

De quelques éléments à considérer avant de choisir un niveau d'analyse ou un langage documentaire
Factors to consider in selecting a level of document analysis or documentation language
Algunos elementos que merecen consideración antes de escoger un nivel de análisis o un lenguaje documental

Suzanne Bertrand-Gastaldy

Volume 32, Number 1-2, January–June 1986

URI: <https://id.erudit.org/iderudit/1052706ar>

DOI: <https://doi.org/10.7202/1052706ar>

[See table of contents](#)

Article abstract

Information professionals are often placed in a position of choosing a system of document analysis. The final decision must rest upon several interrelating factors. The author outlines these factors as well as their relationship one to another.

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Bertrand-Gastaldy, S. (1986). De quelques éléments à considérer avant de choisir un niveau d'analyse ou un langage documentaire. *Documentation et bibliothèques*, 32(1-2), 3–23. <https://doi.org/10.7202/1052706ar>

Tous droits réservés © Association pour l'avancement des sciences et des techniques de la documentation (ASTED), 1986

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

Érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

<https://www.erudit.org/en/>

De quelques éléments à considérer avant de choisir un niveau d'analyse ou un langage documentaire

Suzanne Bertrand-Gastaldy*

Université de Montréal

Les professionnels de l'information documentaire sont souvent confrontés au choix d'un système d'analyse. La décision qu'ils doivent prendre implique la prise en considération d'un grand nombre d'éléments en interaction. L'auteure passe en revue les principaux éléments à considérer ainsi que leurs interrelations.

Factors to consider in selecting a level of document analysis or documentation language

Information professionals are often placed in a position of choosing a system of document analysis. The final decision must rest upon several interrelating factors. The author outlines these factors as well as their relationship one to another.

Algunos elementos que merecen consideración antes de escoger un nivel de análisis o un lenguaje documental

Los profesionales de la información documental se encuentran a menudo con la necesidad de elegir uno entre varios sistemas de análisis. En la toma de decisión deben tener en consideración innumerables elementos interrelacionados. La autora hace un repaso de los principales elementos que se deben considerar así como sus interrelaciones.

Les personnes chargées de la création d'une banque de données bibliographiques sont confrontées à un grand nombre de décisions, dont le choix du niveau d'analyse et du langage documentaire. Les mêmes questions reviennent presque toujours et le spécialiste, lorsqu'il est consulté, éprouve une certaine gêne à devoir avouer qu'il ne peut fournir de réponse immédiate ni de solution miracle.

Le choix résulte en effet de la prise en compte de nombreux éléments en interrelation à l'intérieur de systèmes et de sous-systèmes complexes et la décision n'est bien souvent que le résultat d'un compromis entre tendances diverses, voire contradictoires.

Nous nous proposons, dans un premier article, de passer en revue les principaux éléments à considérer ainsi que leurs interrelations, tout en mentionnant les données recueillies au cours d'expériences relatées dans la littérature. Dans un second article, nous nous pencherons sur les

caractéristiques des différents types de langages documentaires afin de dégager leur impact sur l'optimisation recherchée.

Une des sources de complexité: le nombre de systèmes et de sous-systèmes en interaction

Trois systèmes en interaction

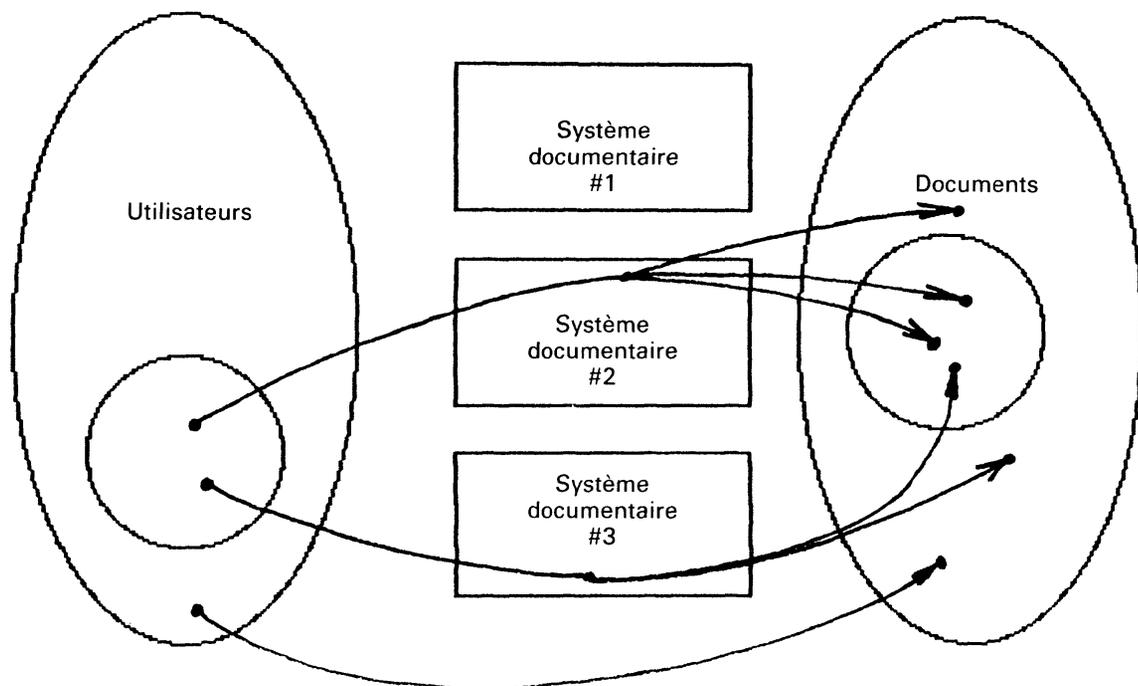
On sait que le but d'un système d'information documentaire est de procurer aux utilisateurs les références des documents (ou les documents eux-mêmes) contenant l'information dont ils ont besoin.

Les opérations effectuées à l'intérieur d'un système documentaire donné visent donc à mettre en relation un sous-ensemble d'utilisateurs avec un sous-ensemble de documents¹ (schéma 1). Ce qui caractérise surtout ces deux sous-ensembles, c'est leur variété et leur complexité.

* L'auteure est professeure agrégée à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal.

1. Tous les utilisateurs ne sont pas obligés de passer par les systèmes documentaires pour obtenir les documents ou les références aux documents.

Schéma 1



Trois types de systèmes en interaction

Mentionnons seulement, pour les documents, la variété des circuits de production et de distribution, la variété des supports, la variété des types de contenu, la variété de l'expression des contenus; pour les utilisateurs, la variété des besoins, des motifs d'utilisation, des niveaux de connaissance, de la fréquence d'utilisation des systèmes documentaires, etc.

Afin de rendre l'information accessible, le système documentaire cherche donc à contrôler cette variété d'une part en sélectionnant, regroupant, condensant, et organisant les documents, d'autre part en trouvant les moyens de standardiser les questions.

C'est ce que l'on représente habituellement dans le schéma de la chaîne documentaire (schéma 2). Le système documentaire y apparaît comme composé de sous-systèmes en interaction étroite. Chacun d'eux comporte des opérations (ou activités) effectuées avec des ressources humaines, financières et matérielles, de façon à transformer les intrants en produits ou extrants qui deviennent eux-mêmes des intrants pour le sous-système subséquent.

Nous verrons surtout, dans ce qui suit, comment, par une série de filtres successifs, on réduit la variété relative au contenu des documents et la variété relative aux questions des utilisateurs, dans les sous-systèmes de l'analyse et de l'exploitation. Nous n'étudierons pas le sous-système de la collecte qui, pourtant, constitue le premier filtre important grâce à la sélection d'un nombre réduit de domaines du savoir, de publications, de supports, en fonction des objectifs poursuivis (schéma 3).

Le sous-système de l'analyse documentaire

Nous laissons de côté le traitement descriptif pour ne nous occuper que du traitement analytique (ou analyse de contenu). Cette opération vise à substituer au texte de départ un texte d'arrivée, plus court et mieux organisé, plus économique à mémoriser et à manipuler. Puisque c'est sur ce nouveau texte que s'effectue la recherche documentaire, on peut comprendre que de la qualité de l'analyse découle en grande partie la qualité du service rendu à l'utilisateur.

Schéma 2

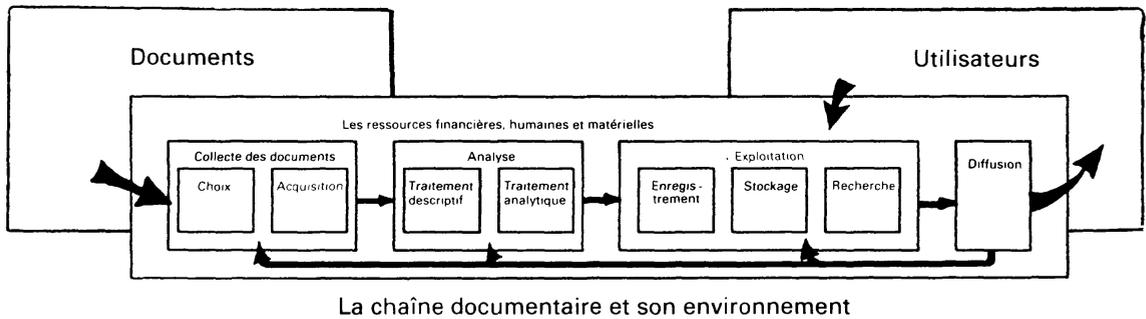
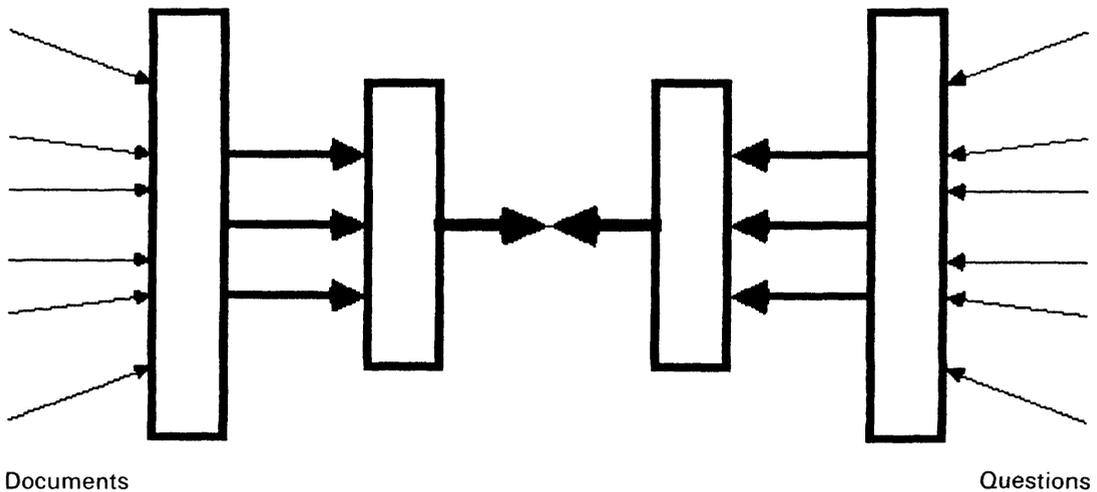


Schéma 3



Réduction de la variété des documents et des questions à l'intérieur du système documentaire

Les opérations:

L'analyse comporte les étapes de reconnaissance et de représentation du contenu.

— Reconnaissance du contenu: les niveaux d'analyse

Cette première opération implique un phéno-

mène de filtration, de mise à l'écart, puisqu'elle est basée sur l'élimination d'informations considérées comme redondantes, inutiles ou peu importantes. Les notions non retenues n'entrent pas dans cette «banque de données où est déposé un capital d'informations sur lequel on peut faire des retraits à la demande»².

Plus on veut représenter le contenu d'un docu-

2. Robert Escarpit, *Théorie générale de l'information et de la communication*, Paris, Hachette, 1976, p. 149 (Langue, linguistique, communication)

ment de façon détaillée, plus la lecture (le visionnement ou l'audition, le cas échéant) doit être complète. On distingue généralement, du niveau le plus superficiel au niveau le plus profond, les opérations suivantes:

- la classification ou catégorisation;
- l'indexation superficielle et l'indexation approfondie;
- la réduction;
- la syntaxe (de plusieurs documents).

On peut également analyser de façon indirecte le contenu, en procédant à une analyse des citations³.

— Traduction du résultat de l'analyse de contenu:

La seconde opération consiste à traduire en symboles et à mémoriser les notions sélectionnées et parfois les relations que ces notions entretiennent entre elles. Le résultat constitue la représentation formalisée du contenu du document (texte d'arrivée ou extrait du sous-système d'analyse)

Outils d'analyse et de traduction:

Les réservoirs de symboles dans lesquels puisent les indexeurs consistent en outils d'analyse qui, selon la nature même des symboles, leur contrôle et la façon dont ils sont coordonnés (avant ou après le stockage), peuvent être soit des tables de classification, soit des plans de classement, soit des répertoires de vedettes-matière, soit des thésauri, soit l'ensemble des termes du langage naturel, avec ou sans syntaxe. Les règles, normes, politiques d'analyse constituent une autre sorte d'instruments d'analyse.

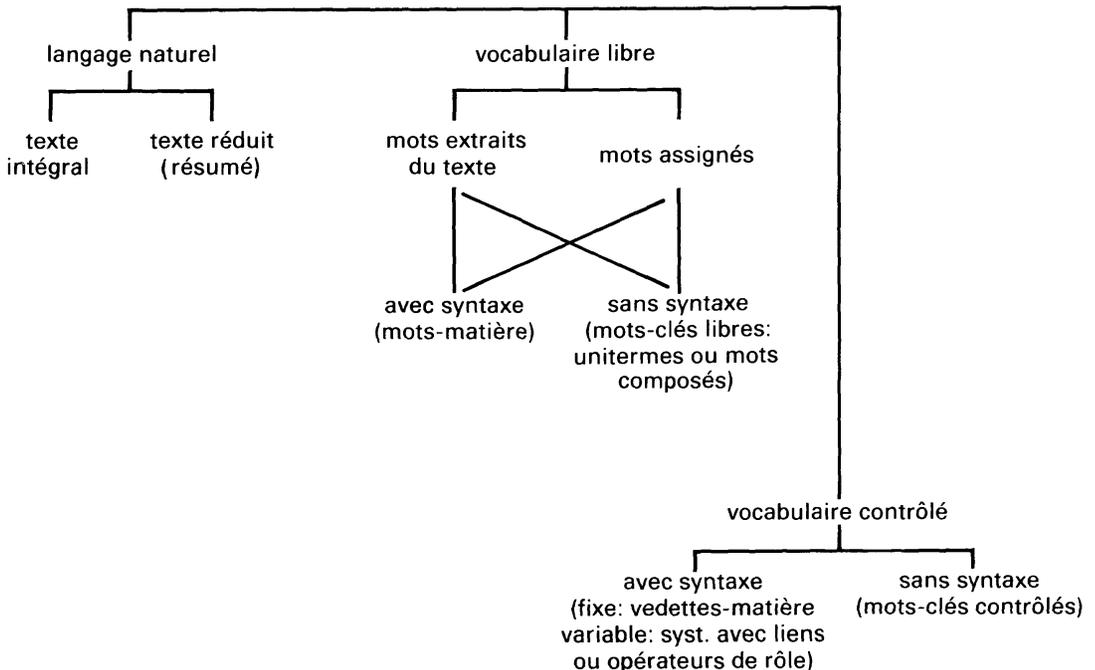
La traduction du contenu s'effectue donc au moyen d'un langage documentaire, qui peut être libre ou contrôlé, précoordonné ou postcoordonné (schéma 4)⁴.

Produits obtenus:

Donc selon le niveau d'analyse, on n'emploie pas les mêmes outils et on n'obtient pas les mêmes représentations formalisées: indices de classification (langage contrôlé et précoordonné);

Schéma 4

Les différents types de langages documentaires



3. Hubert Fondin, «La structure et le vocabulaire de l'analyse documentaire; contribution pour une mise au point», *Documentaliste*, vol. 14 (mars 1977), 11-16

4. Schéma adapté de F.W. Lancaster, «Perspective paper: information science», In Donald E. Walker et al., *Natural Language Information Science: Perspective and Direction of Research*, Stockholm, Skriptor / FID, 1977. p.30.

indexat composé de mots matière (vocabulaire libre et précoordonné) ou de vedettes-matière (vocabulaire contrôlé précoordonné); de mots-clés libres-unitermes ou expressions composées-ou contrôlés (postcoordonnés); et aussi de langage naturel pour le reste.

Il faut remarquer que, tout comme les langues naturelles découpent et structurent différemment notre vision du monde (les couleurs, les liens de parenté, etc.) les langages documentaires découpent et structurent différemment le contenu des documents. L'indice de classification situe le document à l'intérieur d'une discipline et de ses subdivisions; la vedette-matière exprime un sujet (combinaison de concepts), les mots-clés (simples ou composés) traduisent les concepts élémentaires sans indiquer la façon dont l'auteur a choisi de les relier dans le document, le langage naturel peut indiquer des relations complexes entre événements, causes, circonstances de temps et de lieu, etc.

Modalités d'analyse:

Les symboles utilisés pour traduire le contenu des documents peuvent être:

- soit extraits directement de ceux-ci (s'il s'agit de vocabulaire libre ou de langage naturel),
- soit assignés, dans le cas de concepts implicites dans le texte (c'est assez fréquent dans les documents juridiques), de concepts non verbalisés (dans des documents visuels par exemple), ou de notions traduites au moyen d'outils d'analyse.

Pour compléter ce panorama, il faudrait ajouter que les deux étapes de l'analyse (reconnaissance du contenu et traduction) peuvent s'effectuer en mode manuel ou automatique.

Alors que l'analyse «manuelle» faite par les indexeurs peut consister en:

- extraction de mots-clés (unitermes ou expressions composées) du titre ou du texte des documents, sans aucune transformation;
- assignation de mots-clés supplémentaires pour traduire les concepts implicites;
- assignation de langage contrôlé;
- rédaction d'annotations, d'analyses indicatives ou informatives pour compléter les mots-clés et les mettre en contexte,

l'analyse automatique procède surtout par extraction:

- des mots des titres pour constituer des index KWIC et KWOC,
- des mots les plus significatifs des documents (grâce au recours à des méthodes statistiques) que

l'on peut ensuite regrouper par des méthodes de classification automatique (on ne traite en général que les chaînes de caractères, vu les difficultés de la reconnaissance automatique des mots composés);

- des phrases clés qui donnent des extraits;

Mais elle peut aussi consister en la traduction des mots d'un texte en descripteurs d'un thésaurus, grâce à des correspondances établies a priori.

L'analyse manuelle peut être assistée par ordinateur. Cela signifie que l'indexeur choisit ses mots-clés, mais que l'ordinateur lui affiche des mots-clés connexes, la fréquence d'utilisation, ou des documents déjà indexés avec les mêmes mots-clés. L'ordinateur peut aussi intervenir, après l'étape de l'analyse, dans le stockage, la manipulation et l'édition des représentations formalisées choisies par l'indexeur (ex: systèmes d'indexation PRECIS, NEPHIS, POPSI, CIFT, etc.).

Il est important de noter que, selon la procédure, on ne travaille pas sur le même type de représentation, puisqu'on ne manipule pas les mêmes unités d'information: l'extraction automatique s'effectue essentiellement sur les unitermes ou chaînes de caractères, alors que l'assignation humaine peut traduire aisément des concepts, des sujets, des classes à l'intérieur d'une discipline.

Puisque, selon les niveaux, modalités et langages adoptés à l'étape de l'analyse, on ne retient pas le même type d'information, il faut être conscient qu'à l'étape du repérage, on ne pourra pas fournir le même type d'information.

Le sous-système de l'exploitation documentaire

Les opérations d'exploitation comprennent:

- enregistrement, organisation des représentations formalisées (qui sont les intrants de ce sous-système) sous forme d'index, de fichiers inversés et stockage sur supports imprimés ou ordinolingues.

- recherche dans les instruments de repérage dans lesquels ont été enregistrés et stockés les résultats de l'analyse.

La chaîne documentaire s'achève par la diffusion des résultats de la recherche.

Il existe un parallélisme exact entre le processus de traitement analytique et le processus de la recherche documentaire⁵.

5. Chai Kim, *Thesaurus Updating: an Investigation of the Theoretical Criteria for Thesaural Structure Revision in Political Science*, PhD thesis. University of Pittsburg, 1973, p. 4-5.

Traitement analytique

1. L'analyste parcourt le document pour sélectionner un ensemble de mots, d'expressions ou de phrases qui représentent le sujet exprimé par l'auteur.

2. Il décide quels sont les éléments représentatifs du sujet qui valent la peine d'être retenus parce qu'ils correspondent aux capacités du système d'information.

3. Il traduit les éléments retenus en éléments du langage documentaire utilisé dans le système.

Recherche documentaire

1. Le documentaliste parcourt (ou écoute) la question pour sélectionner un ensemble de mots, d'expressions ou de phrases qui représentent le besoin de l'utilisateur.

2. Il décide quels sont les concepts qui valent la peine d'être retenus parce qu'ils répondent au contenu de la question posée.

3. Il transforme les concepts pertinents en termes de recherche normalisés et utilisés dans le système.

Dans un système documentaire, on utilise donc les mots ou symboles suivants:

traitement analytique

1. mots ou symboles du document employés par l'auteur

2. mots ou symboles du *langage documentaire* retenus par l'analyste.

recherche documentaire

3. mots ou symboles de la question employés par l'utilisateur

4. mots ou symboles du *langage documentaire* retenus pour interroger le système

L'analyse et le repérage s'articulent donc autour de ce noyau commun qu'est le langage documentaire. Tout comme le langage naturel, mais d'une façon plus simplifiée, le langage documentaire permet la communication.

C'est en effet grâce à lui que l'on peut réduire et organiser la masse des informations enregistrées dans les documents et la diversité de leur expression d'une part, la diversité des formulations des questions d'autre part.

C'est par l'intermédiaire du langage documentaire que s'établit, avec plus ou moins de bonheur, la communication entre la pensée d'un producteur d'information et celle d'un utilisateur.

Lorsque la recherche s'effectue automatiquement sur les mots des titres, des résumés ou du texte intégral, c'est bien encore de langage documentaire qu'il s'agit: dans la plupart des systèmes documentaires, la question est posée non pas en langage naturel, avec toute la complexité et la diversité (syntaxe de la phrase, structure générique du texte) que cela implique, mais seulement avec les mots du langage naturel reliés par des opérateurs syntaxiques rudimentaires (opérateurs booléens et opérateurs de proximité). Il existe donc, dans ce cas, une différence, dont on ne saisit peut-être pas toujours l'importance, entre le langage de la représentation et le langage de l'interrogation.

Autre lieu de convergence entre analyse et repérage: les supports de recherche. On les choisit principalement en fonction de leur maniabilité, de leur capacité de stockage, de leur coût et de la facilité des mises à jour.

Selon le mode de recherche effectuée, on ne pourra pas utiliser le même langage ni les mêmes supports.

Recherche manuelle et recherche automatisée

La recherche manuelle peut s'effectuer:

— dans des instruments traditionnels, assez rigides au niveau de la manipulation: catalogues imprimés ou fiches traditionnelles de bibliothèques, index de livres ou de périodiques à partir des éléments (assez peu nombreux) mis en vedette. Ce type d'instruments convient bien aux systèmes précoordonnés;

— dans des instruments plus flexibles permettant une plus grande possibilité de combinaison des clés d'accès: fichiers (ou index) à colonnes, fichiers à coïncidence optique, fichiers ordinaux pour les langages postcoordonnés.

Recherche rétrospective et diffusion sélective de l'information

La recherche rétrospective consiste à interroger la base de données sur un ensemble de notices bibliographiques accumulées depuis un laps de temps assez grand (plusieurs années).

Cette recherche, effectuée en une seule fois, est dite en direct ou en conversationnel lorsqu'elle utilise l'ordinateur et la stratégie peut alors être modifiée en cours de route. Afin que l'utilisateur

ne soit pas noyé par le nombre de références, il est nécessaire d'utiliser, pour ce type de recherche, au moins un langage spécifique exprimant des nuances fines.

La diffusion sélective de l'information compare un profil (stratégie de recherche fixe) à un nombre relativement restreint de documents, ceux qui ont été ajoutés à la base de données lors de la dernière mise à jour. Généralement la recherche est faite en différé. Les notices repérées sont signalées régulièrement au demandeur (toutes les semaines, tous les mois). Il est intéressant, pour ce type de recherche, de pouvoir disposer d'un langage peu étendu afin de récupérer au moins quelques références chaque fois.

La taille du corpus soumis à l'investigation a donc un impact direct sur le type de langage à utiliser.

Recherche personnelle et recherche déléguée

La recherche personnelle est menée par l'utilisateur lui-même. Elle nécessite de sa part une bonne connaissance des clés d'accès, du langage documentaire et des commandes du système, s'il est automatisé.

La recherche peut être déléguée à un documentaliste. Le recours à un intermédiaire pose des problèmes supplémentaires de communication, d'interprétation du besoin d'information, mais dispense l'utilisateur d'apprendre les techniques du langage et du logiciel d'interrogation qui peuvent être fort complexes.

Nous pouvons constater combien, à l'intérieur des deux sous-systèmes que nous venons de survoler, les composantes sont multiples et inter-reliées. À la diversité des opérations s'ajoute la diversité des modes d'analyse et de repérage, la diversité des représentations du contenu, des langages, des outils et des supports de recherche.

Tous ces éléments s'influencent mutuellement et le langage documentaire joue un rôle pivot.

Le succès d'un système d'information dépend non seulement de l'harmonisation de tous ces éléments, mais aussi du degré de correspondance entre deux notions fondamentalement opposées:

— l'efficacité maximum de son fonctionnement en termes de rapidité, de facilité («smoothness») et d'économie d'une part,

— et l'efficacité maximum en termes de compatibilité avec les besoins et les préférences de ses utilisateurs d'autre part⁶.

Personnel administratif d'un côté, utilisateurs de l'autre, se dotent de critères pour évaluer le service documentaire en termes de qualité, de temps et de coût. Contraintes des uns, besoins ou exigences des autres doivent être cernés. C'est ce que nous tenterons de faire dans ce qui suit.

Critères de qualité

Les performances du système, rappel et précision, sur lesquelles nous reviendrons à propos des usagers, sont déterminées en premier lieu par l'exhaustivité de l'analyse et la spécificité du vocabulaire d'indexation⁷. Nous examinerons donc ces critères ainsi que d'autres critères tout aussi importants.

Qualité de l'analyse

L'exhaustivité (ou la profondeur) de l'analyse

L'exhaustivité de l'analyse correspond au nombre de renseignements ou de mots-clés retenus pour rendre compte du contenu d'un document⁸.

Même une bonne politique de développement peut être annihilée si l'indexation ne représente

6. S. Herner et K.J. Snapper, «The application of multiple-criteria utility theory to the evaluation of information systems», *Journal of the American Society for Information Science*, vol. 29, no. 6 (November 1975), 289.

7. John Balnaves, «Is PRECIS worth the price?», *Lasie (Library Automated System Information Exchange)*, vol. 5, no. 1 (1974), 4.

8. Cyril Cleverdon, «On the inverse relationship of recall and precision», *Journal of Documentation*, vol. 28, no. 3 (September 1972), 195-201.

et:

Maron, «Depth of indexing», *Journal of the American Society for Information Science*, vol. 30 (1979), 224-228.

pas adéquatement le contenu des documents qui seront alors inutilisés⁹.

On peut se demander s'il existe un niveau optimum d'exhaustivité.

En comparant le repérage à partir du titre, du résumé et du texte intégral, Cleverdon¹⁰ concluait que les titres n'étaient pas assez exhaustifs (7 termes) et que les résumés l'étaient trop (60 termes). Le seuil se situait quelque part entre 22 et 60 termes et tout ce que l'on pouvait dire c'est que 33 termes donnaient de meilleurs résultats que 22 ou 60.

Mais comment se fait-il que certains index n'attribuent qu'une demi-douzaine de mots-clés par document?

En fait, ce n'est pas tant la quantité d'éléments retenus qui importe que leur qualité. L'indexation avec 7 termes du titre donne de meilleurs résultats qu'avec 25 termes pris au hasard¹¹, ce qui plaiderait en faveur de l'indexation humaine laissant place au jugement, contrairement à l'indexation automatique ou au repérage en vocabulaire libre sur le texte intégral.

En réalité, il n'existe pas de seuil immuable, car celui-ci varie en fonction de l'environnement du système¹². Font partie de cet environnement:

- le langage documentaire qui sert à transcrire les concepts isolés au moment de l'analyse. À quoi bon faire une analyse exhaustive si le seul outil dont on dispose est un système de classification, un répertoire de vedettes-matière ou un thésaurus restreint?

- le support d'enregistrement et de stockage. La recherche manuelle ne peut fonctionner si l'analyse est trop approfondie; c'est le contraire pour une recherche automatisée. Certains déplorent qu'«on continue d'indexer, à l'âge de l'ordinateur, avec la même superficialité que si le seul et éternel support de la représentation des unités documentaires était la fiche»¹³.

- les ressources du service d'analyse: plus l'analyse est profonde, plus elle prend de temps, plus elle demande d'expertise, donc plus elle coûte cher.

- les besoins de l'utilisateur. Si les questions sont exhaustives, l'analyse doit l'être également. Mais, dans les systèmes postcoordonnés, plus l'indexation est approfondie, plus le bruit augmente à cause des fausses liaisons entre des termes trop nombreux, et plus on risque de décevoir l'utilisateur en lui signalant des documents qui ne traitent que très superficiellement des concepts qu'il recherche.

Puisque l'exhaustivité de l'analyse devrait être déterminée en fonction de l'exhaustivité des questions, Lancaster suggère de procéder ainsi¹⁴: en phase d'implantation on retient une série de questions réellement posées par les utilisateurs (50 par exemple) et on fait varier l'exhaustivité de l'indexation, les stratégies de recherche, le langage d'indexation pour une collection-test, puis on relève les performances des requêtes (taux de rappel et de précision). On remarquera qu'on atteindra un certain seuil (point d'inflexion). Par exemple, pour une collection-test donnée, 15 mots-clés par document entraîneront un taux de rappel de 73%, 20 mots-clés élèveront le rappel à 90%, mais il faudra aller jusqu'à 35 mots-clés pour atteindre 95%. 20 mots-clés dans ce cas-là pourraient alors constituer un bon compromis.

L'exactitude de l'analyse

L'exactitude de l'analyse correspond à la fidélité avec laquelle est traduit le contenu du document. Elle détermine la confiance de l'utilisateur dans le système.

L'exactitude peut être trahie au moment du choix des concepts, à cause d'une connaissance insuffisante du sujet de la part de l'indexeur ou d'un manque d'objectivité, à cause de la difficulté de lecture du document et aussi au moment de la traduction en langage documentaire, de même

9. Jean de Bonville, «L'efficacité du système de repérage dans la perspective de la désuétude documentaire», *Argus*, vol. 10, no 1 (janvier-février 1981), 17-18.

10. Cyril W. Cleverdon, «On the inverse relationship...»

11. Elaine Svenonius, *Effect of Indexing Specificity on Retrieval Effectiveness*, Ph.D. Dissertation, University of Chicago, 1971.

12. K. Sparck Jones, «Does indexing exhaustivity matter?», *Journal of the American Society for Information Science*, vol. 24, no.5 (1973), 313

13. Jean de Bonville, «L'efficacité du système de repérage...», 17.

14. D'après Donald W. King et al., «Evaluation of indexing processes and indexing language», *In the Evaluation of Information Services and Products*, Washington, Information Resources Press, 1971, p. 154-155.

qu'au moment de l'enregistrement (fautes d'orthographe).

La cohérence de l'analyse

La cohérence de l'analyse, bien qu'elle ne soit pas un gage d'exactitude, est fortement liée à l'exactitude et à l'avantage de se mesurer plus facilement¹⁵.

La cohérence correspond à la capacité chez plusieurs indexeurs, ou chez un même indexeur à différents intervalles, d'analyser un même document de la même façon. En somme la cohérence de l'indexation mesure la similarité de réactions de plusieurs individus — ou d'un même individu — traitant la même information¹⁶.

On distingue plusieurs sortes de cohérence et il existe toute une panoplie de formules pour les mesurer (on peut en retrouver un bon nombre dans la thèse de Leonard¹⁷):

— la cohérence entre plusieurs indexeurs. On la mesure en faisant indexer un même document par plusieurs indexeurs¹⁸, ou par les indexeurs et le réviseur, avec les mêmes outils¹⁹.

— la cohérence d'un même indexeur. On la mesure en lui faisant indexer un même document ou une partie de document à son insu, après suffisamment de temps pour que le phénomène de rétention consciente de l'information ne joue pas²⁰.

— Certains auteurs étudient la cohérence de l'indexation non pour l'ensemble de ce processus mais pour chacune des étapes du processus:

1. cohérence dans le choix des concepts

2. cohérence dans le choix de la terminologie apte à traduire ces concepts²¹.

— la cohérence entre indexation et recherche. Si pour traduire un concept, 30% des indexeurs emploient tel terme et que 30% des usagers utilisent le même terme pour le même concept, cette cohérence sera parfaite. En augmentant la cohérence des indexeurs, on risque cependant de provoquer simultanément une chute du taux de cohérence avec la recherche et les performances du système diminueront. Il existe en effet une relation étroite entre cohérence de l'indexation et efficacité de la recherche. C'est seulement si l'analyse est cohérente que l'utilisateur peut prévoir la façon dont l'information a été entrée dans l'index.

Comment garantir une bonne cohérence?

D'après les expériences, les facteurs suivants affectent la cohérence:

— facteurs environnementaux comme la chaleur, le bruit;

— facteurs pragmatiques liés à la personnalité de l'indexeur, à son état d'esprit, à sa formation et à son expérience;

— facteurs sémantiques liés au document («langage» du document — l'image est plus difficile à saisir que l'écrit²² —, vocabulaire de l'auteur, lisibilité du texte, domaine, spécificité du discours, style de l'auteur et clarté de la présentation, redondance de l'information, longueur du document) d'une part; facteurs sémantiques liés à

15. Ibid., p. 138.

16. Pranas Zunde et Margaret Dexter, «Indexing consistency and quality», *American Documentation*, vol. 20, no.3 (July 1969), 259.

17. Laurence Edwards Leonard, *Interindexer consistency and retrieval effectiveness: measurement of relationships*, Ph. D. Dissertation, Urbana, Ill., University of Illinois, Urbana Champaign, 1975.

18. Frances J. Hurwitz, «A study of indexer consistency», *American Documentation*, vol. 20, no.1 (January 1969), 92.

19. Johanne J. Herr, «Use of data base access for interindexer communication and for indexer training», *Proceedings of the American Society for Information Science: the Information Conscious Society*, 33rd Annual Meeting, Philadelphia, October 11-15, 1970. Washington, D.C.: ASIS, 1970, p. 164.

20. W.R. Mullison et al. «Comparing indexing efficiency, effectiveness, and consistency with or without the use of roles», *Proceedings of the American Society for Information Science: Cooperating Information Societies*, 32nd Annual Meeting, San Francisco, Cal., October 1-4, 1969. Westport, Conn., Greenwood Pub. Corp., 1969, p. 301-302.

21. Barbara M. Preschel, «Improved communication between information centers through a new approach to indexer consistency», *Communication for Decision-Makers; Proceedings of the American Society for Information Science 34th Annual Meeting*, Denver, November 7-11, 1971., Westport, Conn. Greenwood Pub. Corp., 1971, p.363.

et

Daniel Tarr et Harold Borko, «Factors influencing inter-indexer consistency», *Information Utilities: Proceedings of the 37th Annual Meeting of the American Society for Information Science*, Atlanta, October 13-17, 1974. Washington, D.C., ASIS, 1974, p.50.

22. Karen Markey, «Interindexer consistency test: literature review and report of a test of consistency in indexing visual materials», *Library and Information Science Research*, vol. 6, no.2 (April-June 1984), 155-177.

l'analyse d'autre part (spécificité du langage rendant trop ténue la différence entre deux mots-clés, contraintes imposées à l'indexeur dans le choix des concepts et de la terminologie, aides à l'indexation).

Les responsables ne devraient donc rien négliger pour offrir aux indexeurs:

— le plus d'instruments susceptibles de les guider dans leur travail. Une grille d'analyse soigneusement conçue et des politiques claires minimisent les divergences d'approches et d'interprétations. La possibilité de consulter la base de données au cours de l'indexation ou de la révision est recommandée, de même que la fourniture d'une liste de termes reliés lors des indexations antérieures (concurrence des mots-clés).

— un environnement favorisant la concentration intellectuelle

— au moins un langage documentaire contrôlé restreint.

Si l'analyse est décentralisée, l'importance de la révision est indéniable et il faudrait y affecter les ressources nécessaires.

En effet, bien que l'étude de Léonard ait montré que la révision de l'indexation n'avait qu'un effet minime, il faut interpréter les résultats avec prudence et tenir compte du contexte. À l'inverse, certains soutiennent que la révision serait le facteur pouvant créer les plus grandes différences dans le repérage²³.

Qualité du langage documentaire

Puisque la qualité de l'analyse est tributaire du langage documentaire, penchons-nous maintenant sur les indicateurs de qualité de celui-ci.

La spécificité (ou précision) du langage

Les indexeurs doivent, en principe, indexer les

documents à leur niveau exact de précision. Dès 1876, Charles Ammi Cutter avait émis, pour l'attribution des vedettes-matière, la règle suivante: un livre sur les chats doit être décrit par le mot «Chat» et pas par «Zoologie. Mammifères» et il ajoutait: «à moins qu'il n'y ait pas de termes suffisamment précis, auquel cas il faut recourir à un terme plus générique.»

Par spécificité on entend la mesure selon laquelle chaque symbole du langage documentaire constitue une étiquette précise et exacte pour la classe de documents à laquelle il a été attribué²⁴.

La spécificité est aussi comprise comme la capacité d'un terme à discriminer un petit nombre de documents²⁵.

La spécificité du langage documentaire est donc étroitement liée à son étendue. Plus les descripteurs sont nombreux, plus l'indexeur a de la chance de pouvoir représenter des nuances fines. Sinon, en l'absence d'un descripteur adéquat, même s'il reconnaît un concept spécifique, il sera obligé d'indexer à un niveau de généralité plus élevé et trahira le texte, sans que son jugement soit en cause.

Deux comparaisons illustreront l'importance du vocabulaire pour la qualité de l'analyse²⁶.

— une thèse de Yassin Ahmed Mostapha ayant reçu pour unique vedette-matière tirée de la liste de la Library of Congress: «Hydraulics», aurait pu être indexée, à l'aide du *Thesaurus of Engineering and Scientific Terms (TEST)*, par «Flow», «Channel flow», «Open channel flow», «Roughness», ou par les descripteurs du *Water Resources Thesaurus: «Roughness coefficients», «Roughness (Hydraulics)», «Head loss»*²⁷.

— Si un langage contient les trois termes suivants, du plus général au plus spécifique: «Métal», «Acier» ou «Chrome», l'indexeur aura le choix d'attribuer celui qui représente le mieux le contenu du document. Cependant, s'il ne contient que «Métal», l'indexeur sera obligé d'adopter ce

23. Frances J. Hurwitz, «A study of indexer consistency»..., 92-95.

24. John Balnaves, «Is PRECIS...», 4.

25. K. Sparck Jones, «Does indexing...», 313.

26. Hans Wellish, «Subject retrieval in the seventies. Methods, problems, prospects», *In Subject Retrieval in the Seventies: Proceedings of an International Symposium Held at the center of Adult Education, University of Maryland, College Park, May 14-15, 1971.*

et

Donald W. King, «Evaluation of indexing processes...», p. 150.

27 *Ibid.*, p. 152.

terme qui ne sera pas forcément représentatif du contenu réel²⁸.

À la limite, seul le langage naturel, avec sa palette infini de mots, permet de traduire les nuances les plus subtiles de la pensée.

Les exigences de spécificité varient:

— selon le type de documents analysés: les monographies, qui épuisent l'ensemble des facettes d'un même sujet, sont indexées à un niveau général, alors que les articles qui traitent de questions plus pointues nécessitent un vocabulaire plus spécifique. On attribue donc généralement des indices de classification ou des vedettes-matière aux premières et des descripteurs tirés de thésauri spécialisés aux seconds²⁹.

— selon le niveau d'analyse qui devrait varier selon le niveau des questions: l'étude du niveau de spécificité requis est en effet indissociable de celle qui est proposée pour l'exhaustivité. Dans la pratique, il faut accorder le niveau de spécificité du langage au niveau de spécificité des requêtes. L'idéal serait que le système puisse répondre à la fois à des demandes générales et à des demandes très spécifiques.

Un examen minutieux d'un échantillon de questions posées au fichier test permet de déterminer si, à l'hétérogénéité des questions, correspond effectivement une grande variation dans le niveau de spécificité du vocabulaire employé.

L'indexation des documents à partir de différents niveaux d'analyse aide à examiner les performances de chaque langage ou de chaque combinaison de langages et à voir s'il est bien nécessaire d'adapter à la fois des clés d'accès générales et des clés d'accès spécifiques.

L'étendue (ou ampleur) du vocabulaire

L'étendue du vocabulaire correspond au nombre de mots-clés disponibles pour traduire les concepts couverts par le corpus documentaire.

L'étendue du vocabulaire est liée à sa spécificité. Elle est également fonction de l'exhaustivité (ou profondeur) de l'analyse (pour une analyse

superficielle de 3 à 5 mots clés, le langage documentaire nécessaire sera restreint). Elle dépend aussi de la couverture de la banque de données (interdisciplinaire ou spécialisée).

L'objectivité du langage

Le choix des termes de même que le choix des relations entre ceux-ci sont extrêmement importants pour rendre avec exactitude et objectivité le contenu d'un document.

On a pu montrer comment les vedettes-matière de la Library of Congress trahissaient les préjugés à l'égard des minorités ethniques, des femmes, des marginaux, des religions, etc.³⁰.

Le regroupement de certains sujets sous un même indice classificatoire peut révéler bien des préjugés d'école, de même que l'inégalité dans le développement des indices pour certains sujets. On a aussi étudié le sexisme dans le système de classification Dewey³¹.

Donc, avec un langage qui manque d'objectivité, l'exactitude de l'analyse n'est qu'un leurre.

Les quelques critères de qualité énumérés ci-dessus pourraient être regroupés de façon à correspondre à deux des axiomes de la théorie de l'indexation proposée par Fugmann³²: la *fidélité* de la représentation (assurée essentiellement par l'exhaustivité, l'exactitude, la spécificité, l'étendue et l'objectivité) et la *prévisibilité*, garantie surtout par la cohérence (en plus du contrôle du vocabulaire).

Les ressources humaines, financières et matérielles

En principe, puisque la qualité du repérage dépend de la qualité de l'analyse et du langage, les administrateurs sont conscients de la rentabilité des investissements (en temps, en personnel et en équipement) qu'il faut y consentir.

Temps requis pour l'analyse

L'analyse est l'opération de la chaîne documentaire qui prend le plus de temps: de l'ordre de 60%

28. M. Bloomfield, «Evaluating indexing 5. Discussion and summary», *Special Libraries*, vol. 62, no. 2 (February 1971), 94.

29. M. Bloomfield, «Evaluating indexing 2. The simulated machine indexing experiments», *Special Libraries*, vol. 61, no.9 (November 1970), 509.

30. Jessica L. Milstead Harris et Doris H. Clack, «Treatment of people and peoples in subject analysis», *Library Resources and Technical Services*, vol. 23, no. 4 (Fall 1979), 374-390.

31. Sanford Berman, *The Joy of Cataloguing: Essays, Letters, Reviews, and other explosions*, Phoenix, Orynx Press, c1981.

32. Robert Fugmann, «On the practice of indexing and its theoretical foundations», *International Classification*, vol.7, no. 1 (1980), 13-20.

pour l'indexation et de 10% pour la gestion d'un thésaurus³³.

Lors d'une étude effectuée dans plusieurs établissements sur la durée d'indexation, Cleverdon a constaté qu'il fallait 20 minutes pour indexer des rapports techniques et que 90% des termes d'indexation étaient attribués en 80% du temps³⁴. Les derniers 20% du temps ne sont donc pas très productifs et il est inutile d'aller au-delà d'un certain seuil d'exhaustivité.

Ces chiffres doivent être considérés en valeur relative, puisque tout dépend du contexte, en particulier:

- du type de documents analysés:

les monographies et les articles ne requièrent pas le même niveau d'exhaustivité et de spécificité, avons-nous mentionné;

la littérature grise, les rapports de recherche, sont moins facilement accessibles, donc nécessitent une analyse plus approfondie (de l'ordre de plusieurs heures, parfois) pouvant se substituer au texte original;

les documents audiovisuels que l'on ne peut visionner directement méritent le même soin;

- de la présentation du document: il est facile d'analyser les documents structurés toujours de la même façon, comme les brevets (LANDAU, 1969, p.306);

- des outils d'analyse: de leur complexité, de la façon dont ils sont consultables par les indexeurs (l'affichage sur écran facilite la tâche);

- de la clarté des politiques d'analyse.

Niveau d'expertise du personnel

L'analyse étant une opération d'interprétation, l'indexeur doit:

- bien comprendre le document, donc être spécialisé dans le domaine;

- bien connaître les techniques documentaires;

- bien connaître les besoins des usagers.

Il va de soi que plus la base est spécialisée, plus l'expertise des utilisateurs est grande, plus l'analyse est approfondie, plus le langage documentaire est étendu, alors plus l'indexeur doit être qualifié.

En pratique, on cherche à engager les gens qui coûteront le moins cher. Peut-on engager des indexeurs non-spécialistes? ou des spécialistes du domaine non formés aux techniques d'indexation? Pour régler ce vieux débat, on a proposé d'essayer, pendant la phase d'expérimentation, des gens d'expérience et de formation différentes afin de juger des résultats³⁵. Peut-être vaut-il mieux une approche par essais et erreurs pendant un court laps de temps qu'une mauvaise décision définitive.

En tout cas, une bonne analyse des tâches s'impose afin d'aboutir à une solution judicieuse qui ne gaspille pas les ressources:

- L'élaboration de politiques d'analyse, d'outils documentaires, la révision sont les opérations les plus délicates.

- Les tâches de saisie et d'enregistrement ne requièrent pas le même niveau de formation.

Une chose est certaine: il faut minimiser l'intervention humaine qui coûte le plus cher³⁶. C'est pourquoi on cherche à automatiser certaines tâches et à transférer certaines opérations intellectuelles vers le repérage, donc vers les utilisateurs.

Ressources matérielles

La question que l'on se pose le plus souvent est celle de l'équipement informatique. La taille du corpus, son rythme de croissance, le genre de produits et de services offerts nécessitent-ils le recours à un ordinateur? À quel type d'ordinateur? micro- ou mini-? Envisage-t-on la recherche automatique seulement? Veut-on procurer à l'uti-

33. J.S. Sessions, «The cost and costing of information storage and retrieval», *Information Storage and Retrieval*, vol. 6, no. 2 (June 1970), 165.

34. Herbert B. Landau, «The cost analysis of document surrogation: a literature review», *American Documentation*, vol. 20, no. 4, (October 1969), 304.

35. Donald W. King et al., «Evaluation of indexing processes ...», p.153-155.

36. L.H. Campey, «Costs of producing KWIC/KWOC indexes», *Information Storage and Retrieval*, vol. 10, no. 9-10 (September-October 1974), 293-307.

et:

Julius Frome, «Amortization of indexing input costs», *Journal of Chemical Documentation*, vol. 12, no.2 (May 1972), 98-101.

lisateur des outils d'aide au repérage, comme des thésaurus de recherche? Veut-on gérer automatiquement les outils d'analyse, le thésaurus par exemple? Fera-t-on de l'indexation assistée par ordinateur? Ou même de l'indexation automatique?

L'indexation automatique fait l'objet de nombreuses recherches. Les expériences de laboratoire relatives par Stibic donnent une idée des économies possibles. On a inclus divers degrés d'analyse:

- a- extraction des mots
- b- troncature à droite
- c- isolation des mots-outils et de quelques chaînes de caractères alphanumériques et numériques
- d- division des mots composés avec trait d'union
- e- reconnaissance des mots composés
- f- consultation du thésaurus
- g- pondération
- h- création d'un vecteur de documents et de fichiers inversés.

On a pu conclure que l'indexation automatique coûtait trois fois moins cher que l'indexation manuelle (qui prenait 6 minutes par document, dans le contexte étudié) si on utilisait toutes les options énumérées, à l'exception de d et e. Effectuée sur les chaînes de caractères seulement, elle coûtait 5 fois moins³⁷.

Besoins des utilisateurs

Les utilisateurs sont souvent négligés dans la conception des systèmes d'information. Ils cons-

tituent une source additionnelle de variété exigeant une flexibilité qui s'oppose à la tendance conservatrice, simplificatrice et rigide de l'organisation administrative³⁸.

Tout un courant de réflexion tend à remettre en cause les modèles stéréotypés dans lesquels les bibliothèques ont enfermé le savoir³⁹ et les utilisateurs (l'usager du catalogue de la Bibliothèque du Congrès est perçu comme un «American/Western European, Christian, White, Heterosexual, and Male»⁴⁰).

On prend peu à peu conscience que l'information n'a de valeur que relative, selon les objectifs poursuivis par l'individu, l'état de ses connaissances, ses capacités de traiter les données, de les mémoriser, de les intégrer à sa propre structure cognitive.

Cependant même si l'on reconnaît désormais la nécessité de tenir compte des besoins des utilisateurs, de leur fonctionnement individuel, il est encore très difficile de le faire en pratique, parce que d'une part les variables humaines individuelles sont nombreuses mais n'ont parfois que peu d'influence prises isolément et qu'on dispose seulement de modèles très lacunaires⁴¹. On peut glaner çà et là des renseignements utiles et le nombre d'études sur le sujet tend à augmenter. On cherchera à la fois du côté de la psychologie cognitive et du côté des études sur l'utilisateur final.

Les utilisateurs «naïfs», comme on les appelle parfois, commencent à bénéficier de l'attention des spécialistes de la science de l'information. Bon nombre des exposés prononcés lors des congrès Online y sont désormais consacrés (45 sur un total de 101, en 1984, à New York)⁴². Des statistiques datant déjà de quelques années révélaient qu'ils ne formaient que 5% de la clientèle des systèmes bibliographiques⁴³. À la fin des années 70, on pensait que l'avenir immédiat ne laissait pas

37. V. Stibic, «Remarks on the economic feasibility of automatic indexing». In *EURIM 11*, European Conference On Applied Research for Information Services and Libraries, London, Aslib, 1977, p. 130-133.

38. David Smith, *Systems Thinking in Library and Information Management*, New-York, London, K.G. Saur, Clive Bingley, 1980, p.63.

39. D. Stratfield, «Moving towards the information user: some research and its implications», *Social Science Information Studies*, vol.3 (1983), 223-240.

40. Joan K. Marshall, comp., *On Equal Terms: A thesaurus for Nonsexual Indexing and Cataloging*, New-York, Neal Schuman Publishers Inc., 1977

41. Ray Fidel et Dagobert Soergal, «Factors affecting online bibliographic retrieval: a conceptual framework for research», *Journal of the American Society for Information Science*, vol. 34, no.3 (May 1983), 163-180.

42. Ibid.

43. M.L. Moureau, «Problems and pitfalls in setting up and operating an on line information service», *Online Review*, vol. 2, no. 3 (1978), 243.

prévoir de changement notable⁴⁴. L'avènement du vidéotex et de l'ordinateur personnel a accéléré l'évolution. Pour qui veut s'emparer du marché, il est nécessaire de connaître les besoins.

Variété des comportements

Les besoins peuvent en effet varier:

— selon les individus, en fonction de leur expérience de l'interrogation des systèmes documentaires, de leur connaissance du domaine ou du contenu de la base, de leur familiarisation avec le langage documentaire. Un novice doit être guidé par étape, un utilisateur occasionnel a besoin d'une assistance minimum qui lui rafraîchisse la mémoire, un habitué se débrouille seul⁴⁵.

— pour un même individu, selon que le sujet cherché est au centre de ses préoccupations ou appartient à un secteur marginal (l'utilisateur peut, en effet, être novice dans certains domaines et expert dans d'autres)⁴⁶ et selon le temps: à la longue, il deviendra habile, mais s'intéressera au même fichier, au même système.

La prise en compte du grand public et de l'utilisateur final entraînera la simplification des systèmes. On pense en particulier que, n'ayant pas la même motivation que les clients habitués des systèmes bibliographiques, le grand public sera moins tolérant qu'eux et ne sera pas prêt à investir autant de temps, d'argent et d'effort pour récupérer un nombre médiocre de références pertinentes⁴⁷.

Il semble bien que l'utilisateur final soit moins patient que le bibliothécaire⁴⁸. Par exemple, toute démarche intermédiaire entre l'expression du besoin et l'accès au document, comme celle qui consiste à consulter d'abord une liste bibliographique, le conduit à une frustration. Si la frustration est trop grande (difficulté d'utiliser le produit: commandes complexes, langage documentaire contraignant ou lourd; dispositif de communication des documents inefficace et inapproprié; piètres résultats du repérage), il rejettera le système d'information et préférera s'adresser à des collègues⁴⁹. De façon générale, il faudra se souvenir, si l'on veut offrir un accès direct à l'utilisateur final que:

Les études réalisées sur le comportement des scientifiques et d'autres professionnels en matière de recherche d'information ont constamment démontré que l'accessibilité et la facilité d'usage sont les facteurs déterminants du choix d'une source d'information. En général, on choisit la source d'information la plus commode, que cette source soit perçue ou non par l'usager comme la plus complète, la plus sûre ou, de quelque manière, la meilleure⁵⁰.

C'est le comportement du débutant qui est le plus difficile à cerner, car contrairement à celui des utilisateurs confirmés, il n'est absolument pas structuré⁵¹. Son problème, mal défini en général, s'adapte mal à un système très rigide⁵². N'étant pas au courant des nombreuses règles suivies par les indexeurs⁵³, et ne voulant pas l'être la plupart du temps, il risque de poser des questions trop triviales pour la banque de données.

44. F. Gebhardt et I. Stellmacher, «Desing criteria for documentation retrieval languages», *Journal of the American Society for Information Science*, 29 (1978), 191-199.

45. A. Avner et H. G. Friedman, «Interacting with computer users: desing considerations» *In Clinic on Library Applications of Data Processing*, University of Illinois 80, Proceedings: Public Access to Library Automation, Illinois, University of Illinois, Graduate School of Library and Information Science, 1981, p.16.

46. John Bennett, «Ease of use as an acceptance factor in interactive search system?», *In Charles T. Meadow, ed., Forum on Interactive Bibliographic Systems*, Oak Ridge, Tennessee, U.S. Atomic Energy Commission, 1971, p. 166.

et:

P. Leggate *et al.*, *The On Line Search Guide Project: a Prototype for a Mini-Computer Band Aid to the Construction of Search Profiles for Natural Language Data Bases*, Oxford, Engl., Cairns Library, March 1976, p.11.

47. Ann H. Schabas, «Optimizing Telidon tree structures / Maximiser les structures arborescentes de Télidon» *In Comptes rendus de la 9e Conférence annuelle de l'Association Canadienne des Sciences de l'Information*, Pointe-au-Pic (Québec), 27-30 mai 1981, ACSI, 1982, p.176.

48. G. K. Ostrum et D.K. Yoder, «Chemists as end-user searchers — Training and follow up», *In 9th International Online Information Meeting*, London, 3-5 December 1985, p. 280.

49. R.R. Blanken et B.T. Stern, «Planning and design of on-line systems for the ultimate user of biomedical information», *Information Processing and Management*, vol. 11 (1975), 208.

et:

Christine Snow, «Architects wants and needs for information, demonstrated through a university-based information service», *Aslib Proceedings*, vol. 27, No. 3 (March 1975), 112.

50. F. W. Lancaster, *Principes directeurs pour l'évaluation des systèmes et services d'information*, Paris, Unesco, 1978, p. 17.

51. Thomas P. Moran, «An applied psychology of the user; guest editor's introduction», *Computing Surveys*, vol. 13, no. 1 (March 1981), 4.

52. David A. Thompson, «Interface design for an interactive information retrieval system: à literature survey and a research system description», *Journal of the American Society for Information Science*, (November-December 1971), 362.

53. Robert Fugmann, «The glamour and the misery of the thesaurus approach», *International Classification*, vol. 1, no. 2 (1974), 76.

On a pu observer que le débutant commence souvent ses explorations avec les classes les plus générales et descend l'arbre hiérarchique du système de classification jusqu'au degré de spécificité requis⁵⁴. Au fur et à mesure qu'il prend de l'expérience, ses questions deviennent plus spécifiques. Cette démarche est cependant mise en doute⁵⁵ et il faudra encore beaucoup de recherches interdisciplinaires, notamment en psychologie ergonomique⁵⁶, avant qu'on puisse être affirmatif sur le comportement documentaire de l'utilisateur final et du grand public ainsi que sur le type d'assistance qu'on doit leur fournir.

Quelles sont les implications de ces constatations au niveau de l'analyse et du langage documentaire?

On a remarqué que, lorsqu'on met sur pied des stratégies de formation des utilisateurs, on a l'habitude d'identifier plusieurs types d'utilisateurs, alors qu'on ne retrouve pas cette distinction au moment où l'on conçoit l'interface homme-système (le langage documentaire est un élément de cette interface). Si la clientèle est hétérogène, il est impossible de trouver une solution unique et il faudrait, idéalement, une interface pour chaque catégorie d'utilisateurs⁵⁷.

De même qu'on conclut qu'il n'est pas raisonnable de concevoir un système documentaire impliquant une tâche aussi complexe que la recherche bibliographique⁵⁸, pour un utilisateur moyen, de même il faudra prévoir plusieurs niveaux d'analyse et un ensemble de langages qui assureraient le maximum de flexibilité et préserveraient la liberté d'action de la clientèle, surtout si celle-ci est hétérogène⁵⁹.

It is suggested that 'User-Friendly' is not simply a way of describing systems orientated towards novices but that truly User-Friendly system must be friendly to all types of user at all level of sophistications⁶⁰.

Tolérance face au délai de signalement

Le délai de signalement peut dépendre de la

discipline dans laquelle oeuvre l'utilisateur, de la nature des documents qu'il veut consulter ou du type de recherche qu'il effectue.

Par exemple, si la durée d'une invention est courte, un fichier sur les brevets devra minimiser le délai.

Pour les films et vidéogrammes, dans certains secteurs c'est non seulement l'information qui devient vite désuète, mais aussi et surtout l'image qui la véhicule.

En ce qui concerne la diffusion sélective de l'information:

L'utilisateur attend d'un service signalétique courant qu'il l'informe des dernières nouveautés, «en rapport avec ses intérêts professionnels courants. Le système d'information doit donc veiller à fournir fréquemment et régulièrement des informations qui soient aussi à jour que possible»⁶¹. Une enquête à propos de 11 services de diffusion sélective nous donne une idée des exigences des usagers: ceux-ci, conscient des facteurs retardant le signalement (opérations de la chaîne documentaire précédant l'enregistrement et le stockage) acceptent en moyenne 6 semaines de délai.

Plus l'analyse sera approfondie, contrôlée, exhaustive et homogène, plus le langage documentaire sera complexe à appliquer et à mettre à jour, plus le délai de signalement sera long et plus certains utilisateurs seront insatisfaits.

En ce qui concerne la recherche rétrospective:

L'utilisateur qui demande une recherche bibliographique exhaustive est habituellement quelqu'un qui travaille à un projet de recherche relativement long. La célérité de la réponse n'est en général pas ce qui compte le plus pour lui (sauf s'il y a une date limite au-delà de laquelle les résultats de la recherche perdraient beaucoup de leur valeur); il est prêt à attendre plus longtemps pour obtenir une documentation complète (la rapidité compte beaucoup moins que l'exhaustivité)⁶².

54. David A. Thompson, «Interface design ...», 172.

55. O. Riondet, «Grand public et bases de données de presse», *Documentaliste*, vol. 20, no 6 (novembre-décembre 1983), 188-192.

56. Bisseret, «Pour une psychologie ergonomique des systèmes documentaires», *Documentaliste*, vol. 20, no 1 (janvier-février 1983), 6-10.

57. M. Goldstein et W.H. Ford «The user-cordial interface», *Online Review*, vol. 2, no. 3 (September 1978), 269.

58. John Bennett, «Ease of use ...», p. 166-167.

59. R. R. Blanken et B. T. Stern, «Planning and design ...», 209.

60. F. W. Lancaster, *Principes directeurs ...*, p.17.

61. S. Herner et K. J. Snapper, «The application of multiple-criteria utility theory ...», 293.

62. F. W. Lancaster, *Principes directeurs...*, p. 17, 19.

Types de produits et de supports documentaires désirés

Les représentations formatisées du contenu des documents peuvent être enregistrées sur support papier pour une recherche manuelle ou en mémoire d'ordinateur pour une recherche automatique, on l'a vu.

À cause de la difficulté de manipulation des imprimés, de temps et du coût d'édition, on n'enregistre généralement que les résultats d'une analyse superficielle dans un bulletin bibliographique et l'on se contente de langages documentaires contrôlés relativement restreints: les notices sont rangées selon un plan de classement rudimentaire permettant le bulletinage, les analyses sont signalétiques plutôt qu'informatives, les index ne comportent que les mots-clés principaux (de trois à cinq) ou les vedettes-matière (une ou deux).

Avec les fichiers automatisés on peut offrir une analyse plus approfondie. Grâce à la puissance et à la vitesse de combinaison des mots-clés, on peut en attribuer davantage, une dizaine ou plus. La recherche en langage naturel sur les titres ou sur de plus grandes portions de textes, comme les analyses informatives, devient possible, à condition qu'on dispose d'un bon logiciel d'interrogation.

Donc un même service d'analyse voulant mettre sur le marché toute une gamme de produits devra pour chacun adopter le niveau d'analyse et le langage les plus appropriés. À titre d'exemple, chaque document d'ERIC était affecté d'une dizaine de descripteurs tirés du thésaurus et d'un identificateur principal interrogeables en direct⁶³. Seuls les descripteurs principaux sont reproduits dans les bulletins imprimés. INSTEC (Information Services in Physics, Electrotechnology, Computers and Control) utilise simultanément des langages contrôlés et libres dont certains sont réservés au service de diffusion sélective, à la recherche rétrospective ou aux bulletins, mais tous sont intégrés.

Performances de la recherche bibliographique

Mesures qualitatives: rappel et précision:

Les principales mesures qualitatives du succès de la recherche sont au nombre de deux:

«1. L'utilisateur obtient-il ou non ce qu'il cherche?

2. Avec quel degré de précision et d'exhaustivité l'obtient-il?»⁶⁴

La première question se pose lorsque l'utilisateur est à la recherche d'une référence ou d'un fait précis, unique. Les critères majeurs d'évaluation sont alors l'exhaustivité et la validité de la réponse. Nous ne nous attarderons pas à cet aspect.

La deuxième s'applique «lorsqu'il faut évaluer une recherche bibliographique pour laquelle il n'existe pas de réponse correcte type⁶⁵». Il faut alors recourir à d'autres mesures de qualité: le rappel et la précision surtout.

Le taux de rappel est une mesure de l'exhaustivité de la recherche. Il donne le pourcentage des documents retrouvés par rapport au nombre total de documents pertinents contenus dans le système. «...lorsqu'on parle d'un rappel à 80%, cela signifie que 80% des documents pertinents (contenus dans la base de données) ont été retrouvés⁶⁶». On est alors face à un silence de 20% (pourcentage de documents pertinents non repérés).

Le taux de précision correspond au pourcentage de documents pertinents repérés par rapport au nombre total de documents retrouvés par le système. «Une recherche bibliographique qui repère, par exemple, 50 documents dont 10 sont jugés pertinents par l'initiateur de la recherche se verra attribuer pour cette opération un taux de précision de 10/50, soit 20%⁶⁷». Le bruit (pourcentage de documents non pertinents repérés) sera de 80%.

Les exigences en rappel et précision sont fonction de la discipline, des utilisateurs et de leur mode de recherche.

Pour un utilisateur chevronné désirant rédiger un état de la question, pour un inventeur visant à déposer un brevet, pour un médecin qui cherche une contre-indication à une chimiothérapie, par exemple, les conséquences d'un taux de rappel inférieur à 100% sont graves⁶⁸. Ce type de recherche a pour but de confirmer qu'il n'existe aucune documentation sur un sujet particulier. Une fois la preuve acquise, l'inventeur aura la confirmation que son invention est brevetable, le médecin l'assurance qu'il pourra prescrire son traitement sans danger, le scientifique qu'il pourra proposer un nouveau projet de recherche.

63. Judith Yarborough, «How to prepare for a computer search of ERIC: a non-technical approach», *Information Reports and Bibliographies*, vol. 4, no. 5 (1975), 15-17.

64. F. W. Lancaster, *Principes directeurs...*, p. 19.

65. *Ibid.*, p. 42.

66. *Ibid.*, p. 19.

67. *Ibid.*

68. R. R. Blanken et B. T. Stern, «Planning and Design ...», 208.»

Au contraire, un étudiant de niveau secondaire ou de premier cycle universitaire sera pleinement satisfait de quelques documents très pertinents. Dans ce cas,

Le taux de rappel absolu ne signifie rien. Par exemple, cet usager peut trouver trois ouvrages (sur un total de 18 contenus dans la collection), qui correspondent à ce qu'il cherche et le satisfont entièrement. Il ne s'inquiètera pas outre mesure de fait qu'il n'a pas eu accès aux quinze autres livres, à moins que les références qui lui ont échappé ne soient d'une manière quelconque «meilleures» que celles qu'il a trouvées (c'est-à-dire) plus directement pertinentes ou plus récentes)⁶⁹.

Rares sont les utilisateurs qui se préoccupent d'un taux de rappel de 100%, pensent certains auteurs⁷⁰. Mais peut-être faut-il se méfier des intentions qu'on leur prête: d'après une enquête menée auprès de chimistes, on a découvert que, contrairement aux hypothèses, 22% voulaient une recherche exhaustive⁷¹.

Donc, dans une recherche rétrospective, les experts qui exigent un taux de rappel élevé, c'est-à-dire qui ne tolèrent pas de silence ou en tolèrent un peu, auront peut-être la motivation nécessaire pour s'accompagner d'un bruit important.

Au contraire l'utilisateur peu spécialisé ne sera pas prêt à investir beaucoup de temps pour trier les références pertinentes. Il éprouvera un sentiment de frustration si le bruit est élevé.

Il convient de noter, en passant, que le taux de bruit ne dépend pas seulement de la tolérance de l'usager dans l'absolu, mais bien du nombre de références repérées. Une chose est de devoir éliminer 3 documents sur 10, une autre 30 documents sur 100, bien que le taux reste constant: 30%.

Après des tentatives pour éliminer du processus d'évaluation des systèmes documentaires les éléments aussi subjectifs que la pertinence (notamment en faisant la distinction entre la pertinence d'un document repéré et la pertinence du terme de la question pour traduire le contenu de ce document) on constate désormais la nécessité de tenir compte de la valeur que les utilisateurs accordent à l'information fournie:

Although value cannot be quantified within the

information system this is not sufficient for putting it beyond the concern of the information profession....

It can also be seen to follow, that just as the value of information is a quality of the users systems, the property that is known as relevance within the information system is also really a users' measure. That is to say: what is 'relevant' in the information system has 'value' in the users' system⁷².

En fait la pertinence permet de mesurer si le système atteint son objectif: fournir à l'utilisateur l'information dont il a besoin.

En quoi le rappel et la précision dépendent-ils de l'analyse du langage documentaire?

Ils dépendent directement de l'exhaustivité et de la spécificité, comme nous l'avons mentionné plus haut. Lors de l'évaluation de systèmes d'information, on combine les mesures de rendement de la recherche avec l'étude des échecs en matière de rappel et de précision.

L'analyse des échecs comporte elle-même l'examen de chaque document en cause, celui de l'enregistrement de l'indexage de ces documents, l'étude des demandes qui ont suscité les recherches, celle des stratégies de recherche, de la terminologie du système, et les estimations de pertinence de la part des usagers. L'examen de chacun de ces éléments doit permettre de déterminer quel composant du système a été principalement responsable des échecs constatés. Les principales catégories d'échecs sont imputables à l'indexage, à la stratégie de recherche, à la terminologie, ou à une interaction défectueuse entre l'usager et le système.⁷³

On voit que le langage documentaire à lui seul peut intervenir dans trois de ces quatre catégories, puisque la qualité de l'indexation dépend en partie de lui (capacité de traduire fidèlement les concepts retenus et leurs relations), la qualité de la stratégie de recherche aussi (capacité de suggérer les mots exacts et les mots reliés).

Il faut cependant reconnaître que certains facteurs ne dépendent pas de l'étape d'analyse, mais d'étapes antérieures comme le choix de documents. En effet, même si le langage documentaire permettait un rappel de 100%, une couverture

69. W. F. Lancaster, *Principes directeurs...* p.44

70. Elaine Svenonius et Helen F. Schmierer, «Current issues in the subject control of information», *Library Quarterly*, vol. 47, no. 3 (1977), 333.

71. G. K. Ostrum et D. K. Yoder, «Chemist as end-users ...», commentaire oral au cours de la conférence.

72. David Smith, *Systems Thinking...*, p. 79.

73. W. F. Lancaster, *Principes directeurs...*, p.23.

inadéquate du fichier fournirait un faible pourcentage de documents pertinents existants susceptibles de satisfaire les besoins des utilisateurs. C'est pourquoi l'usager qui désire effectuer une recherche rétrospective complète se préoccupe non seulement du taux de rappel (qui est une mesure de l'exhaustivité de la recherche) mais aussi de la couverture de la base de données (qui est une mesure de l'exhaustivité de la collection).

Mesures quantitatives: temps et coût

Avec les systèmes interactifs, les mesures de précision et de rappel ne sont plus tellement significatives. En effet, lors d'une recherche directe, «l'usager mène sa propre recherche, et il formule des jugements de pertinence au fur et à mesure de sa progression...»⁷⁴. On peut «supposer que si l'usager consentait à travailler assidûment et durant assez longtemps, il finirait par découvrir tous les documents pertinents existant dans le «système» et aurait alors une réussite égale à 100%»⁷⁵.

C'est pourquoi Lancaster propose de

remplacer la mesure de précision par le calcul du temps requis de l'usager pour effectuer sa recherche. Le temps pris par le lecteur est ensuite divisé par le nombre total de références jugées pertinentes afin de déterminer le coût unitaire (en temps) de chaque référence pertinente repérée. La durée de la recherche fournit une mesure plus adéquate de l'effort demandé à l'usager»⁷⁶.

Mais le comportement de l'usager à l'égard du coût est très peu étudié.⁷⁷

En quoi l'analyse et le langage documentaire peuvent-ils affecter le temps requis?

Si le vocabulaire contrôlé est utilisé, il faudra tenir compte du temps nécessaire à sa consultation au moment où on établira la stratégie et chaque fois qu'on la modifiera. Dans le cas d'un langage restreint, le rappel risque d'être plus élevé que la précision et il faudra longuement trier les références. Inversement, plus le langage sera complet, plus la précision sera grande, mais plus la manipulation du langage et l'interrogation du fichier seront longues.

Avec le vocabulaire libre, il faudra tenir compte de la complexité des stratégies pour éliminer les problèmes de synonymie, de polysémie et d'homographie, et des tâtonnements inévitables si aucune aide à la recherche n'est fournie. La longueur du texte sur lequel sera faite la recherche aura une influence sur le taux de bruit. Il en sera de même de sa structure syntaxique. Plus celle-ci sera libre, plus elle contiendra de périphrases, de verbes, de locutions prépositives, plus la recherche sera laborieuse et le bruit élevé.

On constate, en tout cas, que toute implantation d'un système d'analyse devrait être précédée d'une étude des attentes de la clientèle et de son comportement. Cette clientèle est généralement assez méconnue et mériterait qu'on lui offre des mécanismes plus souples pour satisfaire ses divers besoins documentaires.

Le corpus documentaire

Le corpus documentaire, qui n'est qu'un sous-ensemble des documents disponibles dans le domaine, présente des traits importants pour le choix du niveau d'analyse et du langage documentaire. D'ailleurs les besoins des usagers peuvent être appréhendés en partie par le biais des documents.

En effet en se renseignant sur le type de documents dont les utilisateurs se servent, sur ceux qu'ils produisent, sur les caractéristiques qu'ils retiennent pour mener à bien leur travail, sur la structuration de l'information à l'intérieur de ces documents, sur la nature et le rythme d'évolution des concepts, du vocabulaire et de la discipline, les responsables de l'analyse documentaire seront plus à même de prendre des décisions adéquates pour établir la grille d'indexation (type de concepts retenus), le niveau d'exhaustivité, le type de langage documentaire, etc.

Nature des documents selon le type de contenu

Le type de contenu peut être un facteur important pour le choix d'un système d'analyse et d'un langage. Nous avons mentionné plus haut que la cohérence de l'indexation était plus difficile à obtenir avec des images. En effet «La difficulté à

74. W. F. Lancaster, *Principes directeurs...*, p. 45.

75. *Ibid.*, p. 44.

76. André Cossette. «Evaluation de l'efficacité de la bibliothèque: analyse des études majeures», *Documentation et bibliothèques*, vol. 24, no 3 (septembre 1978), 124.

77. Ray Fidel et Dagobert Soergel, «Factors affecting ...», 165.

rendre compte du contenu de l'image tient à sa polysémie, à son caractère extra-linguistique et à la multitude de ses usages latents ...»⁷⁸. L'utilisation d'une grille d'analyse construite à partir des découvertes de la sémiologie pallierait en partie cet inconvénient. On s'est aperçu que les classifications traditionnelles étaient tout à fait inadéquates pour ce genre de documents:

Plusieurs classes ou divisions du savoir ne se prêtent pas aux représentations figurées: philosophie, religion ... Les classifications de livres accordent une place prépondérante aux sujets généraux, alors que les images prises isolément doivent être décrites en termes spécifiques. ... Enfin les systèmes traditionnels sont conçus sur le principe de la verbalisation et du développement séquentiel des idées. Dans le cas de l'image, la classification doit reposer sur le contenu visuel et non sur le contenu intellectuel supposé.⁷⁹

Au contraire, «l'indexation répond parfaitement au problème de la polysémie de l'image et assure souplesse et précision dans la recherche.»⁸⁰

Dans un même domaine du savoir, on distinguera plusieurs types de documents selon leur contenu: en droit, par exemple, on reconnaît la doctrine, la jurisprudence et la théorie.

En examinant les documents consultés et les documents produits par les architectes, dans les différentes phases de leur travail, il est possible de reconnaître une grande variété: catalogues de manufacturiers, manuels de génie, documents juridiques, etc. d'une part, esquisses, plans, devis, correspondance échangée en cours de réalisation de mandat, etc. d'autre part⁸¹.

On peut dégager de chacun de ces types de documents une structuration différente de l'information qui devrait aboutir à des politiques d'analyse différentes. C'est ainsi qu'une analyse du contenu des jugements québécois et canadiens a permis de discerner une structuration tripartite qu'il faudrait respecter dans la rédaction des résumés de jurisprudence⁸².

Avec l'interdisciplinarité, la diversité des documents utilisés pour remplir une mission est de plus en plus grande, et leur provenance est de plus en plus dispersée et hétérogène. Il n'est pas rare que l'utilisateur doive consulter à la fois des coupures de presse, des rapports de recherche, des décisions juridiques, des brochures, des comptes rendus de conférences dans un grand nombre de disciplines (les domaines de la jeunesse, des droits de la personne, de l'aphasie sont quelques illustrations de cette interdisciplinarité).

Nature du domaine

En ce qui concerne le domaine, il est possible de distinguer entre disciplines dures et disciplines molles.

Sont appelées disciplines dures des disciplines qui ont un vocabulaire stable et non ambigu, comme les mathématiques et les sciences naturelles⁸³ et, de façon générale, tous les domaines scientifiques et techniques⁸⁴, avec cependant quelques nuances.

On entend par discipline molle, une discipline dans laquelle les concepts sont mal fixés et qui n'a pas de frontière bien délimitée.

Les indices de dureté d'une discipline sont variés et on a même suggéré qu'ils pouvaient se trouver dans la qualité des rapports sociaux que les scientifiques entretiennent entre eux⁸⁵. Certains de ces indices sont susceptibles d'être pertinents pour le choix d'un langage documentaire, notamment le contrôle et l'évolution de la terminologie ainsi que, dans une certaine mesure, le taux d'obsolescence des documents.

Le rythme d'évolution du vocabulaire dépend à la fois de l'application de nouveaux termes à de vieux concepts et de l'addition de nouveaux concepts ou de la disparition d'anciens concepts. On n'a cependant que des données partielles sur cette question. On sait, par exemple, que les concepts n'évoluent pas rapidement en physique⁸⁶ et que le vocabulaire de la science de l'information change de 4% par an à peu près⁸⁷.

78. Maryse Pierrard-Binie, «Analyse documentaire de l'image fixe: état de la question» *Argus*, vol. 10, no 5 (septembre-octobre 1981), 83.

79. *Ibid.*, 84.

80. *Ibid.*, 85.

81. Lucie Gendron et Jeanne Bazinet, «L'information et les architectes», *Argus*, vol. 12, nos 3-4 (mai-août 1983), 75-81.

82. Diane Poirier, *Pour des résumés adéquats de jurisprudence québécoise et canadienne: une étude du document jurisprudentiel, de sa structure, de ses citations, de son rôle et de sa spécificité*, Mémoire de maîtrise, Université de Montréal, 1985.

83. A. Ghose et A. S. Dhawle, «Between traditional classification and coordinate indexing», *Information Processing and Management*, vol. 15, no. 1 (1979), 214.

84. K. Bhattacharyya, «The effectiveness of natural language in science indexing and retrieval», *Journal of Documentation*, vol. 3, no. 3 (September 1974), 239.

85. Norman W. Storer, «The hard sciences and the soft: some sociological observations», *Journal of the Medical Library Association*, vol. 51, no. 1 (1967), 75-84.

86. K. Bhattacharyya, «The effectiveness...».

87. Jessica L. Harris, «Terminology change: effect on index vocabularies», *Information Processing and Management*, 15 (1979), 77-88.

La stabilité du vocabulaire tient en partie au contrôle exercé par la communauté scientifique sur l'utilisation des termes connus et sur l'adoption de nouveaux termes (les physiciens semblent accepter un mot et un seul pour un nouveau concept).

Un rythme rapide de changement peut être compensé par une bonne structuration de l'information: en zoologie, par exemple, ou cette structure du type genre-espèce, les mises à jour des nomenclatures doivent être fréquentes, mais ne posent pas de problème particulier.

Inversement, un domaine dont l'organisation logique n'est pas très solide peut être doté d'un vocabulaire relativement stable. Les études d'Aberystwyth sur la science de l'information prouvent que le vocabulaire naturel sans contrôle est très satisfaisant lors de l'interrogation, alors que dans le domaine de l'aérodynamique, les études de Cranfield avaient révélé qu'un contrôle des synonymes et des terminaisons donnait de meilleurs résultats que le langage naturel⁸⁸.

Quelles sont les conséquences au niveau de l'analyse et des langages documentaires?

D'une part, plus une discipline est dure, plus il est facile de constituer un thésaurus⁸⁹, puisque les termes spécifiques sont déjà normalisés. Mais aussi plus une discipline est dure, moins il est nécessaire d'établir des outils de contrôle du vocabulaire. En effet, les termes étant non ambigus, les synonymes étant peu nombreux, la structure logique de l'information étant évidente, les stratégies de recherche seront faciles à établir et l'interrogation en langage naturel donnera de bons résultats⁹⁰.

Dans les sciences humaines, on emploie souvent des mots de tous les jours avec des significations spéciales mais non normalisées. Il faudrait alors faire appel à des intervenants de formation variée. En effet seule une lecture multiple, selon des points de vue différents, pourrait permettre de saisir l'ensemble des concepts explicites et implicites présents dans un document et de le traduire sous une forme qui respecte leur signification technique précise souvent éloignée de l'usage courant. On aura plus de difficulté à élaborer un vocabulaire contrôlé en science politique ou en sociologie. Les spécialistes ne seront pas d'accord

entre eux et il serait vain de prétendre établir une hiérarchie stricte entre les termes. Une solution préconisée parfois consisterait à les associer par groupes sémantiques après les avoir soumis à des utilisateurs⁹¹. Cependant, en l'absence de tout contrôle du vocabulaire, on obtiendra des taux de rappel et de précision très faibles lors des interrogations.

D'autre part il ne faudrait pas se contenter de l'approche-sujet, mais faire ressortir les différentes facettes du contenu, y compris l'approche idéologique ou théorique lorsque cela peut être utile aux usagers⁹².

Enfin, plus le domaine est dur, plus la demi-vie des documents est longue (en chimie et en physique la demi-vie est plus longue qu'en biologie) et moins le nombre de documents publiés est important⁹³. En conséquence, il vaudra d'autant plus la peine d'analyser en profondeur — et donc de choisir un langage qui permettra cette profondeur — les documents d'une discipline dure. Il sera rentable, aussi, d'établir des politiques et procédures d'analyse pour assurer un bon contrôle et une bonne homogénéité de l'analyse.

L'histoire, qui est une discipline molle, est un cas intéressant. Dans ce domaine les documents ne deviennent jamais périmés, à proprement parler. Pourtant le vocabulaire change, les concepts également. Un contrôle strict est impossible. Au fil des années et des siècles, la façon de désigner les indiens s'est modifiée, les changements de frontières et de régimes politiques ont affecté les dénominations de pays et, si l'on veut pouvoir repérer seulement les documents pertinents pour une période précise, il va falloir garder les différentes terminologies. Mais, pour assurer un bon rappel lors d'une recherche diachronique, il est indispensable d'établir un réseau de renvois.

Conclusion

En somme, le choix d'un niveau d'analyse et d'un langage est une décision qui s'appuie sur un réseau complexe de facteurs. On voit bien qu'il ne peut être question de solution unique, de recette miracle, mais que la prise de décision doit toujours être précédée d'une analyse de système, dans un contexte particulier. La solution adoptée impli-

88. K. Bhattacharyya, «The effectiveness...», 250-251.

89. K. Bhattacharyya, «The effectiveness...», 250-251.

90. Jason Farradane et D. Thompson, «The testing of relational indexing procedures by diagnostic computer programs», *Journal of Information Science*, vol. 2, no. 6 (December 1980), 287.

91. P. Leggate et al., *The On Line Search Guide ...*

92. A. Ghose et A. S. Dawle, «Between traditional classification ...», p.7

93. D. F. Swift *et al.*, «'Aboutness' as a strategy for retrieval in the social sciences», *Aslib Proceedings*, vol. 30, no. 5 (May 1978), 182-187.

quera la plupart du temps la recherche d'une certaine flexibilité et d'un équilibre entre des éléments conflictuels. Des tests seront nécessaires en phase d'implantation pour mesurer l'interdépendance de tous les éléments. On pourra aussi, dans la prise de décision, se référer à des expériences déjà réalisées dans d'autres contextes et à l'avis des experts.

Les critères de qualité de l'analyse, des langages et de la recherche documentaires sont bien connus car étudiés depuis longtemps. Il en va de même des contraintes administratives. Tous ces éléments sont étroitement interreliés. De plus ils ne peuvent être envisagés dans l'absolu car ils dépendent avant tout de la nature des documents et des besoins des utilisateurs. Il est clair que ce qui importe le plus pour que le système documentaire remplisse ses objectifs, c'est de parvenir à

maîtriser ces deux composantes externes. Mais sur ces deux composantes et sur leur interaction (interaction que le système d'information a précisément pour but de favoriser), nous savons bien peu de choses. Il reste à comprendre, entre autres:

- comment sont structurés les textes,
- comment extraire et représenter le sens,
- comment les usagers attribuent un sens et une valeur aux informations qui leur sont fournies.

Vaste programme qui va bien au-delà des stéréotypes dans lesquels on a souvent confiné ces deux éléments. Il appartient tout à la fois à la recherche théorique et à l'expérimentation de faire progresser nos connaissances en ce domaine.



de **A** à **Z**

Nous comblons tous vos besoins d'abonnements

LES SERVICES D'ABONNEMENT
CANEBSCO
 SIX BOUL. DESAULNIERS
 SUITE 308
 ST. LAMBERT, QUE J4P 1L3
 (514) 672-5878
 Ligne directe pour Québec:
 (800) 361-7322

SERVICE AU QUÉBEC:

- Centre de traitement a St.-Lambert
- Equipe spécialisée
- Banque de données
- Avis professionnels
- Représentant des ventes