difficulté d'utilisation du langage de commande.

MOTS ET VEDETTES MATIÈRES

La reconnaissance des mots et des expressions (à bien différencier des chaînes de caractères situées entre deux blancs) reste donc l'une des difficultés majeures des techniques d'accès dans les OPACs comme dans les systèmes de recherche documentaire. Mais cette difficulté n'est pas seulement d'origine technique. Elle tient à la nature de l'information manipulée : il ne s'agit pas ici de manipuler des chiffres, des codes figés mais bien des termes de langue.

Ceux-ci obéissent à des règles de fonctionnement de la langue — en l'occurrence, le français, mais cela reste vrai pour n'importe quelle langue.

Les mots de notre langue présentent des variations flexionnelles (masculin, féminin, singulier, pluriel, conjugaison des verbes : participes passés, par exemple). Un nombre important d'échecs provient du fait que l'utilisateur n'a pas utilisé la forme flexionnelle stockée dans le catalogue (usage fréquent du pluriel alors que le terme est le plus souvent stocké au singulier). Pour un autre système, c'est l'usage du pluriel qui prédomine⁴⁶. Dans certains cas, les données sont hétérogènes, tantôt au singulier, tantôt au pluriel (et pas seulement quand il s'agit de variations de sens).

Les expressions et termes composés forment un tout d'un point de vue linguistique (constructions syntaxiques aboutissant à des contenus sémantiques précis). Leur éclatement arbitraire en unités indépendantes en brise la signification. Ce phénomène est beaucoup plus gênant quand il s'agit d'indexer automatiquement des textes de plusieurs lignes.

Les différentes constructions syntaxiques possibles d'une langue (constructions grammaticales, relations entre les mots) permettent de produire des variations d'énoncés qui compliquent encore la reconnaissance des notions manipulées dans les questions.

Chaque terme utilisé véhicule un sens ou plusieurs (polysémie). Ces termes peuvent être combinés entre eux pour créer de nouvelles unités sémantiques.

Les listes d'autorité matières et les répertoires de vedettes matières représentent une tentative de normalisation de ces ensembles de notions. Ces listes contrôlées obéissent à une syntaxe qui leur est propre (ponctuation, subdivisions diverses). Pour des raisons historiques qui tiennent à l'organisation des fichiers manuels, ces listes de vedettes matières sont précoordonnées (plusieurs notions présentes côte à côte dans une même vedette matières) et donnent donc une représentation sémantique plus ou moins riche correspondant au contenu d'un document. Ces listes représentent une mine d'or par leur richesse sémantique,

42. Utilisée automatiquement après un terme au singulier (pour ramener les termes au pluriel), la troncature échoue devant des formes flexionnelles fréquentes en français : -al/-aux, -eur/-euse, -teur/-trice, les pluriels irréguliers, etc. Quand elle fonctionne, elle peut provoquer beaucoup de bruit.

43. Dont voici la liste : VILLE, VILLEFRANCHE, VILLÉGIATURE, VILLEHARDOUIN, VILLEMAR, VILLEMER, VILLEMEUR, VILLEMOT, VILLENEUVE, VILLERS, VILLERVALLES, VILLES, VILLET, VILLETTE, VILLEURBANNE.

44. La troncature a l'avantage de présenter le terme *ville* dans différents contextes et donc d'aider l'utilisateur à choisir son sujet. EUROPE* (avec la troncature) ramène plus de 200 vedettes matières.

45. Si, par exemple, les fonctions de recherche permises par les opérateurs booléens s'avèrent utiles, y compris dans un OPAC, elles devront être implantées dans l'OPAC. Mais ceci ne signifie en rien que l'utilisateur doive manier explicitement ces opérateurs booléens.

46. Même si ce n'est habituellement pas la règle dans le cas des bibliothèques.

mais elles présentent de réelles difficultés d'utilisation pour les usagers (parfois aussi pour les bibliothécaires) et de gestion par des systèmes informatiques traditionnels.

Or un constat s'impose : les recherches par sujet sont les plus fréquentes dans les OPACs et, parallèlement, ce sont celles qui obtiennent les moins bons résultats. L'OPAC de demain devra donc proposer des solutions plus satisfaisantes.

LA COMMUNICATION HOMME/MACHINE

L'OPAC pose donc de front le problème suivant : comment affranchir l'utilisateur du langage documentaire utilisé ? Comment lui permettre de retrouver l'information tout en l'exprimant dans son langage ? Cette interrogation appelle deux commentaires.

Pour traiter correctement ces difficultés, il ne faut pas limiter la réflexion à un type de recherche particulier (par sujet), mais bien analyser l'OPAC comme un tout offert à l'utilisateur : d'une part, il faut rappeler que les recherches par auteur et titre posent aussi bon nombre de difficultés spécifiques ou existant dans les recherches par sujet (variation des noms d'auteurs, interrogation par un élément d'un nom de collectivité, etc.); d'autre part, les différents champs possibles d'interrogation ne sont pas hermétiques, comme nous l'avons rapidement montré à travers l'analyse du corpus, et une interaction entre eux devrait être exploitée en fonction de la requête de l'utilisateur et des résultats obtenus.

Pour concevoir un système d'interrogation cohérent, il faut prendre en compte différentes composantes d'une communication homme/machine : la *compréhension*, soit reconnaître les termes de la requête de l'utilisateur et faire expliquer les ambiguïtés éventuelles; la *coopération*, soit proposer des pistes à l'utilisateur pour qu'il puisse se retrouver dans l'univers qu'il interroge; enfin la *présentation*, ou afficher des résultats intermédiaires ou finals dans des formes adéquates à l'utilisateur.

Ces trois composantes ne sont pas obligatoirement séparées et ne se suivent pas séquentiellement de manière immuable selon un schéma simpliste compréhension-coopération-présentation.

Un exemple simple : à la suite d'une requête sur MICHEL TOURNIER, un OPAC pourrait, grâce à une stratégie de recherche simultanée sur différents critères de recherche, indiquer à l'utilisateur qu'il existe des documents « sur » et

d'autres « écrits par » MICHEL TOURNIER. Dans ce cas, le système lève l'ambiguïté (recherche par auteur ou recherche par sujet) par un dialogue de coopération qui résout en même temps la difficulté de compréhension.

S'il s'agit d'une requête infructueuse par sujet, le système pourrait indiquer à l'utilisateur qu'il existe seulement des ouvrages écrits par cet auteur et les proposer (dialogue de coopération). De la même manière, une recherche par sujet non aboutie pourrait être reprise par le système sur les titres et proposée à l'utilisateur s'il y a des réponses.

Ces stratégies de dialogue, avec des propositions d'élargissement ou d'affinement sur différents critères (ouvrages les plus récents, par exemple) présentées à l'utilisateur, éviteraient dans bien des cas des recherches infructueuses. Elles ont des conséquences en termes de réalisation de systèmes, de création d'index, de charge machine, etc.

Les recherches par sujet sont les plus fréquentes dans les OPACs et, parallèlement, ce sont elles qui obtiennent les moins bons résultats

A côté des principes ergonomiques de base qui relèvent de l'ergonomie « de surface » (lisibilité des écrans, messages clairs, cohérence des touches fonction ou des commandes d'un écran à l'autre, possibilité de retour en arrière, etc.), des stratégies de dialogue plus élaborées pourraient être implantées dans les OPACs. Ces stratégies, fondées sur des représentations de connaissances, introduiraient « l'ergonomie cognitive » dans les OPACs. Meilleure sera l'ergonomie profonde (cognitive) du système, moins l'ergonomie de surface (de présentation) sera utilisée pour suppléer aux insuffisances du système et plus elle sera appréciée par l'utilisateur. A quoi sert de disposer d'écrans bien conçus si ceux-ci ramènent trop souvent des réponses erronées ou du silence ? La conception d'un OPAC plus facile à utiliser passe par la définition du niveau adéquat de représentation de connaissances dans le système. La brièveté des questions posées par les utilisateurs n'exige pas d'analyseur syntaxique très poussé (à l'opposé de ce qui peut être nécessaire quand il faut analyser plusieurs pages de

textes⁴⁷). En revanche, des processus de dialogue plus ou moins sophistiqués impliquent de disposer de connaissances suffisantes sur l'univers sémantique du catalogue informatisé concerné par l'OPAC.

Comment le système peut-il retrouver des documents sur la RUCHE s'il n'a aucune information lui indiquant une relation avec l'APICULTURE et les ABEILLES, sur les MANIPULATIONS GÉNÉTIQUES et la MORALE si les documents traitant de ce sujet sont indexés sous BIOLOGIE:MORALE et que rien n'indique le lien entre BIOLOGIE et MANIPULATIONS GÉNÉTIQUES, sur SHAMPOOING quand il faut chercher à CHEVEU:SOIN, etc. ?

Seule une expertise détaillée à partir d'analyses de questions d'utilisateurs, de sessions d'interrogation, d'observation et d'enquête auprès des utilisateurs permettra de définir le niveau *ad hoc* de représentation de connaissances (analyses morpho-lexicale et syntactico-sémantique, sémantique et parfois pragmatique) permettant d'obtenir le meilleur ratio investissement/qualité des résultats.

Il y a probablement un seuil (80 % de réponses satisfaisantes ?) difficile à dépasser sans investissement disproportionné par rapport au bénéfice. Mais atteindre les 70 ou 80 % de résultats représenterait un saut qualitatif indéniable par rapport à la moyenne des systèmes actuels, moyenne très difficile à mesurer. À partir de quel moment peut-on considérer le résultat d'une recherche comme satisfaisant ? Les enquêtes auprès des utilisateurs donnent un résultat subjectif intéressant, mais insuffisant s'il n'est pas croisé avec d'autres données. Le stockage des requêtes non abouties (telles celles du corpus examiné) est précieux et pourrait être comparé au nombre de requêtes ayant donné un résultat. Il ne faudrait cependant pas en conclure trop vite qu'une recherche ayant donné un résultat doit être systématiquement considérée comme une recherche aboutie dont le résultat est satisfaisant. S'il est clair que la logique de l'utilisateur n'est pas une logique d'exhaustivité⁴⁸, la difficulté réelle à

surmonter est d'éviter le bruit et le silence et de présenter des réponses (références) ou des orientations pertinentes à l'utilisateur.

DES APPROCHES NOUVELLES

Il existe aujourd'hui différentes techniques qui permettent de lever certaines des difficultés présentées. Il est remarquable de noter que, dans le corpus analysé, la plupart des fautes d'orthographe (et non des fautes de frappe) auraient pu être redressées par un logiciel de recherche phonétique. Il existe des algorithmes permettant de redresser des fautes de frappe. Il existe également des méthodes permettant de retrouver des termes, simples et composés, indépendamment de leurs flexions, de traiter les expressions et bien d'autres phénomènes plus ou moins perturbants dans les recherches (mots à préfixe avec ou sans tiret, sigles, dates).

Les difficultés massives liées au lexique et à la sémantique pourraient trouver leur résolution par l'utilisation d'automates exploitant des grammaires syntaxique et sémantique. L'information lexicale et la représentation sémantique nécessaires (relations entre les mots, relations hiérarchiques, de voisinage, classes de termes géographiques, de dates,...) pourraient être en grande partie récupérées des listes matières existantes. Les classifications pourraient servir de base à une organisation hiérarchique des domaines de la connaissance couverts par les catalogues. Cette récupération de données devrait être complétée d'informations provenant de dictionnaires de langue générale.

Ces techniques sont maintenant passées du laboratoire au stade industriel et mises en œuvre dans des applications réelles⁴⁹ sur des configurations informatiques peu coûteuses. L'OPAC de demain intégrera ces techniques dans une conception générale repensée de la communication homme/machine intégrant des procédures de relâchement « intelligent » de contraintes de recherche, de propositions de pistes de recherches à l'utilisateur⁵⁰.

47. Sur l'analyse des problèmes posés par l'indexation automatique des textes, cf. Bruno MENON « Indexation automatique et intelligence artificielle : quelques questions de stratégie » dans l'ouvrage collectif *Image et intelligence artificielle dans l'information scientifique et technique* (cours INRIA, Bénodet, 1988), Le Chesnay Rocquencourt (78), INRIA, 1988.

48. Trouver toutes les références concernant un sujet donné n'est généralement pas le problème de l'utilisateur. Quelques références, voire une seule, peuvent répondre à une demande.

49. Cf. les développements réalisés par GSI-ERLI pour l'annuaire électronique et le Minitel guide des services, dont bon nombre des problèmes posés se retrouvent dans les OPACs. Cf. Grégoire CLEMENCEAU, « Querying the French yellow pages : natural access to the directory », *Information processing & management*, 1988, vol. 4, n° 6, p. 633-649.

50. Pour des commodités de langage, nous avons souvent eu tendance à parler de l'utilisateur, de l'utilisateur comme si cet être mythique formait une personne idéale ou, à parler des utilisateurs, des utilisateurs comme si ceux-ci formaient un ensemble cohérent dont chaque élément se ressemblerait. De même que le Français moyen n'existe pas, si ce n'est dans les études statistiques, de même l'utilisateur moyen est une notion abstraite. Les OPACs devront s'adapter à la diversité des utilisateurs et non correspondre à un modèle unique de représentation de l'utilisateur.

La situation française nous semble intéressante à plus d'un titre. L'informatisation grandissante des bibliothèques françaises et, partant, de leurs catalogues, l'existence de listes d'autorité matières informatisées, le développement des OPACs⁵¹, l'existence du vidéotex, des accès publics aux services télématiques et des réseaux appropriés donnent l'infrastructure nécessaire pour réaliser les OPACs de demain. Ceux-ci pourront devenir des services Télétel⁵² offerts parmi d'autres.

Dans le même prolongement, le catalogue informatisé de demain évoluera aussi dans la nature de son offre : il ne sera plus limité aux seuls documents, mais pourra devenir aussi un catalogue de ressources indiquant des organismes, des services, des associations face à certaines demandes d'information. La combinaison de l'offre entre informations primaires et informations secondaires dans les mêmes services ou *via* des passerelles (reroutage à venir entre services télématiques) se développera et permettra de mieux répondre aux besoins immédiats des usagers.

51. Ce thème devient porteur en France. Il suffit de noter le succès remporté par la journée d'étude organisée sur « les catalogues en ligne » par la Médiathèque de la Cité des sciences et de l'Industrie (Paris, la Villette) le 21 septembre 1988. Ce numéro du *BBF* en est une autre illustration.

52. C'est déjà le cas de plusieurs OPACs, dont celui de la Médiathèque de la Villette (*via* le 3614). La plupart des OPACs français sont accessibles par un numéro de téléphone local direct (et non par Télétel). C'est le cas, par exemple, des BM de Valence et de Charenton-le-Pont (Val-de-Marne). Que les autres bibliothèques nous excusent : la BM de Valence nous a fourni le corpus de questions et GSI-ERLI est situé à Charenton !

Une autre évolution envisageable concerne l'analyse de l'ensemble des documents d'une bibliothèque. Aujourd'hui, l'utilisateur comprend difficilement que les périodiques ne soient, en général, pas analysés dans les bibliothèques au niveau de leurs unités réellement significatives, à savoir les articles publiés. Il est probable que l'on verra des passerelles entre bases de données documentaires et catalogues de bibliothèques, voire de véritables intégrations par extraction de bases de données pour la partie correspondant aux titres de périodiques possédés par la bibliothèque (ceci ne sera pas sans poser des problèmes de cohérence entre les structures de données existantes, les langages documentaires utilisés).

Mais toutes ces évolutions impliquent que le mode de recherche le plus utilisé, à savoir l'accès par sujet, soit beaucoup mieux adapté aux usagers et intégré dans une conception générale de l'OPAC. Ceci exige de mener des études et des recherches précises sur l'utilisation des systèmes existants, d'entreprendre des études de conception, la réalisation de maquettes et de prototypes pour aboutir à la réalisation des OPACs de demain. Il serait illusoire de croire que ces évolutions rendront l'ensemble du savoir directement accessible et transparent à l'utilisateur en faisant abstraction des connaissances et des pratiques nécessaires à sa compréhension et à son usage. Du moins peut-on espérer que les catalogues informatisés seront moins opaques, éviteront d'être des barrières supplémentaires dans l'accès aux connaissances et deviendront des systèmes plus Ouverts, Publics, Accessibles, et Conviviaux.

Fin mars 1989