

1984

24

UPERIEUR DE BIBLIOTHECAIRE

MEMOIRE DE FIN D'ETUDES

Michel NETZER

ETUDE PRELIMINAIRE EN VUE DE LA CONSTITUTION
D'UN THESAURUS COMMUN A LA BIBLIOTHEQUE DE
L'E.N.S.B., A LA BIBLIOTHEQUE DE L'I.N.T.D.
ET AU CENTRE DE DOCUMENTATION DE LA
D.B.M.I.S.T.

ANNEE : 1984

20^{ème} PROMOTION



ECOLE NATIONALE SUPERIEURE DES BIBLIOTHEQUES

17-21, Boulevard du 11 Novembre 1918 - 69100 VILLEURBANNE

Je tiens à remercier tout particulièrement
Madame Madeleine Wagner et Monsieur Jean-Louis
Taffarelli, Conservateurs à l'E.N.S.B.,
Madame Catherine Lermyte, Bibliothécaire-documentaliste
à l'I.N.T.D.,
et le personnel du Centre de documentation de la
D.B.M.I.S.T.
pour la gentillesse et la compétence avec lesquelles
ils ont répondu à toutes mes questions.

DIPLOME SUPERIEUR DE BIBLIOTHECAIRE

MEMOIRE DE FIN D'ETUDES

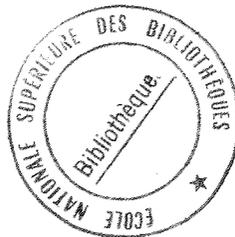
Michel NETZER

ETUDE PRELIMINAIRE EN VUE DE LA
CONSTITUTION D'UN THESAURUS COMMUN A
LA BIBLIOTHEQUE DE L'E.N.S.B., A LA BIBLIOTHEQUE
DE L'I.N.T.D. ET AU CENTRE DE DOCUMENTATION
DE LA D.B.M.I.S.T.

Mémoire présenté sous la direction de

*Madame Madeleine WAGNER
et Monsieur Jean-Louis TAFFARELLI*

1984
24



ANNEE : 1984

20ème PROMOTION

ÉCOLE NATIONALE SUPERIEURE DES BIBLIOTHEQUES
17-21, Boulevard du 11 Novembre 1918 - 69100 VILLEURBANNE

II

NETZER (Michel) - Etude préliminaire en vue de la constitution d'un thésaurus commun à la Bibliothèque de l'E.N.S.B., à la Bibliothèque de l'I.N.T.D. et au Centre de documentation de la D.B.M.I.S.T. : mémoire/présenté par Michel Netzer. - Villeurbanne : Ecole Nationale supérieure de bibliothécaires, 1984. - 57 p. ; 30 cm.

Mémoire E.N.S.B. : Techniques documentaires : Villeurbanne : 1984

Thésaurus, sciences de l'information

Sciences de l'information, thésaurus

Recherche documentaire, Ecole nationale supérieure de bibliothécaires

Recherche documentaire, Institut national des techniques de la documentation

Recherche documentaire, Direction des bibliothèques, des musées et de l'information scientifique et technique

Les possibilités d'accès par la recherche documentaire aux fonds de la Bibliothèque de l'E.N.S.B., de la Bibliothèque de l'I.N.T.D. et du Centre de documentation de la D.B.M.I.S.T. pourraient être améliorées par la réalisation d'un thésaurus commun aux trois organismes. Ce mémoire s'efforce de justifier ce projet et de montrer ce que pourraient être les caractéristiques formelles d'un tel thésaurus, son mode d'élaboration et son exploitation.

S O M M A I R E

	PAGES
INTRODUCTION	1 - 2
I - JUSTIFICATION DU PROJET DE THÉSAURUS COMMUN AUX BIBLIOTHÈQUES DE L'E.N.S.B., DE L'I.N.T.D. ET DE LA D.B.M.I.S.T.	3 - 17
1 - POURQUOI UN THÉSAURUS ?	4 - 10
a - LA SITUATION ACTUELLE EN MATIERE DE RECHERCHE DOCUMENTAIRE A L'E.N.S.B., A L'I.N.T.D. ET A LA D.B.M.I.S.T. ..	4
b - LE THESAURUS COMPARE AUX AUTRES TYPES DE LANGAGES DOCUMENTAIRES,	6
c - LE THESAURUS ET L'ORDINATEUR	9
2 - POURQUOI UN THÉSAURUS COMMUN AUX TROIS BIBLIOTHÈQUES ?	11 - 17
a - AVANTAGES	11
b - CONDITIONS DE POSSIBILITE	12

II - LES CARACTERISTIQUES FORMELLES DU THÉSAURUS	18	-	33
1 - LES UNITES LEXICALES	19	-	25
a - REGLES DE SENS	19		
b - REGLES DE FORME	22		
LES DESCRIPTEURS	22		
LES IDENTIFICATEURS	25		
2 - LA STRUCTURE DU THÉSAURUS	25	-	29
a - LES RELATIONS ENTRE DESCRIPTEURS,...	25		
b - LA STRUCTURE D'ENSEMBLE,.....	26		
3 - LE MODE DE PRÉSENTATION DU THÉSAURUS ..	30	-	33
a - L'INTRODUCTION	30		
b - LES PARTIES DU THESAURUS	30		
LE DICTIONNAIRE CONCEPTUEL	31		
LES ARBRES HIERARCHIQUES	32		
L'INDEX PERMUTE	33		
III - ÉLABORATION DU THÉSAURUS ET POLITIQUE			
D'INDEXATION	34	-	52
1 - RECHERCHE DE THÉSAURUS EXISTANTS	35	-	38
2 - LE THÉSAURUS DE L'UNESCO ET LE THÉSAURUS			
DES SCIENCES DE L'INFORMATION DU C.N.R.S	38	-	44
3 - MODE D'ÉLABORATION DU THÉSAURUS	45	-	48
4 - POLITIQUE D'INDEXATION	48	-	52

CONCLUSION 53

BIBLIOGRAPHIE 54 - 57

I N T R O D U C T I O N

L'Ecole nationale supérieure de bibliothécaires et l'Institut national des techniques de la documentation sont deux établissements que rapproche leur vocation à former, l'une le personnel scientifique des bibliothèques, l'autre les cadres de l'information et de la documentation. L'E.N.S.B. dépend de la Direction des bibliothèques, des musées et de l'information scientifique et technique au Ministère de l'Education Nationale. L'I.N.T.D. est un institut du Conservatoire national des arts et métiers, établissement créé sous l'autorité du ministre de l'Education nationale. L'E.N.S.B., l'I.N.T.D. et la D.B.M.I.S.T. ont en commun de disposer tous trois d'une bibliothèque ou d'un centre de documentation spécialisés dans le même champ du savoir - un domaine difficile à étiqueter, celui de l'information, de la communication, de la documentation, des bibliothèques.

Or ces trois bibliothèques - fait paradoxal étant donné les missions des organismes dont elles dépendent - n'offrent pas actuellement à leurs utilisateurs de système de recherche documentaire vraiment satisfaisant. Il se trouve que le Directeur des bibliothèques a émis la suggestion et le souhait que s'instaure une coopération entre son centre de documentation et les bibliothèques de l'E.N.S.B. et de l'I.N.T.D. Cette coopération pourrait précisément se concrétiser dans la constitution d'un système de recherche documentaire commun aux trois bibliothèques, dont le noyau serait en l'occurrence un thésaurus, outil particulièrement efficace et approprié à un fonds documentaire spécialisé.

Ce mémoire souhaite poser les bases indispensables à la réalisation d'un tel thésaurus. Il ne constitue pas en lui-même le thésaurus, mais une réflexion préalable. Il ne s'agit pas d'une réflexion théorique sur la manière d'élaborer un thésaurus en général : il s'agit d'examiner le cas précis d'un thésaurus destiné aux bibliothèques de l'E.N.S.B., de l'I.N.T.D. et de la D.B.M.I.S.T., de se demander à quelles conditions ce thésaurus doit répondre, quelles pourraient en être les caractéristiques formelles, quelle démarche il faudra suivre pour l'élaborer. Bref, il s'agit de poser quelques garde-fous, de présenter les diverses possibilités et de faire quelques propositions : l'important n'est pas que ces propositions soient suivies mais qu'elles puissent aider dans leur tâche ceux qui élaboreront le thésaurus.

I

JUSTIFICATION DU PROJET DE THÉSAURUS COMMUN
AUX BIBLIOTHEQUES DE L'E.N.S.B., DE L'I.N.T.D.
ET DE LA D.B.M.I.S.T.

I - POURQUOI UN THÉSAURUS ?

a - LA SITUATION ACTUELLE EN MATIÈRE DE RECHERCHE DOCUMENTAIRE DANS LES BIBLIOTHÈQUES DE L'E.N.S.B., DE L'I.N.T.D. ET DE LA D.B.M.I.S.T.

Comme il a été dit en introduction, aucun des systèmes utilisés dans les trois bibliothèques qui nous intéressent ne permet une recherche documentaire qui soit vraiment satisfaisante, tant en matière de livres qu'en matière de périodiques. La situation se présente de manière diverse.

L'E.N.S.B. possède pour les livres un catalogue alphabétique des auteurs, un catalogue alphabétique des matières (1) (selon la norme NF Z 44070) et un catalogue topographique (la bibliothèque a un classement systématique propre). D'autre part, les acquisitions françaises de l'E.N.S.B. (livres et périodiques) ont fait l'objet, à partir de Janvier 1981, d'une indexation de la part des enseignants. Ont été indexés en langage libre d'une part les livres français, d'autre part les articles de périodiques français jugés les plus intéressants, au fur et à mesure qu'ils étaient acquis par la bibliothèque. Les notices des documents (signalement et mots-matières) étaient systématiquement enregistrés en mémoire d'ordinateur (d'abord au centre de calcul de Lyon, puis au C.N.U.S.C. (2) à Montpellier) et donnaient lieu à une édition bimestrielle (listing d'ordinateur avec un index des descripteurs) arrêtée en Mars 1983. Il existe également un index général qui fait correspondre à chaque descripteur les numéros identifiant

(1) Pour remédier aux quelques incohérences que peut présenter le choix des vedettes, le conservateur responsable de la bibliothèque, *Madame WAGNER*, a constitué petit à petit un fichier de contrôle : index où figurent les descripteurs et les renvois

(2) Centre national universitaire *Sud* de calcul

les documents indexés au moyen de ce descripteur . La consultation de cette petite mémoire documentaire est, du fait de la simplicité de l'index, plus aisée "en ligne" que sur le papier. Elle est néanmoins compliquée par l'incohérence du lexique employé (*se posent en particulier des problèmes de synonymie et de forme des descripteurs*) et une certaine hétérogénéité de l'indexation. La liste des descripteurs n'est ni contrôlée ni structurée : il ne s'agit pas d'un thésaurus mais d'un simple lexique documentaire.

L'I.N.T.D. ne donne accès aux livres de son fonds (à l'exception des mémoires d'élèves, traités comme les périodiques) que par l'intermédiaire d'un catalogue alphabétique des auteurs et d'un catalogue topographique (plan de classement propre à la bibliothèque). En revanche, elle offre en matière de périodiques un système de recherche documentaire plus satisfaisant que celui de l'E.N.S.B. . Le système pratiqué est intermédiaire entre le simple lexique documentaire et le thésaurus. Les articles de périodiques sont indexés par la bibliothécaire-documentaliste selon une liste de descripteurs adaptés du Thésaurus sciences de l'information du C.N.R.S. (1). Cette liste est contrôlée mais peu structurée : elle n'admet que deux niveaux de vocabulaire. La recherche documentaire se fait sur un appareil à fiches superposables à sélection visuelle SPHINXO complété par un fichier UNITERM qui permet d'affiner la recherche. L'I.N.T.D. publie chaque mois depuis Juin 1976 des bulletins bibliographiques multigraphiés qui donnent un résumé de chaque article indexé et comprennent un index des descripteurs produit par ordinateur.⁽²⁾ Sur les soixante-quinze périodiques reçus actuellement par la bibliothèque, un tiers sont analysés de façon exhaustive, un tiers de façon sélective et un tiers occasionnellement.

Le centre de documentation de la D.B.M.I.S.T. est en réalité constitué provisoirement d'au moins deux fonds documentaires distincts. La D.B.M.I.S.T. a en effet hérité du fonds du centre de documentation spécialisé en bibliothéconomie, documentation et sciences de la communication qui se trouvait rue Dutot et du fonds du centre de documentation de la Division

(1) Pour les références bibliographiques des documents cités, voir la bibliographie en fin de volume.

(2) : note reportée en fin de volume, après la bibliographie .

de la coopération et de l'automatisation . Ces deux fonds sont actuellement séparés dans les deux salles du centre de documentation du boulevard Pasteur et cotés différemment, mais doivent être prochainement regroupés. Les documents venant de la rue Dutot sont recensés dans un catalogue alphabétique des auteurs et un catalogue alphabétique des matières. Idem pour le fonds issu de la D.I.C.A. augmenté des acquisitions récentes de la D.B.M.I.S.T. . Les périodiques actuellement reçus ne font l'objet d'aucun dépouillement. Comme cela n'a pas toujours été le cas, on trouve ~~des~~ articles de périodiques signalés, voire analysés, dans le catalogue-matières de la D.B.M.I.S.T.

La situation peut donc se résumer ainsi : d'une part les systèmes de recherche documentaire en place ne couvrent pas de manière égale l'ensemble des collections possédées par les bibliothèques. Les lacunes concernent plus particulièrement les périodiques à la D.B.M.I.S.T. et dans une certaine mesure à l'E.N.S.B., et les livres à l'I.N.T.D. . D'autre part, les possibilités de recherche documentaire sont limitées par les types mêmes de systèmes documentaires utilisés : catalogues alphabétiques de matières, plans de classement et lexiques documentaires - même améliorés - sont moins performants qu'un outil documentaire comme le thésaurus. Une comparaison rapide s'impose ici entre les principaux types de langages documentaires.

b - LE THESAURUS COMPARE AUX AUTRES TYPES DE LANGAGES DOCUMENTAIRES

Les classifications documentaires (non adoptées par nos biblio-

thèques) (1) sont des langages précoordonnés qui procèdent par emboîtement successif de notions : elles traduisent mal ou difficilement les notions complexes dont les éléments appartiennent à des arbres hiérarchiques différents. Elles imposent un cheminement obligatoire dans un domaine de la connaissance pour atteindre les concepts contenus dans un document. De plus, si le catalogue systématique se prête à une recherche s'intéressant à une discipline ou un domaine du savoir, il est peu commode à utiliser en cas de recherche sur un sujet précis.

L'indexation alphabétique de matières permet au contraire les recherches ponctuelles puisque le catalogue alphabétique des matières exprime directement, dans l'ordre alphabétique des vedettes-matières, les notions les plus spécifiques. Elle constitue toutefois un langage précoordonné (ordre des vedettes, sous-vedettes et précisions de lieu, de temps, de forme) qui n'autorise pas la recherche sur chaque élément de l'ensemble-descripteur, puisque toutes les permutations ne sont pas faites. D'un autre côté, si la structure obligatoire du descripteur équivaut dans une certaine mesure à un lien syntaxique, il subsiste des ambiguïtés (par exemple : "*Dessin, enfant*" peut signifier dessin fait par un enfant, ou pour un enfant).

Enfin, l'indexation alphabétique de matières a l'inconvénient d'établir peu de relations sémantiques entre les descripteurs : outre les relations d'équivalence, on trouve dans le catalogue alphabétique des matières des relations de voisinage (par exemple : Bibliothèque voir aussi Centre de documentation) mais très peu de relations hiérarchiques (par exemple : Chien voir aussi aux noms des différentes races).

Parmi les langages post-coordonnés, il faut d'abord mentionner les lexiques documentaires simples. Ils peuvent être totalement post-coordonnés : l'indexation se fait alors au niveau du mot, de la plus petite unité d'information significative. C'est le cas du système des "Uniterms", dans lequel le concept "Droit à la liberté d'expression" est exprimé à l'aide des trois Uniterms : Droit, Liberté, Expression. Le lexique se présente dans ce cas sous la forme d'une liste alphabétique de mots.

(1) Les trois bibliothèques ont simplement élaboré chacune un plan de classement qui ne peut servir (et n'est pas destiné) à l'indexation des documents : il est tout juste assez développé pour permettre le classement systématique des livres sur les rayons et la fabrication des cotes.

Mais il existe aussi des lexiques documentaires simples qui, bien que fondamentalement post-coordonnés, admettent un certain degré de précoordination. L'indexation se fait alors au niveau du concept ("Droit à la liberté d'expression") et le lexique se présente sous la forme d'une liste alphabétique de descripteurs.

Ces deux types de lexiques possèdent la souplesse liée à la post-coordination : la recherche est possible sur chaque descripteur, les notions complexes sont exprimées par le croisement de notions simples. Mais ils ont l'inconvénient de ne pas présenter de relations sémantiques entre les termes et de ne pas être nécessairement des langages contrôlés (en particulier on peut trouver des synonymes parmi les descripteurs).

Les thésaurus (1) combinent en principe les avantages des langages documentaires à structure hiérarchique (classifications) et ceux des langages à structure combinatoire. Ce sont des langages post-coordonnés. L'indexation consiste à énumérer les descripteurs traduisant les concepts relevés dans le document. Lors de la recherche, l'utilisateur combine à sa guise les descripteurs pour exprimer les notions complexes, objet de sa recherche. Mais il existe entre les descripteurs tout un réseau de relations sémantiques qui constituent une aide précieuse à la recherche documentaire. Les descripteurs sont regroupés soit par thèmes ou domaines, soit par facettes.

Les descripteurs de chaque domaine ou de chaque facette sont liés entre eux par des relations hiérarchiques (sur un ou plusieurs niveaux) et associatives. Des relations associatives sont aussi établies entre descripteurs de domaines ou de facettes différents. Lors de la recherche, une question peut ainsi être posée à plusieurs niveaux de vocabulaire ou encore étendue à des notions voisines, l'utilisateur étant en quelque sorte guidé dans ces opérations par le thésaurus lui-même.

Autre caractéristique constitutive des thésaurus : ce sont des langages contrôlés. Ils obéissent à des règles de forme, de sens et de structure qui visent à écarter les accidents du langage naturel et améliorent ainsi l'efficacité de l'indexation et donc le rappel de l'information.

(2) La forme plurielle employée dans ce mémoire sera "thésaurus", selon le principe adopté par exemple pour "médias"

Un thésaurus se présente sous la forme de listes de termes (descripteurs ou non) : listes alphabétiques, listes systématiques, présentées de manières diverses. Aux listes structurées sous forme d'arbres hiérarchiques se substitue ou s'ajoute parfois un mode de présentation graphique (schémas ou tableaux).

c - LE THESAURUS ET L'ORDINATEUR

Le thésaurus est le fruit d'un travail intellectuel mené par l'homme. L'ordinateur peut apporter une aide appréciable, mais secondaire à sa construction. Il peut en particulier assurer certaines opérations de manipulation de vocabulaire (par exemple la fabrication de l'index permuté), de contrôle du langage, et l'édition du thésaurus (1).

Par contre, l'ordinateur se révèle extrêmement utile voire indispensable lorsqu'il s'agit de mettre en correspondance, en vue de la recherche documentaire, le thésaurus et le fonds documentaire. Il serait en effet très fastidieux, d'abord pour le bibliothécaire-documentaliste, d'établir manuellement et de mettre à jour régulièrement une liste alphabétique des descripteurs où devrait figurer, dans le meilleur des cas, chaque descripteur à sa place alphabétique avec derrière lui, dans l'ordre alphabétique également, les autres descripteurs utilisés pour indexer un document donné, enfin le numéro de ce document ; ce qui voudrait dire qu'un document indexé avec huit descripteurs se trouverait signalé huit fois dans la liste à la place alphabétique de chacun des huit descripteurs et que chaque descripteur serait répété dans la liste autant de fois qu'il est utilisé pour l'indexation. Il serait fastidieux également pour l'utilisateur de rechercher dans une telle liste les références qui l'intéressent. Et l'on pourrait certes alléger le travail du bibliothécaire en lui faisant constituer une liste alphabétique simple où chaque descripteur ne figurerait qu'une fois avec, en face de lui, tous les numéros des documents indexés à l'aide de ce descrip-

(1) Les procédures d'indexation automatique ne sont pas encore bien au point. De toute façon, le travail d'indexation ne représente qu'une partie du processus d'élaboration d'un thésaurus

teur, mais ce serait compliquer la tâche de l'utilisateur qui devrait rechercher lui-même les numéros résultant du croisement d'un certain nombre de descripteurs. Bref, il vaut mieux avoir recours à un système informatique qui permette une interrogation sur ordinateur à l'aide des opérateurs booléens et qui édite les index dont il vient d'être question.

D'autre part il est possible d'améliorer les performances d'un thésaurus au moyen de quelques procédés dont l'application est plus facile voire seulement possible sur ordinateur. Parmi ces procédés, dont certains seront décrits en [II-4, on peut citer l'autopostage, la pondération et les techniques des liens et des indicateurs de rôle.

Le thésaurus exploité de manière informatique apparaît donc comme le système documentaire le plus approprié à l'indexation et à la recherche dans une bibliothèque spécialisée (1). C'est pourquoi il a été retenu pour l'E.N.S.B., l'I.N.T.D. et la D.B.M.I.S.T. . Il serait évidemment intéressant que l'indexation porte à la fois sur les livres et sur les périodiques, un seul système de recherche se substituant ainsi à la coexistence actuelle de plusieurs systèmes. L'organisation de la mémoire documentaire et l'utilisation d'un système informatique nous épargneront les problèmes d'inflation de fichiers tels que peut en poser un catalogue par matières sur fiches qui représente souvent un document par plusieurs fiches. Pour cette même raison, on craindra moins, si nécessaire, d'indexer les documents de manière plus exhaustive que cela n'a été le cas jusqu'ici, par exemple à l'E.N.S.B. (voir le chapitre sur la politique d'indexation).

(1) Les listes de vedettes-matières comme celles de la B.P.I. ou de la B.N. sont destinées à des bibliothèques encyclopédiques.

2 - POURQUOI UN THÉSAURUS COMMUN AUX TROIS BIBLIOTHÈQUES ?

a - AVANTAGES

La pratique commune d'un thésaurus présente des avantages certains, d'abord pour les bibliothécaires. L'élaboration d'un thésaurus est une tâche délicate qui demande du temps. Le travail de compilation du vocabulaire, voire de hiérarchisation des domaines (ou des facettes) du thésaurus peut être réparti entre les trois organismes. D'autre part, le thésaurus peut bénéficier de la mise en commun de l'expérience professionnelle d'indexeurs travaillant dans les mêmes disciplines mais dans des bibliothèques différentes où les besoins documentaires sont à la fois multiples et comparables (voir ci-dessous).

Mais on peut aussi envisager, dans des conditions qui seront décrites plus loin, qu'une fois le thésaurus constitué, les trois bibliothèques se partagent l'indexation des documents qu'elles possèdent et qu'elles acquièrent. Cela allègerait là encore la tâche des indexeurs.

Bien évidemment, le public utilisateur y gagnerait aussi. D'abord parce que la connaissance d'un seul langage documentaire lui suffirait pour avoir accès aux fonds des trois bibliothèques. Ensuite parce qu'en cas de mémoire documentaire unique et partagée, consultable en ligne à partir des trois établissements, l'utilisateur pourrait même à partir d'une bibliothèque faire porter sa recherche documentaire sur les fonds des trois bibliothèques.

b - CONDITIONS DE POSSIBILITE

L'idée d'un thésaurus commun est donc séduisante ; encore faut-il s'assurer qu'elle est dans l'ordre des choses possibles. La première condition réside dans la coïncidence sémantique des différents fonds. On s'aperçoit que si, à l'intérieur du vaste domaine commun que représentent les sciences relatives à l'information, à la communication, à la documentation et aux bibliothèques, les centres de gravité des trois fonds ne se superposent pas exactement, le décalage n'est pas tel qu'il fasse obstacle à la réalisation d'un thésaurus commun. Il convient néanmoins d'examiner la situation, ce qui nous permettra en même temps de définir le domaine qui doit être couvert par le thésaurus. La comparaison des trois fonds documentaires entre eux nous sera, d'autre part, utile lorsqu'il s'agira de répartir certaines tâches entre les bibliothèques (par exemple la compilation du vocabulaire, le dépouillement des périodiques).

On peut dégager, dans les trois fonds de livres, un tronc commun quantitativement important constitué par la bibliothéconomie (1) (à laquelle on peut adjoindre l'histoire des bibliothèques et l'archivistique) et les sciences de l'information et de la documentation (2). Ce domaine est largement représenté à la bibliothèque de l'E.N.S.B. . Il constitue la plus grosse partie du fonds de l'I.N.T.D. . A la D.B.M.I.S.T., il correspond essentiellement au fonds issu du centre de la rue Dutot auquel il faut cependant ajouter, dans l'autre fonds, tout le secteur de l'automatisation des bibliothèques et de l'informatique documentaire, plus développé qu'à l'E.N.S.B. et à l'I.N.T.D.

- (1) telle que la définit la norme AFNOR Z 40-001 : "Science, techniques et activités relatives à l'organisation, la gestion, la législation et la réglementation des bibliothèques".
- (2) Je désigne par là tout ce qui a trait aux techniques documentaires et aux systèmes, services, réseaux et politiques d'information, bref l'ensemble des questions traitées par exemple dans l'Introduction générale aux sciences et techniques de l'information et de la documentation de Claire GUINCHAT et Michel MENOÛ.

A côté de ce large tronc commun, on distingue quelques branches de développement variable. Pour mieux en discerner l'agencement et les rapports avec les sciences de l'information et de la documentation, on peut s'aider ici de la tentative de classification "des sciences de l'information et de la communication" présentée par Jean MEYRIAT dans un article de la revue Schéma et schématisation (1). Si l'on résume la pensée de Jean MEYRIAT, les sciences de l'information et de la communication recouvrent les disciplines suivantes :

1 - *La médialogie* qui étudie les médias, "systèmes technosociaux" organisés pour rendre possible la communication. La médialogie se décompose en à peu près autant de disciplines qu'il y a de médias (radio, télévision, cinéma, image, disque etc ...). A la médialogie il faut adjoindre, si l'on ne veut pas l'y inclure, la bibliologie, "*étude systématique des conditions de production-diffusion et utilisation des écrits imprimés sous toutes leurs formes (...)*" (définition de Jacques BRETON) (2) (3).

2 - *L'informatologie*, "*science du contenu sémantique de la communication, donc de l'information, dont elle veut expliquer la nature, les conditions dans lesquelles elle est produite, transmise, reçue, conservée ou transformée*". L'informatologie s'intéresse donc d'une part aux propriétés de l'information, d'autre part aux mécanismes de transmission du sens, et ce aussi bien au niveau du système primaire qu'à celui du système secondaire de transfert de l'information. La documentation en particulier, fait partie, en tant que processus d'extraction et de traitement de l'information, du champ d'investigation de l'informatologie (4)

3 - *La communicologie fonctionnelle* concentre son attention sur "*les systèmes de communication chargés de remplir une fonction particulière*". J. MEYRIAT distingue entre autres la communication persuasive (publicité, re-

(1) Voir Bibliographie, référence n°13

(2) Voir réf. bibliographique n°5

(3) L'étude de la presse écrite appartient-elle à la bibliologie ou à la médialogie, celle-ci s'intéressant à la presse radiodiffusée et télévisée ?

(4) Voir à ce sujet certains autres articles de J. MEYRIAT dans la revue citée, en particulier : réf. bibliographique n°12

lations publiques, propagande ...), la communication didactique, la communication artistique, l'animation socio-culturelle.

4 - A la périphérie des sciences de l'information et de la communication se situent d'une part *la technologie de la communication* (télécommunications, télématique, moyens audio-visuels, techniques d'impression, reprographie ...), d'autre part *les sciences sociales de l'information et de la communication* (psychosociologie de la lecture, sociologie de la création littéraire, de la littérature, sciences juridiques de l'information, etc ...).

5 - J. MEYRIAT distingue enfin des "*sciences connexes*" qui ne font que fournir aux sciences de l'information et de la communication des outils et des méthodes utilisables dans de tout autres domaines : d'une part *les sciences formelles* (mathématiques, logique, cybernétique, systématique ...), d'autre part *les sciences des outils* (linguistique, informatique, etc ...).

Si l'on excepte les sciences connexes, c'est l'E.N.S.B. qui dans les domaines autres que celui de la bibliothéconomie et des sciences de l'information et de la documentation possède le plus grand nombre de documents (1). La D.B.M.I.S.T. a toutefois hérité dans ces domaines des collections non négligeables de la rue Dutot. A l'I.N.T.D. ces disciplines sont représentées de manière lacunaire. Dans les sciences connexes par contre, le centre de gravité se déplace de l'E.N.S.B. à la D.B.M.I.S.T. et à l'I.N.T.D. . La science informatique est nettement plus développée à la D.B.M.I.S.T. que dans les deux autres bibliothèques. L'I.N.T.D. semble plus riche en livres de mathématiques que l'E.N.S.B., en livres de linguistique que l'E.N.S.B. et la D.B.M.I.S.T. .

(1) Il est préférable de signaler ici les différences quantitatives dans la représentation de tel domaine dans les trois bibliothèques puisqu'il y a un rapport - non systématique toutefois - entre le nombre de documents possédés dans un domaine donné et le nombre de descripteurs nécessaires pour traduire le contenu de ces documents. En informatique par exemple la D.B.M.I.S.T. aurait sans doute besoin de plus de noms de langages informatiques que l'E.N.S.B. et l'I.N.T.D.

Pour terminer ce tour d'horizon documentaire, il faut mentionner des domaines divers plus ou moins directement liés aux bibliothèques et aux sciences de l'information et de la communication : l'administration et le management, l'éducation, les arts graphiques, l'action culturelle, la littérature ... Ces domaines, pratiquement absents de l'I.N.T.D., sont surtout présents à l'E.N.S.B., dans une moindre mesure à la D.B.M.I.S.T. (fonds de la rue Dutot) sauf en ce qui concerne le management, secteur beaucoup plus développé qu'à l'E.N.S.B. .

Quant à la "bibliothèque générale" de l'E.N.S.B., elle est bien évidemment exclue de notre projet de thésaurus nécessairement spécialisé. On pourrait envisager d'utiliser pour l'indexation des livres de cette partie de la bibliothèque une liste d'autorité-matières.

Si l'on considère maintenant les périodiques, la situation reste inchangée. En effet, le contenu sémantique des périodiques reflète à peu de choses près, dans chaque bibliothèque, celui des livres. Parmi les périodiques reçus et dépouillés par l'I.N.T.D. prédominent ceux qui traitent de documentation et systèmes d'information, de bibliothéconomie, d'informaticque documentaire. Même prédominance à la D.B.M.I.S.T. à laquelle il faut ajouter un nombre non négligeable de périodiques spécialisés dans l'informatique en général et la bureautique. A l'E.N.S.B. aussi les disciplines citées pour l'I.N.T.D. sont largement représentées mais la bibliothèque de l'E.N.S.B. se distingue des deux autres par le nombre plus élevé de titres de périodiques qu'elle reçoit et dépouille dans des secteurs comme l'histoire du livre, la fabrication de l'imprimé, l'audio-visuel, l'administration, la littérature pour enfants ou l'éducation. L'I.N.T.D. en particulier reçoit très peu de périodiques spécialisés dans ces secteurs, dans certains cas pas du tout. Mais cela ne veut pas dire que certaines des notions spécifiques de ces domaines ne soient pas présentes, dans son fonds, dans des périodiques d'autres spécialités ou dans des livres.

Il faut enfin signaler que parmi les périodiques dépouillés à l'E.N.S.B. se trouvent des titres non spécialisés dans les disciplines présentées jusqu'ici : par exemple La Recherche, Pour la science, L'Histoire, La Revue des sciences humaines, la Revue internationale des sciences sociales,

etc ... Il ne semble pas toutefois que cela nous oblige à inclure dans notre thésaurus une liste hétéroclite de descripteurs de toutes provenances ou à exclure ces titres de périodiques du système documentaire projeté, dans la mesure où les articles qui y ont été sélectionnés jusqu'ici par les indexeurs intéressent pour la plupart la profession de bibliothécaire-documentaliste.

Une évaluation rapide du fonds de livres et de périodiques des trois bibliothèques montre que les différences, tant qualitatives que quantitatives, dans la représentation documentaire des différents domaines à couvrir ne suffisent pas à justifier des thésaurus séparés pour la recherche documentaire. Il faudra seulement essayer de donner au thésaurus une structure telle que les habitués de chaque bibliothèque puissent s'y retrouver sans trop de peine. Autrement dit, il faudrait d'abord faire en sorte que les différents (sous)-domaines qui constituent les trois fonds soient aisément repérables dans la structure d'ensemble du thésaurus (1). Il serait ensuite souhaitable que les domaines du thésaurus qui seront légèrement sous-utilisés par une ou deux bibliothèques sur les trois puissent, précisément, être facilement exploités à ce niveau, c'est-à-dire que l'indexeur et l'utilisateur ne se sentent pas noyés dans une liste trop longue de descripteurs : d'où le soin à apporter à la structure interne des domaines (établissement des relations hiérarchiques). Il faudra enfin veiller à ce que le thésaurus dans son ensemble conserve une dimension raisonnable.

Pour évaluer la pertinence de l'utilisation d'un même thésaurus par plusieurs bibliothèques de spécialité identique, il convient également de s'interroger sur la recherche documentaire qui y est pratiquée. En effet, l'indexation courante des documents est nécessairement fonction des besoins des utilisateurs documentaires et le thésaurus lui-même, outil d'indexation, est en partie déterminé par les modalités et le niveau de la recherche documentaire. Ainsi par exemple, deux centres de documentation spécialisés dans le même champ du savoir mais pratiquant exclusivement l'un la diffusion

(1) Je me place ici dans l'hypothèse d'un thésaurus organisé par domaines ou thèmes et non par facettes. Ce choix est justifié en II,2,b.

sélective de l'information, l'autre la recherche rétrospective, pourraient utiliser des descripteurs qui ne se situent pas, dans l'ensemble, à un même niveau de vocabulaire. On pourrait envisager dans ce cas un thésaurus à deux niveaux (différenciation nette du "macro-vocabulaire" et du vocabulaire plus spécifique) (1) ou, à la limite, deux thésaurus distincts, l'un se maintenant à un certain niveau de généralité.

Mais il semble que se manifestent dans nos trois bibliothèques, en matière de recherche documentaire, des besoins assez comparables dans l'ensemble, donc propres à être satisfaits par un thésaurus commun. (Ces besoins appelleront du reste un thésaurus qui présente toute la gamme des niveaux de vocabulaire). Les élèves de l'I.N.T.D. font pour leur mémoire de fin d'études des recherches du même type que celles qui sont pratiquées à la bibliothèque de l'E.N.S.B. par les élèves de l'Ecole pour leur propre mémoire, par les étudiants inscrits au D.E.S.S. d'informatique documentaire pour leur note de synthèse et par les étudiants et stagiaires du cours post-universitaire de l'UNESCO. Le centre de documentation de la D.B.M.I.S.T. n'est pas encore fréquenté par les étudiants mais envisage de leur ouvrir ses portes à l'avenir (actuellement aucune publicité n'est faite à l'extérieur). D'autre part, le même type de recherche documentaire assez précise peut être pratiqué par certains utilisateurs extérieurs des bibliothèques de l'I.N.T.D. et de l'E.N.S.B., par les enseignants de l'E.N.S.B. (par exemple pour rédiger un article) et par le personnel du Bulletin des bibliothèques de France au centre de documentation de la D.B.M.I.S.T. (2).

(1) Voir CHAUMIER (Jacques) Analyse et langages documentaires, p. 41 et p. 128

(2) Le personnel travaillant à la D.B.M.I.S.T., surtout les informaticiens, représente à l'heure actuelle la grosse majorité des utilisateurs du centre de documentation

II

LES CARACTERISTIQUES FORMELLES DU THESAURUS

Langage contrôlé et normalisé, un thésaurus obéit à un certain nombre de règles qui le caractérisent formellement : règles de sens et de forme des unités lexicales qui le composent, règles de structure et de présentation du thésaurus. Ces règles sont énoncées, parfois sous forme de simples recommandations, voire de choix laissé à la volonté du concepteur, dans la norme française homologuée NF Z 47.100 de Décembre 1981 (1). L'usage qui en est fait varie, parfois sensiblement, d'un thésaurus à l'autre. Il nous faudra donc, à partir de la norme AFNOR et dans la mesure où le choix nous en est laissé, déterminer précisément les caractéristiques formelles que nous souhaitons donner à notre thésaurus, comme je vais essayer de le faire ci-dessous en rappelant en même temps les principales règles à observer. Je me laisserai guider dans ce choix par l'examen comparatif de quelques thésaurus existants, en particulier le Thésaurus de l'UNESCO (2) et le Thésaurus sciences de l'information du C.N.R.S. (C.D.S.T.).

1 - LES UNITES LEXICALES

a - REGLES DE SENS

Il existe des thésaurus qui admettent pour l'indexation et l'interrogation tous les termes représentant une notion, autrement dit qui traitent

(1) Ainsi que dans la norme Iso 2788-1974. Il est souhaitable de prendre connaissance des deux normes, car certains points y sont inégalement développés.

(2) Réf. bibliographique n° 31 . C'est toujours de ce thésaurus qu'il s'agira lorsque je parlerai simplement du "thésaurus de l'UNESCO"

tous les termes comme descripteurs, faisant correspondre à tous les descripteurs exprimant la même notion un code identique. Ces thésaurus exigent l'emploi de la machine pour la mise à jour et la recherche documentaire. Il me semble préférable que nos bibliothèques se dotent d'un thésaurus à termes préférentiels, c'est-à-dire où une même notion n'est représentée que par un descripteur. Il faut donc, pour chaque ensemble de synonymes linguistiques, faire le choix d'un descripteur qui seul pourra être utilisé à l'indexation comme à l'interrogation. Les synonymes ne sont pas pour autant éliminés, mais figurent dans la liste alphabétique ou le dictionnaire conceptuel comme "*termes équivalents*" ("*non-descripteurs*") renvoyant nécessairement aux descripteurs uniques correspondants. Le descripteur peut être choisi pour sa conformité à certaines des règles de forme et de sens énoncées ci-dessous, mais aussi pour des raisons diverses (adéquation à la notion exprimée, position hiérarchique, fréquence d'utilisation, mot ou expression composée, connotation ...) auxquelles le concepteur du thésaurus doit être sensible s'il veut faire le choix le plus judicieux. Il ne faut pas négliger, pour un descripteur donné, de réunir tous les synonymes qui pourraient raisonnablement venir à l'esprit de l'utilisateur du système documentaire. Selon Jacques CHAUMIER il est souhaitable qu'un thésaurus comprenne à peu près autant de termes équivalents que de descripteurs (taux d'équivalence $\simeq 1$) (1).

Autre règle à observer : un descripteur ne sert à exprimer qu'une notion. Il s'agit donc de résoudre les cas de polysémie (et d'homographie) par l'adoption de deux descripteurs quand les deux concepts désignés ont leur place dans le thésaurus ou quand il risque d'y avoir confusion (quand un seul concept figure dans le thésaurus, le contexte suffit parfois à lever l'ambiguïté). On peut avoir recours à un synonyme linguistique, à l'adjonction d'un qualificateur (BRUIT (informatique)) ou à la transformation du terme en un syntagme (PRESSE et PRESSE A IMPRIMER). Il faut parfois user en outre d'une note d'application.

(1) CHAUMIER (Jacques) Analyse et langages documentaires, p. 104

D'une manière générale, celle-ci doit être accolée au descripteur toutes les fois que subsiste une ambiguïté, pour en préciser la définition, en limiter l'emploi, ou expliciter une abréviation.

La règle qui vient d'être énoncée peut subir une légère entorse lorsque des termes fins qui seraient d'une utilisation trop rare sont renvoyés sur un terme plus général qui les représente.

Le choix des descripteurs implique également qu'ait été résolue la question parfois délicate de leur degré de coordination : jusqu'où faut-il décomposer les notions pour leur traduction en descripteurs ? On sait qu'en utilisant au maximum des descripteurs simples, on réduit le volume du thésaurus et l'on diminue le risque de silence lors de la recherche, mais qu'on augmente le risque de bruit et surtout qu'on restreint les possibilités de hiérarchisation entre les descripteurs. A l'inverse, avec un langage plus précoordonné, on augmente certes le volume du thésaurus, mais aussi le taux de pertinence (on a moins de bruit), et l'augmentation corrélative du risque de silence est en partie compensée par la possibilité qui est offerte d'établir des relations hiérarchiques plus poussées permettant une recherche documentaire plus complète et plus affinée. Il me semble que ce dernier facteur doit en particulier motiver notre choix. Dans le cas d'un thésaurus organisé par thèmes ou domaines, le problème se pose de la manière suivante. Prenons par exemple les notions complexes de "vente des livres" et "vente des disques". On pourrait choisir de les décomposer en VENTE + LIVRE et VENTE + DISQUE ; on économiserait ainsi un descripteur. Mais le descripteur VENTE figurerait alors soit dans la liste des mots-outils, soit dans un arbre hiérarchique où il soit univoque : il ne pourrait pas figurer à la fois sous les rubriques INDUSTRIE DU LIVRE et INDUSTRIE DU DISQUE. Au contraire, les descripteurs VENTE DU LIVRE et VENTE DU DISQUE y ont chacun leur place respective, le premier par exemple dans l'environnement sémantique :

DIFFUSION DU LIVRE

VENTE DU LIVRE

VENTE EN LIBRAIRIE

VENTE PAR CORRESPONDANCE

Or il est intéressant pour la recherche documentaire de disposer de cet ensemble sémantique hiérarchisé. Toutes les fois que l'importance de la position hiérarchique d'un descripteur conjuguée au risque d'ambiguïté l'impose, il faut donc, dans un thésaurus à structure thématique, adopter pour ce descripteur une forme précoordonnée.

Il faut enfin mentionner ici deux types de descripteurs particuliers, sémantiquement distincts des descripteurs ordinaires, qui seront présents dans notre thésaurus : les mots-outils et les identificateurs.

Les mots-outils sont des termes non discriminants et généralement utilisés en association avec un ou plusieurs descripteurs (COMPARAISON, EVALUATION, EVOLUTION). Trop généraux pour figurer dans les listes thématiques, ils doivent être regroupés dans une liste propre qui pourra également comprendre par exemple, les noms de langues et les noms des arts, des sciences et des techniques. Ils figurent également dans le dictionnaire conceptuel.

Les identificateurs sont des noms propres ou assimilés : noms de collectivités, de lieux géographiques, de personnes, de projets ou de systèmes, titres de livres ou de périodiques ... Ils ne figurent que dans des listes annexes qui doivent être régulièrement mises à jour.

b - REGLES DE FORME

LES DESCRIPTEURS

Ils doivent être à la forme substantive. On peut cependant employer parfois un adjectif seul plutôt qu'un adjectif associé à un substantif (par exemple BILINGUE, INTERNATIONAL ; voir la norme à ce sujet). Selon certains auteurs, lorsque les deux genres sont possibles, il faut adopter

le genre masculin (EMPLOYE DE BIBLIOTHEQUE). La norme ne fait pas mention de cette règle. Il me semble nécessaire d'admettre la présence d'un terme au féminin lorsque le signifié de celui-ci n'est pas équivalent au signifié de son homologue masculin. Ainsi, l'on trouve, dans le thésaurus de l'UNESCO, comme terme spécifique de TRAVAILLEURS, le descripteur TRAVAILLEUSES sur le même plan que TRAVAILLEURS ENFANTS etc ... La règle du singulier appelle un développement. La norme AFNOR recommande d'écrire les descripteurs au singulier sauf lorsque la seule forme plurielle existe (ARCHIVES) ou lorsque le mot n'a pas le même sens au singulier et au pluriel (ECHECS). Or, on constate qu'un très grand nombre des descripteurs du thésaurus de l'UNESCO sont au pluriel. Comme seule explication, l'introduction signale que le thésaurus suit à cet égard les règles de l'Engineers Joint Council (E.J.C.). Recherches faites, il s'avère que l'E.J.C. emploie dans son thésaurus le pluriel pour toutes les notions que l'on peut dénombrer (1). Cela donne par exemple dans le thésaurus de l'Unesco : TELESPECTATEURS, CATALOGUES, PERTES DE DOCUMENTS, SERVICES DE PRET. Restent donc au singulier, dans ce thésaurus, les descripteurs désignant une action ou un ensemble d'actions (INDEXATION), un processus ou un ensemble de processus (AUTOMATISATION DES BIBLIOTHEQUES), une propriété (EXHAUSTIVITE DE L'INDEXATION), un matériau spécifique (PAPIER). Quel avantage cette fidélité aux règles anglaises offre-t-elle en français ? On peut en déceler au moins un : dans le cas où un descripteur peut à la fois désigner un processus et le produit de ce processus (ACQUISITION), ce dernier peut être exprimé, à la différence du premier, par le pluriel sans que la cohérence globale soit rompue puisque tous les descripteurs désignant un objet physique sont au pluriel. Autre exemple : ENREGISTREMENT signifierait le processus technique ou l'ingénierie de l'enregistrement, ENREGISTREMENTS, les produits physiques de l'opération d'enregistrement (bandes sonores, minicassettes, etc ...). Pourtant ce parti pris est loin d'être satisfaisant ainsi que le montre l'exemple du thésaurus de l'UNESCO. D'abord parce que les raisons du choix ne sont pas toujours transparentes. Pourquoi, par exemple, le singulier FICHER INVERSE alors qu'INDEX PERMUTES est au pluriel ? Pourquoi d'un côté FAISCEAU DE SOUS-CLASSES (singulier), de l'autre RELATIONS PARTIES-TOUT ? Pourquoi NOTATION au sens de *"langage artificiel qui exprime les relations entre les classes d'une classification"* figure-t-il au singulier ? Et si la raison en est l'équivocité du pluriel, pourquoi conserver le singulier pour les descripteurs spécifiques (NOTATION

(1) La norme Iso indique qu'en anglais il y a lieu en général d'employer dans un même thésaurus le singulier et le pluriel, selon les cas.

HIERARCHIQUE etc ...) dont l'épithète lève l'ambiguïté ? D'autre part, si l'oeil s'habitue vite à l'emploi du pluriel pour les objets physiques ou les êtres humains, il reste gêné dans certains autres cas, parfois

contestables : par exemple : HEURES DES CONTES. Enfin ce système a tendance à compliquer un peu la recherche dans la liste alphabétique des descripteurs. On peut donc se demander s'il n'est pas préférable de conserver la règle du singulier, d'autant que les cas d'ambiguïté processus-produit sont plutôt rares. Pour lever la polysémie dans ces quelques cas précis, on peut par exemple employer un synonyme pour désigner le produit concret et accompagner le descripteur de nombre singulier d'une note d'application qui en limite l'usage à la notion de processus et qui renvoie, pour l'autre usage, vers le synonyme. Par exemple : IMAGE et REPRODUCTION PHOTOGRAPHIQUE. NA : désigne le processus de reproduction. Pour exprimer l'image photographique, employer IMAGE

On constate d'ailleurs que même dans le thésaurus de l'UNESCO la distinction singulier-pluriel est renforcée par l'emploi de descripteurs différents (par exemple ENREGISTREMENT DU SON et ENREGISTREMENTS PHONIQUES).

Comme autres règles de forme des descripteurs, il faut mentionner l'emploi de la forme développée des sigles - sauf précise la norme AFNOR, lorsque le sigle est universellement connu (O.N.U.) ; de l'orthographe la plus largement admise ; de la séquence normale des termes d'une expression (DROIT D'AUTEUR et non AUTEUR (DROIT D')). La norme conseille également de supprimer les mots de liaison entre les éléments significatifs des expressions composées, quand cela n'entraîne pas d'ambiguïté. Cette solution a été adoptée dans le thésaurus des sciences de l'information du C.N.R.S. (ACCES INFORMATION) ; elle ne l'a pas été dans le thésaurus de l'UNESCO (ACCES A L'INFORMATION). Elle a quelques avantages pratiques. En particulier elle épargne à l'indexeur l'obligation de faire un choix lorsque l'emploi de l'expression est mal fixé (CIRCULATION DE L'INFORMATION ou CIRCULATION D'INFORMATION) et facilite dans le même cas la recherche dans la liste alphabétique. Il est enfin conseillé d'employer comme descripteur le terme le plus couramment utilisé (ORDINATEUR et non CALCULATEUR ELECTRONIQUE), et les noms communs de préférence aux noms de marques et aux dénominations commerciales (par exemple éviter SELECTO).

LES IDENTIFICATEURS

La forme des identificateurs doit être normalisée comme celle des descripteurs. On peut utiliser à cet effet une liste d'autorité, par exemple la Liste des vedettes-matières de la Bibliothèque nationale (noms communs et noms propres), mise à jour chaque trimestre et plus récente que la liste : Autorité matière : noms propres de la Bibliothèque publique d'information.

2 - LA STRUCTURE DU THÉSAURUS

a - LES RELATIONS ENTRE DESCRIPTEURS

La norme NF Z 47.100 signale à juste titre que la valeur d'un thésaurus en tant qu'outil documentaire réside autant dans le choix des descripteurs que dans celui de leurs relations sémantiques : relations d'équivalence, relations hiérarchiques et relations d'association. Aussi convient-il de suivre attentivement les règles et recommandations précises de la norme lors de l'établissement de ces relations entre les descripteurs. On évoquera seulement ici quelques points particuliers.

Les relations hiérarchiques et les relations d'équivalence sont toujours réciproques, alors que ce n'est pas nécessairement le cas des relations associatives. Pour ces dernières par exemple, la réciprocité est systématique dans le thésaurus des sciences de l'information du C.N.R.S., variable dans le thésaurus de l'UNESCO. La réciprocité systématique constitue la solution de facilité : il n'est pas toujours simple de déterminer pourquoi une relation d'association est établie dans un sens et non dans l'autre.

Les relations hiérarchiques, comme le rappelle la norme AFNOR, doivent être établies avec d'autant plus de rigueur qu'elles peuvent être utilisées automatiquement (dans un système automatisé), à l'indexation des documents ou à l'indexation des questions, pour améliorer le taux de rappel du système. Notre thésaurus sera, en raison de sa spécialité, principalement composé de relations génériques, c'est-à-dire "*identifiant le lien entre une classe ou catégorie de notions, et chacun de ses cas particuliers*" (norme AFNOR). On pourra toutefois y rencontrer aussi quelques cas de relations partitives, en particulier dans la liste des noms géographiques si l'on juge bon de structurer celle-ci ou bien entre entités administratives. Il faudra à ce sujet compléter la lecture de la norme AFNOR par celle de la norme Iso.

Un thésaurus peut être monohiérarchique ou polyhiérarchique, selon qu'une notion représentée par un descripteur est subordonnée à une seule notion ou éventuellement à plusieurs notions. Les polyhiérarchies, lorsqu'elles sont établies à bon escient (ce qui est parfois délicat), améliorent l'efficacité de la recherche documentaire. Le thésaurus de l'UNESCO est polyhiérarchique.

b - LA STRUCTURE D'ENSEMBLE

Les relations hiérarchiques sont nécessairement établies à l'intérieur d'ensembles ayant une caractéristique commune. Pour structurer un ensemble de notions ou de concepts, on a le choix entre deux approches possibles : le regroupement par facettes et le regroupement par thèmes ou domaines. Le C.D.S.T. a adopté pour son thésaurus des sciences de l'information la première solution. L'ensemble des descripteurs est réparti en neuf facettes parmi lesquelles on retrouve les cinq facettes fondamentales citées dans la norme AFNOR : matériau, équipement, phénomène, processus et propriété. Le thésaurus de l'UNESCO, au contraire, est organisé par domaines, par exemple : Communication, Information, "Bibliothéconomie et

sciences de l'information", Archives, Culture, Education, etc ...

La notion de facette est plus rigoureuse que celle de thème ou de domaine mais moins directement accessible. Surtout le regroupement par facettes a le gros inconvénient de délocaliser la notion de son domaine d'application. La comparaison entre le thésaurus Sciences de l'information et le thésaurus de l'UNESCO est à cet égard instructive. Les descripteurs DESCRIPTION BIBLIOGRAPHIQUE, TITRES, VEDETTES-AUTEURS, sont classés dans le thésaurus de l'UNESCO sous le terme CATALOGAGE lui-même spécifique de la notion TRAITEMENT DE L'INFORMATION . Il faut tourner quelques pages pour trouver le descripteur CATALOGUES , mais celui-ci est tout de même rangé dans la subdivision TRAITEMENT DE L'INFORMATION. Les filiations sont donc claires. Dans le thésaurus Sciences de l'information, les mêmes notions sont complètement dispersées. Certes les descripteurs CATALOGUE et VEDETTE-AUTEUR sont proches l'un de l'autre dans la facette "équipement", mais TITRE et DESCRIPTION BIBLIOGRAPHIQUE se trouvent, en tant qu'"élément de document", sous la facette "matériau" tandis que CATALOGAGE, subordonné à TRAITEMENT INFORMATION , est classé comme "processus". A l'inverse une même facette rassemble évidemment des notions sémantiquement éloignées les unes des autres. Dans la facette "équipement" se côtoient l'auteur, le bibliothécaire ("élément humain"), les archives et les bibliothèques publiques ("organisme"), les classifications ("outils documentaires"), les algorithmes et les caméras de télévision ("théorie et méthode").

Il est clair qu'une telle structuration du langage documentaire est dans notre cas pratiquement inutilisable. On se sert, dans le thésaurus du C.N.R.S., des relations hiérarchiques établies à l'intérieur des facettes mais on ne peut pas cheminer dans un domaine de la connaissance ou le survoler comme on le ferait avec un thésaurus thématique (ajoutons toutefois que la notion de facette conserve son utilité dans un thésaurus à structure thématique : elle peut être utilisée dans l'organisation interne des domaines).

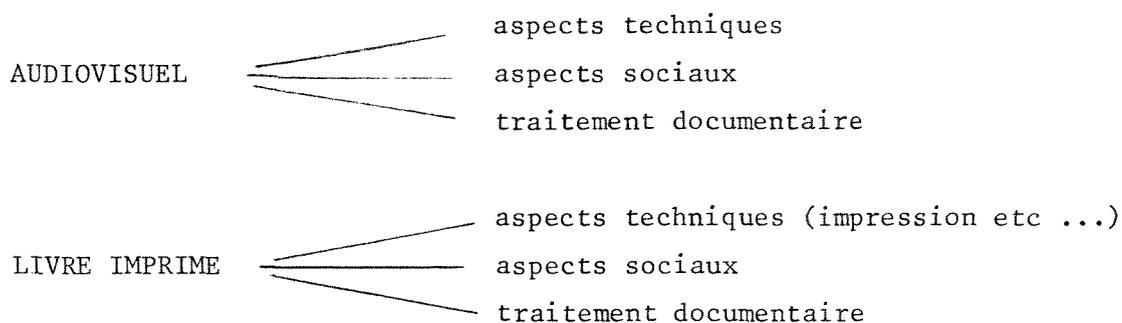
Il me semble donc préférable que nos trois bibliothèques adoptent pour leur thésaurus commun une organisation par "domaines". Toutefois, une précision s'impose ici. La notion de domaine peut être abordée par deux voies légèrement différentes, l'une privilégiant l'idée de "thème" (par exemple l'audiovisuel, le livre imprimé, l'information), l'autre, l'idée de discipline scientifique (par exemple la bibliothéconomie, la bibliologie, les sciences sociales de l'information).

Schématiquement, on pourrait ainsi envisager pour un thésaurus deux modes d'organisation inverses, selon que l'on privilégie l'approche par les thèmes, qui pourraient alors être eux-mêmes subdivisés selon les disciplines scientifiques, ou l'approche par les disciplines, elles-mêmes sous-structurées de manière thématique. Le tableau suivant clarifiera la situation à l'aide d'un exemple limité.

Approche par les thèmes

1er niveau d'organisation

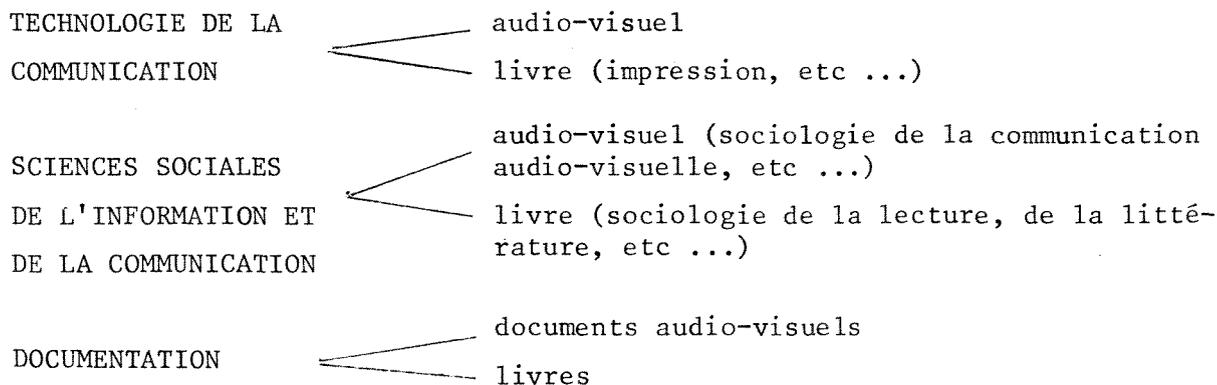
2ème niveau d'organisation



Approche par les disciplines

1er niveau d'organisation

2ème niveau d'organisation



Peut-être est-il possible, dans certaines disciplines, de structurer un thésaurus de manière stricte selon l'un ou l'autre de ces schémas. Dans nos disciplines en tout cas on ne peut que mêler les deux approches. En effet, adopter une structure thématique rigoureuse, c'est s'exposer à devoir employer plusieurs fois les mêmes descripteurs dans des arbres hiérarchiques distincts. Dans les thèmes "audio-visuel" et "livre" par exemple, les sous-parties "sociologie" risquent de comprendre des descripteurs identiques (par exemple GOUT DU PUBLIC, COMPOSITION DU PUBLIC, INFLUENCE DE LA COMMUNICATION s'appliquent à l'un comme à l'autre). Il est donc préférable de créer une section "sociologie de la communication". Mais on ne retombe pas pour autant dans une organisation rigoureuse par disciplines, d'abord parce qu'il y a - à l'inverse de ce qui vient d'être montré - des notions communes à plusieurs disciplines, qu'il est donc commode de regrouper sous un "thème" ou encore une "facette". Par exemple la notion d'"information scientifique et technique" peut être utilisée dans le cadre de ce que J. MEYRIAT appelle la "*communicologie fonctionnelle*" ou encore en "informatologie". Dans un thésaurus elle pourrait figurer sous le thème "Information". De même, la notion de livre de poche intéresse autant la bibliologie que la bibliothéconomie : elle pourrait être rangée dans la facette "matériaux d'information". Ensuite, il faut reconnaître que les frontières entre les disciplines, en particulier dans un secteur en quête de définition comme celui des sciences de l'information et de la communication, ne sont pas toujours nettement tracées : faut-il rattacher la psychosociologie de la lecture à la bibliologie ou aux sciences sociales de l'information ? La bibliométrie est-elle une science autonome ou fait-elle partie de la bibliologie d'une part, de la bibliographie d'autre part ? Sans méconnaître l'intérêt que présentent sur le plan scientifique les tentatives de définition des disciplines, il vaut mieux prendre un certain recul par rapport à ces questions dans l'organisation d'un thésaurus, et y distinguer les domaines et thèmes qui s'avèrent les plus utiles à la recherche documentaire, même si leurs frontières ne coïncident pas avec celles des disciplines elles-mêmes.

3 - LE MODE DE PRESENTATION DU THESAURUS

a - L'INTRODUCTION

Le thésaurus sera nécessairement précédé d'une introduction qui en expliquera brièvement la raison d'être, la méthode d'élaboration, la structure et le mode d'emploi. Le contenu souhaité de l'introduction est précisé par la norme NF Z 47.100.

b - LES PARTIES DU THESAURUS

La facilité d'utilisation d'un thésaurus est conditionnée, entre autre, par son mode de présentation. Le minimum, pour un bon thésaurus, est une double présentation : alphabétique et systématique. Mais il y a diverses manières d'agencer et de présenter ces deux types de classement, dont l'interpénétration réciproque est souhaitable. Il me semble judicieux que nous adoptions pour notre projet de thésaurus une formule de plus en plus répandue qui est celle du thésaurus de l'UNESCO : la coexistence d'un "dictionnaire conceptuel" et de listes sous forme d'arbres hiérarchiques, complétée par un index permuté.

LE DICTIONNAIRE CONCEPTUEL

Le dictionnaire conceptuel donne dans l'ordre alphabétique les descripteurs et les termes équivalents, ceux-ci renvoyant seulement à ceux-là. Pour chaque descripteur, on trouvera dans l'ordre : la note d'application, le ou les termes équivalents, le terme générique immédiatement supérieur, le terme spécifique immédiatement inférieur, enfin le ou les termes associés. La nature de ces relations sémantiques doit être indiquée par un symbole placé devant chacune d'elle. On a le choix entre plusieurs modes de symbolisation (symboles préconisés par la norme AFNOR Z 47103, abréviations françaises, abréviations anglaises) si toutefois le mode d'édition du thésaurus ne nous impose pas une des deux dernières formules. Il me semble préférable de donner les relations hiérarchiques sur un seul niveau, afin de ne pas augmenter démesurément le volume du dictionnaire conceptuel. Il serait toutefois intéressant de pouvoir, suivant l'exemple du thésaurus de l'UNESCO, mentionner sous le terme générique immédiatement supérieur le terme générique du sommet de la hiérarchie à laquelle appartient le descripteur (cette relation devra bien sûr être signalée par un symbole particulier).

La présentation typographique me semble avoir aussi une certaine importance. Il est dommage de rebuter de prime abord l'utilisateur par une typographie qui ajoute à la difficulté de se familiariser avec le thésaurus un problème de déchiffrement. Tout en étant conscient que nous risquons d'être limités aux possibilités graphiques d'un listing d'ordinateur, je préfère présenter ici un projet un peu ambitieux qu'il serait souhaitable de réaliser autant qu'il se peut. Il serait utile de différencier par les caractères typographiques les termes équivalents (écrits par exemple en italique) des descripteurs (caractères romains). Il est impératif, évidemment que les termes génériques, spécifiques et associés soient en retrait par rapport au descripteur. Quant au terme équivalent rappelé dans la notice du descripteur, il faudrait, pour plus de clarté encore, le placer juste au dessous du descripteur, en italique, afin de le détacher des autres termes de l'environnement sémantique (voir à cet égard le thésaurus de la B.I.P.A.). Il est fondamental, enfin, qu'un code placé à droite du descripteur (ou au minimum un numéro de page) permette de retrouver celui-ci rapidement dans les arbres hiérarchiques.

LES ARBRES HIERARCHIQUES

L'édition séparée de ceux-ci est nécessaire, faute de quoi on serait obligé de consulter un à un les différents niveaux de la hiérarchie dispersés par l'ordre alphabétique du dictionnaire conceptuel. Les arbres hiérarchiques donnent, sous chaque descripteur constituant le sommet d'une hiérarchie, l'ensemble complet de cette hiérarchie. Les hiérarchies coïncident avec les domaines et sous-domaines distingués dans le thésaurus. On peut éventuellement faire figurer dans les arbres hiérarchiques les relations associatives les plus importantes (c'est le cas dans le thésaurus de l'UNESCO) : on a ainsi un équivalent de ce qui se fait dans la formule des schémas fléchés où les termes associés figurent soit dans le schéma lorsqu'ils appartiennent à la même hiérarchie, soit en marge du schéma lorsqu'ils appartiennent à d'autres hiérarchies.

Ici encore, la présentation typographique mérite d'être soignée. Il est nécessaire de distinguer les niveaux hiérarchiques par une structure en escalier éventuellement soulignée par l'emploi de caractères gras pour les têtes des hiérarchies. La structure peut être exprimée en outre, à la manière du thésaurus de l'UNESCO, par une notation hiérarchique ou semi-hiérarchique : à chaque descripteur est affecté un code qui désigne sans ambiguïté sa place dans l'arbre hiérarchique (c'est le même code qui renvoie du dictionnaire conceptuel aux arbres hiérarchiques). Il serait également souhaitable de faire précéder les arbres hiérarchiques de leurs sommaires : ceux-ci permettent d'avoir une vue synoptique de la structure du thésaurus. On n'oubliera pas enfin la présence, à côté des arbres hiérarchiques, de listes annexes qui peuvent être aussi dans la mesure du possible hiérarchisées : celle des mots-outils et celle des identificateurs.

Ajouter aux arbres hiérarchiques un mode de représentation graphique (schémas fléchés par exemple) me semblerait faire un peu double emploi et accroître la tâche plus que nécessaire. On aurait pu par contre préférer les schémas fléchés aux arbres hiérarchiques. Les premiers permettent en principe d'avoir instantanément une vue globale d'un groupe sémantique.

Mais ils présentent aussi quelques inconvénients qui ont déterminé mon choix : leur difficulté de réalisation et de mise à jour (à moins de pouvoir disposer d'un logiciel d'édition automatique de schémas fléchés) et le fait qu'il semble moins aisé de "couvrir" par leur moyen l'ensemble des domaines du thésaurus. Mais on pourrait envisager, au moins à des fins pédagogiques, de réaliser des schémas fléchés pour certains domaines du thésaurus qui s'y prêteraient mieux que d'autres.

L'INDEX PERMUTE

Les avantages de l'Index permuté ne sont plus à démontrer dans un thésaurus dont les descripteurs et les termes équivalents sont assez souvent des expressions composées et figurent dans le dictionnaire conceptuel dans l'ordre normal des termes : entrée alphabétique possible sur chacun des mots constituant un descripteur (à l'exception des mots vides), regroupement alphabétique des descripteurs sur une racine commune.

III

ÉLABORATION DU THÉSAURUS

ET

POLITIQUE D'INDEXATION

Il me semble important de constituer le thésaurus a priori, c'est-à-dire de commencer par établir la liste structurée et contrôlée des descripteurs au moyen de laquelle, dans un deuxième temps seulement, on s'appliquera à indexer, en vue de la recherche documentaire, les documents au fur et à mesure de leur acquisition (ainsi que les documents antérieurement acquis). En effet, commencer par indexer les documents à partir de rien, c'est-à-dire constituer d'emblée la mémoire documentaire avec l'intention de contrôler et de structurer la liste des descripteurs ainsi obtenus parallèlement à l'indexation ou à postériori, c'est courir le risque de devoir apporter de nombreuses corrections aux champs "descripteurs" des notices constituant la mémoire documentaire.

1 - RECHERCHE DE THÉSAURUS EXISTANTS

Quand on veut doter d'un thésaurus un centre de documentation donné, il faut avant toute chose se demander s'il n'existe pas déjà, en français ou dans une autre langue, un ou des thésaurus dans le même domaine (ou correspondant à une partie de celui-ci) : selon le cas, on pourrait l'adopter, le reprendre en partie ou s'en inspirer. Notre cas a ceci de particulier que le domaine à couvrir, comme cela a été montré en première partie, ne correspond pas tout à fait à une science ou une spécialité bien définie. Il y a donc de fortes chances pour que nous ne trouvions pas de thésaurus sectoriel dont la couverture sémantique réponde exactement à nos besoins. Nous rencontrerons en gros quatre sortes de thésaurus : ceux qui sont spécialisés en bibliothéconomie et sciences de l'information et de la documentation et qui ne couvrent donc pas, ou insuffisamment, les autres do-

maines de l'information et de la communication (a fortiori des secteurs comme l'administration ou l'éducation) ; des thésaurus limités à un secteur précis des sciences de la communication (par exemple les médias autres que le livre, la communication de masse, les types de documents) ; des thésaurus qui traitent d'un des domaines connexes ou marginaux représentés dans nos bibliothèques (informatique, éducation ...) ; enfin des thésaurus à caractère encyclopédique dont une grande partie ne nous concernera pas et qui risquent par contre d'être un peu pauvres dans certains des domaines qui nous intéressent. Il faut cependant rechercher de tels thésaurus car ils peuvent éventuellement nous être utiles, tant en matière de choix des descripteurs qu'en matière d'organisation hiérarchique.

Il existe des bibliographies de thésaurus. Un certain nombre d'entre elles sont recensées (jusqu'en 1980-1981) dans l'ouvrage de M.T. LAUREILHE sur les thésaurus. J'ai pu consulter trois d'entre elles (réf. bibliographiques n°14 , n°15 , n°17) plus une qui n'y figure pas (réf. bibliographique n°16), qui a l'avantage d'être un peu plus récente que les autres (1982) et de comporter des notices parfois analytiques.

En matière de thésaurus étrangers, j'ai essayé de me procurer quelques uns des ouvrages signalés dans ces bibliographies, ceux qui me paraissaient les plus pertinents : thésaurus structurés, les moins anciens possibles, spécialisés en documentation et en bibliothéconomie, et classifications spécialisées (1). Je me suis limité pour des raisons de connaissances personnelles, aux documents de langue anglaise ou allemande. Il ne s'agissait en effet pas pour moi de faire une recherche exhaustive mais seulement de tester à titre exemplaire les difficultés que nous pourrions rencontrer. Il se trouve en outre que certains thésaurus de langue française que nous possédons (dont il sera question plus loin) constituent déjà des bases de travail tout à fait appréciables. Il serait toutefois souhaitable, si l'on en a les moyens, de mener les recherches un peu plus loin que je ne l'ai fait.

(1) Celles-ci peuvent être intéressantes par les termes qu'elles contiennent et surtout par la structure qu'elles présentent

Dans ces recherches, je me suis heurté à une difficulté d'ordre matériel. Sur les neuf références retenues dans les bibliographies (trois en allemand, six en anglais), après en avoir éliminé une qui s'est avérée inexacte, je n'ai trouvé que trois thésaurus dans les bibliothèques françaises (B.N., B.M. de Lyon, B.P.I., E.N.S.B., I.N.T.D., D.B.M.I.S.T., C.D.S.T.). Au centre de documentation de la D.B.M.I.S.T., cinq références sur les neuf se trouvent dans les catalogues, mais trois thésaurus seulement sont présents sur les rayons. Deux d'entre eux (1) ne semblent pas présenter d'intérêt particulier pour nous. Quant au troisième (2), c'est celui qui a servi de base à la structuration d'une partie de la section "Information, bibliothèques et archives" du thésaurus de l'UNESCO dont il sera question ci-dessous. Pour les cinq documents que je n'ai pas pu trouver en France, j'ai envoyé des demandes de prêt-interbibliothèques à la B.L.L.D. et en Allemagne dans plusieurs bibliothèques. De la B.L.L.D. me sont parvenues deux réponses négatives (3) (dont une après réclamation) et un document incomplet (4) (une brève mise à jour de la classification, au lieu de la classification elle-même). D'Allemagne, j'ai reçu une réponse négative (5) et une classification documentaire de peu d'intérêt pour nous (6).

Quant aux thésaurus de langue française, ils sont peu nombreux dans les disciplines qui nous intéressent. Dans le domaine de la bibliothéconomie et des sciences de l'information et de la documentation, je n'ai trouvé dans les bibliographies mentionnées que deux références vraiment pertinentes : le thésaurus bilingue (anglais/français) Subject descriptors to be used in ISORID de la Fédération internationale de documentation, paru en 1972 (je ne l'ai pas consulté mais il constitue l'une des sources du thésaurus de l'UNESCO, plus récent) et le Thésaurus sciences de l'information du Centre de documentation scientifique et technique du C.N.R.S., édité en 1977. Il faut toutefois y ajouter, bien qu'il soit à caractère encyclopédique, le thésaurus de l'UNESCO dont la version anglaise,

(1) Réf. bibliographiques n°25 et n° 29

(2) Réf. bibliographique n°23

(3) Pour les réf. bibliographiques n°26 et n°28

(4) Correspondant à la réf. bibliographique n°18

(5) Pour la réf. bibliographique n°22

(6) Réf. bibliographique n°27

établie par Jean AITCHISON, est parue en 1977 et la version française (traduction de Jean VIET) en 1983 sous le titre : Thésaurus de l'UNESCO : liste structurée de descripteurs pour l'indexation et la recherche bibliographiques dans les domaines de l'éducation, de la science, des sciences sociales, de la culture et de la communication. Ce thésaurus nous intéresse dans la mesure où, bien qu'offrant "un aperçu général des grands domaines de la connaissance", il traite "en détail les sujets auxquels l'UNESCO s'intéresse plus particulièrement : (...) l'éducation, la communication, la science de l'information et la bibliothéconomie, la culture et certaines branches des sciences sociales et politiques, ainsi que certaines sciences de la terre ..." (Introduction p. X).

Le thésaurus de l'UNESCO ne nous intéresse pas seulement au titre de la bibliothéconomie et des sciences de l'information. A part lui, pour les autres domaines représentés dans nos fonds documentaires, je citerai seulement ici trois thésaurus qui pourront éventuellement nous aider : le Thésaurus : mass communication ... (anglais/français/espagnol) de Jean VIET, le thésaurus de la B.I.P.A., d'orientation assez diverse, et le Thésaurus d'informatique du C.N.R.S.

2 - LE THÉSAURUS DE L'UNESCO ET LE THÉSAURUS DES SCIENCES DE L'INFORMATION DU C.N.R.S.

En attendant de pouvoir éventuellement consulter des thésaurus étrangers, il s'agit de se demander dans quelle mesure les deux thésaurus de langue française qui semblent présenter le plus d'intérêt pour nous, le thésaurus de l'UNESCO et celui du C.N.R.S., répondent, sur le plan sémantique, à nos besoins. Leur pertinence doit être évaluée à quatre niveaux qui correspondent aux quatre étapes principales de l'élaboration d'un thésaurus (1) :

(1) On reviendra sur celles-ci dans le chapitre 3.

couverture sémantique des domaines, c'est-à-dire correspondance entre l'ensemble des notions représentées dans les thésaurus en question et l'ensemble des notions dont nous aurons effectivement besoin en matière de recherche documentaire ; choix et définition des descripteurs ; organisation conceptuelle générale du thésaurus ; relations hiérarchiques et associatives. Dans la pratique, cette évaluation se fera par nécessité, de manière précise, pendant la phase même d'élaboration de notre thésaurus, mais il est préférable ici, pour la clarté de l'exposé, de présenter les éléments séparément. On se contentera toutefois de quelques observations rapides, mais significatives, fondées sur des exemples précis.

COUVERTURE SEMANTIQUE DES DOMAINES

La bibliothéconomie et les sciences de l'information et de la documentation sont abondamment représentées dans les deux thésaurus et plus encore dans le thésaurus de l'UNESCO (Section Z). Par exemple, les diverses notions liées au prêt de documents sont exprimées par seize descripteurs dans ce thésaurus, deux descripteurs seulement dans le thésaurus du C.N.R.S. (PRET et PRET INTERBIBLIOTHEQUES). Le terme "catalogage" est présent dans treize descripteurs du thésaurus de l'UNESCO, six descripteurs du thésaurus du C.N.R.S. . Ce développement poussé du thésaurus de l'UNESCO est certes appréciable : on y trouve par exemple la distinction - non faite par le C.D.S.T. - entre INDEXATION PAR MATIERES et INDEXATION-CLASSIFICATION. Mais il a aussi pour nous des inconvénients. Des descripteurs comme PRETS VENUS A ECHEANCE , RENOUVELLEMENT DU PRET , TERMES COMPOSES (alors qu'il existe PRE-COORDINATION) , RELATION OU auront-ils leur utilité dans notre thésaurus ? Il est permis d'en douter. Il faut noter que toutes les questions relatives au traitement de l'information sont analysées particulièrement en détail (environ 290 descripteurs, c'est-à-dire une centaine de plus que dans le thésaurus du C.N.R.S.).

La situation est nettement moins favorable pour nous dans la plupart des autres disciplines : dans les deux thésaurus, des notions importantes ne sont pas représentées. Les causes en sont claires, bien que paradoxales. D'un côté (C.N.R.S.) une spécialisation plus étroite que celle de nos bibliothèques réunies ; de l'autre côté, une relative non-spécialisation : étant donnée son ambition encyclopédique, le thésaurus de l'UNESCO est obligé dans certaines disciplines de se maintenir à un certain niveau de généralité. On peut illustrer à l'aide de quelques exemples cette insuffisance du thésaurus de l'UNESCO (celle du thésaurus du C.N.R.S. est encore plus flagrante). Dans le domaine de l'administration manquent par exemple les notions : commune, département, déconcentration, fonction publique communale, statut général de la fonction publique, établissement public, etc ... Dans les sciences de l'informatique, on sent que le thésaurus date déjà de sept ou huit ans : on n'y trouve pas entre autre micro-ordinateur, bureautique, robotique, télématique, vidéotex, progiciel. En médiologie, en bibliologie, en sciences sociales de l'information et en technologie de la communication, le thésaurus se révèle également insuffisant pour nous. La section Y (communication) qui correspond à ces disciplines, occupe neuf pages, c'est-à-dire quatre fois moins que la section Z. Sont absentes par exemple les notions : édition à compte d'auteur, vente par correspondance, vente par courtage, offset, machine à écrire ... Par contre, en sciences de l'éducation, de même que par exemple dans les matières culturelles et littéraires, le thésaurus de l'UNESCO est sur-développé par rapport à nos besoins.

CHOIX ET DEFINITION DES DESCRIPTEURS

Il semble qu'à ce sujet le thésaurus de l'UNESCO nous réserve plus de surprises que celui du C.N.R.S. (1) au moins en ce qui concerne la bibliothéconomie et les sciences de l'information et de la documentation.

(1) Le thésaurus du C.N.R.S. ne contient de toute façon pas de définition explicite des descripteurs comme celui de l'UNESCO.

Peut-être cela est-il dû à l'évolution du vocabulaire ou encore au passage éventuel de l'anglais au français : la terminologie de WERSIG et NEVELING, d'où sont extraits les termes et les définitions contenus dans la section Z (sauf pour l'archivistique), date de 1976 ; elle est multilingue. Voici quelques exemples des difficultés qui se présentent : le terme COLLATIONNEMENT est pris au sens de "fusion de fichiers" et non au sens courant de contrôle de la conformité d'un document ou d'un message. Le terme "photogravure" renvoie à IMPRESSION EN CREUX ; dans le thésaurus du C.N.R.S., il renvoie, ce qui est acceptable, au terme plus général IMPRESSION. La notion courante de recherche documentaire automatisée, exprimée dans le thésaurus du C.N.R.S. par le descripteur RECHERCHE AUTOMATISEE (spécifique de RECHERCHE DOCUMENTAIRE) n'a pas d'équivalent exact dans le thésaurus de l'UNESCO. Le descripteur le plus proche RECHERCHE D'INFORMATION EN LIGNE est un peu trop fin ; on trouve sur le même plan RECHERCHE PAR LOTS, notion également incluse dans l'idée de recherche documentaire automatisée. Le terme générique commun à ces deux descripteurs, INTERROGATION, est par contre un peu trop vaste : il comprend par exemple la notion de "bouquinage". La définition du descripteur BASES DE DONNEES est floue. Quant au descripteur BANQUES DE DONNEES, il est assorti de deux définitions dont l'une présente une obscurité syntaxique et l'autre assimile d'une manière peu explicable la banque de données à une institution. Derniers exemples : la notion de référence bibliographique (C.N.R.S. : REFERENCE BIBLIOGRAPHIQUE) est classée, comme terme rejeté, sous le descripteur CITATIONS défini de façon peu claire. Enfin, on ne trouve pas le descripteur BIBLIOTHEQUE CENTRALE DE PRET (qui figure par contre dans le thésaurus du C.N.R.S.) mais celui de BIBLIOTHEQUES DEPARTEMENTALES. On peut toutefois justifier ce dernier choix et peut-être aussi le précédent, en invoquant le caractère encyclopédique du thésaurus qui l'oblige à sacrifier certains termes fins. De la même manière, ROMANS POLICIERS et ROMANS D'ESPIONNAGE renvoient à ROMANS A SENSATION.

En matière de définition des descripteurs, il faut encore mentionner un petit problème que pose le thésaurus de l'UNESCO. Les définitions de la terminologie de WERSIG et NEVELING y sont reprises telles quelles, même lorsqu'elles sont doubles voire triples. Il en résulte quelques cas de

polysémie non compatibles avec un thésaurus. Pour ne prendre qu'un exemple, le descripteur INDEXATION-CLASSIFICATION peut signifier d'une part "Action de préparer une classification", d'autre part "Action d'indexer avec un plan de classification". Il s'agit là de deux opérations distinctes qui devraient, dans un thésaurus, être représentées par deux descripteurs.

ORGANISATION CONCEPTUELLE GLOBALE

Au niveau du regroupement des descripteurs par domaines et sous-domaines, le thésaurus du C.N.R.S. n'entre pas pour nous en ligne de compte du fait de sa structure par facettes. Le thésaurus de l'UNESCO me paraît dans l'ensemble satisfaisant dans notre optique d'utilisation, en particulier la section Z dont la subdivision générale est : "Information" "Transfert d'information" "Matériaux d'information" "Bibliothéconomie et sciences de l'information" "Archives". Quant à la sous-structure de cette section, elle ne semble pas non plus à première vue, présenter d'inconvénient. La section Y (communication) est aussi clairement structurée : elle se décompose en gros en sciences sociales de l'information et de la communication, processus de communication, moyens de communication (principalement mass-média), technologie des communications, industrie des communications. Elle a évidemment l'inconvénient de disperser un peu une discipline comme la bibliologie mais on peut difficilement y échapper étant donné la diversité des facettes qui composent celle-ci.

Dans les domaines marginaux ou connexes, la facilité d'utilisation du thésaurus est variable. On peut tout au moins repérer aisément dans la structure générale du thésaurus des secteurs peu représentés dans nos bibliothèques comme les mathématiques, la linguistique et la littérature (constituant à elles deux la "philologie") et l'action culturelle. L'informatique se détache, elle aussi, clairement dans la division "Sciences et technologie interdisciplinaires". L'éducation forme une division à elle toute seule, solidement charpentée. Par contre, les notions diverses que nous rassemblons

sous le concept un peu large d'administration sont malheureusement dispersées dans des sous-sections distinctes : "Science politique" (par exemple pour l'administration publique), "Finances" (pour l'administration financière et la comptabilité), "Administration, travail et emploi" et même "Environnement humain" (pour les collectivités locales). A cet égard, le caractère encyclopédique du thésaurus redevient gênant pour nous. Le même problème se pose par exemple pour les questions relatives au cinéma, disséminées en "Arts du spectacle", "Arts de la reproduction", "Culture", ainsi que dans le domaine des sciences sociales de l'information et de la communication : celles-ci sont représentées de manière spécifique dans la section Y mais il faut y ajouter des descripteurs classés en sociologie, en psychologie et en droit.

RELATIONS HIERARCHIQUES

Lorsqu'un tel problème se pose au niveau de l'organisation conceptuelle générale, il peut également apparaître à celui des relations hiérarchiques entre descripteurs. Pour ne prendre que deux exemples, on trouve dans le thésaurus de l'UNESCO la hiérarchie

ETABLISSEMENTS HUMAINS

UF Environnement communautaire

TS COLLECTIVITES

TS COLLECTIVITES LOCALES

moins pertinente pour nous que par exemple celle-ci, prise dans le thésaurus de la B.I.P.A. :

ADMINISTRATION

TS COLLECTIVITES LOCALES

TS COMMUNE

Ou encore, thésaurus de l'UNESCO :

SYSTEMOLOGIE

TS CYBERNETIQUE

TS BIONIQUE

TS ERGONOMIE

Thésaurus de la B.I.P.A. :

TRAVAIL

TS CONDITIONS DE TRAVAIL

TS ORGANISATION DU TRAVAIL

TS ERGONOMIE

Mais même dans des domaines qui ne présentent pas l'inconvénient d'une certaine dispersion des notions, on rencontrera quelques difficultés, parfois liées aux définitions qui accompagnent les descripteurs. Ainsi, BIBLIOTHEQUES DEPOSITAIRES est considéré comme terme spécifique de BIBLIOTHEQUES NATIONALES, ce qui peut se justifier si l'on s'en tient au premier sens donné dans la note d'application mais peut s'avérer inexact si on adopte le deuxième sens : toutes les "bibliothèques où des documents sont déposés continuellement sous certaines conditions" ne sont pas des bibliothèques nationales. De même INDEX est défini comme une "liste ordonnée de références au contenu d'un fichier ou d'un document (...)" mais on lui trouve entre autres comme termes spécifiques BASES DE DONNEES et aussi, ce qui est un peu contestable, ARRANGEMENT SUR LES RAYONS. Il semble toutefois qu'il s'agisse là de cas isolés. En tout cas les relations sémantiques entre descripteurs sont nettement plus développées dans le thésaurus de l'UNESCO que dans celui du C.N.R.S.

3 - MODE D'ÉLABORATION DU THÉSAURUS

Après qu'a été défini le domaine à couvrir, que les concepteurs du thésaurus se sont mis d'accord sur les caractéristiques formelles de celui-ci et que l'on a réuni les thésaurus qui peuvent être mis à profit, la construction elle-même du thésaurus peut commencer.

Il ne me paraît pas indispensable de déterminer à priori le volume du thésaurus (nombre de descripteurs), d'autant moins que les formules mathématiques nécessaires à ce calcul supposent que soient connues des valeurs parfois difficiles à établir comme le nombre de documents, le nombre moyen de descripteurs affectés à un document, voire le nombre moyen de documents répondant à une question.

La norme AFNOR NF Z 47.100 ne fait pas mention de cette pratique peu répandue.

Il conviendra de suivre les indications précises données par la norme en matière de construction de thésaurus. On peut rappeler dans leur succession logique les principales étapes de ce travail : compilation du vocabulaire (collecte des candidats-descripteurs), réduction du vocabulaire (sélection des descripteurs parmi les synonymes mais aussi éventuellement rajout de termes équivalents non collectés lors de la première étape), normalisation du vocabulaire (règles d'écriture), structuration du vocabulaire : cette dernière opération consiste dans un premier temps à regrouper les descripteurs par domaines ou par thèmes, dans un deuxième temps, à établir les relations hiérarchiques et associatives entre les descripteurs.

Au cours de ces différentes phases du travail, les thésaurus existants devront être utilisés de la manière décrite ci-dessous.

Pour la collecte des candidats-descripteurs, on combine en général deux méthodes : la méthode synthétique qui consiste à rechercher à priori les termes significatifs dans des listes lexicales diverses (fichiers, terminologies, index, thésaurus ...) et la méthode analytique qui consiste à obtenir les mots significatifs en indexant en langage libre un échantillon de documents représentatifs du fonds, ainsi qu'un échantillon de questions posées par des utilisateurs spécialistes du domaine. En fait de listes lexicales, nous disposons de nos fichiers (catalogues-matières, index BELETTE, liste des descripteurs utilisés par l'I.N.T.D.), de terminologies et d'index divers et bien entendu des thésaurus de l'UNESCO et du C.N.R.S. que nous pourrions compléter dans les domaines qui y sont insuffisamment représentés par d'autres thésaurus, comme celui de la B.I.P.A. pour les questions administratives, le Thésaurus : mass communication pour une partie des sciences relatives à la communication, voire le Thésaurus d'informatique du C.N.R.S. . Dans la mesure où les domaines de notre thésaurus sont assez bien circonscrits et où nous possédons des listes lexicales très fournies, il est plus judicieux pour nous de partir de ces listes et de les adapter par la méthode analytique au contenu réel des documents du fonds, et à celui des questions. Il s'agira donc d'une part de rajouter les notions manquantes, d'autre part de supprimer les notions jugées inutiles ou superflues (parce que trop fines, par exemple).

Les thésaurus cités ci-dessus nous seront également utiles lors du choix des descripteurs, puis lors de l'établissement des relations hiérarchiques et associatives. Il s'agira de les comparer entre eux, dans leurs domaines respectifs : nous retiendrons les solutions qui nous paraissent les meilleures. Mais nous devons faire preuve d'imagination (et de rigueur) pour tous les cas où les thésaurus existants présenteront des solutions qui ne nous satisferont pas, ou n'en présenteront pas du tout.

Pour la détermination des domaines du thésaurus et de leurs principales subdivisions, il me semble souhaitable de prendre pour base la structure du thésaurus de l'UNESCO (1), avec les réserves qui découlent de

(1) On pourra également consulter les classifications documentaires encyclopédiques, mais elles semblent moins intéressantes pour nous à cet égard.

l'analyse faite au chapitre 2. Il faudra de toute façon s'assurer de la pertinence de l'organisation générale des différents domaines. Dans les domaines centraux (correspondant aux sections Z et Y), il faudra procéder à quelques regroupements et suppressions de rubriques. Certains des domaines connexes ou marginaux devront être réorganisés. Pour la disposition logique des grands domaines constituant le thésaurus, il me semble intéressant de conserver à l'esprit une tentative de classification des sciences de l'information et de la communication comme celle qui a été présentée dans le chapitre 2 de la première partie : elle met en évidence certaines relations qui n'apparaissent pas clairement dans le thésaurus de l'UNESCO (mais il ne s'agit évidemment pas de la reprendre telle quelle, pour les raisons mentionnées dans le chapitre 2 (b) de la deuxième partie).

Le léger surcroît de travail que représente la constitution d'un thésaurus commun à trois bibliothèques est compensé par la possibilité de répartir les tâches entre les parties concernées. Les trois bibliothèques peuvent s'attribuer chacune d'un commun accord un certain nombre de domaines dans lesquels elles collecteront les candidats descripteurs, choisiront les descripteurs et établiront les relations sémantiques. Ces deux dernières opérations ainsi que la structuration globale du thésaurus devront évidemment être soumises à examen général. Il est souhaitable que, pour la répartition des domaines entre les bibliothèques, on tienne compte des spécialisations relatives de chacune d'elles (voir I, 2, b).

Il ne faudra pas négliger, en dernier lieu, les essais du thésaurus ni sa mise à jour ultérieure qui devra être faite régulièrement en fonction de l'évolution du vocabulaire et des besoins de la recherche documentaire. Il serait utile à ce propos, de compléter les recommandations de la norme AFNOR par celles que donne J. CHAUMIER dans Analyse et Langages documentaires : elles sont plus précises en matière de maintenance et d'évaluation du thésaurus. CHAUMIER conseille entre autre, à la différence de la norme, de tester le vocabulaire avant même sa structuration, en le comparant à une liste de descripteurs provenant de l'indexation libre, par deux personnes n'ayant pas participé à l'élaboration du thésaurus, d'un échantillon de documents et de questions : au moins 75 % du vocabulaire des tests

doit se retrouver dans le vocabulaire du thésaurus. On évite ainsi de devoir rajouter des descripteurs au thésaurus après que les relations sémantiques ont été établies. Cela ne permet pas pour autant de faire l'économie des essais finals.

Le thésaurus une fois achevé, doit être édité, de préférence selon la présentation en trois parties présentée ci-dessus. On aura probablement recours, pour ce faire, à l'ordinateur qui nous épargnera certaines tâches répétitives comme l'établissement des relations réciproques. Il serait intéressant d'employer un logiciel qui soit capable de fabriquer l'index permuté.

Mais le choix du logiciel est surtout lié à la manière dont on exploitera le thésaurus pour l'indexation et la recherche documentaire.

4 - POLITIQUE D'INDEXATION

Le thésaurus doit permettre le stockage sous une forme condensée de l'information contenue dans les documents et le "rappel" de cette information au gré des utilisateurs (de préférence avec l'aide des techniques informatiques). Il s'agira donc, à partir du moment où l'on disposera d'un thésaurus achevé, d'indexer les documents des trois fonds à l'aide de ce thésaurus ; ce qui signifie principalement : indexer les documents qui rentrent (livres et articles de périodiques), au fur et à mesure de leur acquisition. L'E.N.S.B. devrait également inclure le contenu de ses fichiers BELETTE dans la nouvelle mémoire documentaire, ce qui suppose une réindexation des documents selon le nouveau thésaurus. On pourrait toutefois se limiter à ce qui paraît le plus important. L'I.N.T.D. pourra par con-

tre, à des fins pédagogiques, conserver ses fichiers actuels (SPHINXO et UNITERM), se contentant de les clore. L'idéal serait qu'ensuite - mais le manque de temps et de personnel s'y opposera probablement - on enregistre aussi dans la mémoire documentaire les livres qui font déjà partie du fonds des trois bibliothèques. Cela aurait surtout son importance à l'I.N.T.D., du fait de l'absence d'un catalogue alphabétique de matières, mais serait intéressant aussi dans les deux autres bibliothèques dans la mesure où le thésaurus combinant ses possibilités à celles de l'ordinateur devrait permettre une recherche documentaire plus fine et plus complète que les catalogues-matières.

La finesse et l'exhaustivité de la recherche documentaire dépendent en particulier d'un paramètre variable que nous devons essayer de fixer d'emblée : l'étendue (ou exhaustivité) de l'indexation, qui n'est pas sans rapport avec la spécificité de l'indexation. L'étendue de l'indexation se mesure au rapport quantitatif entre les informations contenues dans le document et celles qu'exprime l'indexation ; la spécificité de l'indexation exprime la *"mesure dans laquelle le terme d'indexation attribué correspond au concept traité dans le document"* (1). (la règle générale est d'indexer au niveau le plus spécifique). La question se posera pour les articles de périodiques - on constate ainsi que le nombre moyen de descripteurs attribués à un document par l'I.N.T.D. est supérieur au même nombre moyen dans le fichier BELETTE - mais surtout pour les livres, du moins pour un grand nombre d'entre eux. Assurément, devant un recueil de statistiques brutes sur les B.C.P., on n'hésitera pas longtemps (encore qu'on puisse se demander s'il ne faut pas faire ressortir dans l'indexation certaines des fonctions bibliothéconomiques sur lesquelles portent les statistiques - par exemple le prêt). Mais on peut avoir affaire à une réflexion assez générale qui porte sur une douzaine de points d'importance égale (la presse, la télévision, la télématique, les banques de données, etc ...). On peut lui attribuer douze descripteurs voire plus, encore faut-il avoir préalablement défini les règles à suivre dans un pareil cas. Un cas extrême est constitué par les recueils de textes traitant de notions trop nombreuses

(1) Définition extraite de la terminologie de WERSIG.

pour que l'on puisse les exprimer toutes mais dont le dénominateur commun est si vaste qu'il traduit mal à lui seul le contenu de l'ouvrage.

L'indexation peut également être caractérisée par quelques procédés destinés à améliorer l'efficacité du système de recherche documentaire. Il en est deux au moins que nous pourrions utiliser si toutefois le logiciel que nous aurons choisi le permet et si nous l'estimons nécessaire : la pondération qui permet, par l'affectation d'un poids aux descripteurs, de préciser l'importance relative des notions relevées dans le document ; et l'"autopostage", procédé qui permet de récupérer automatiquement, lors d'une interrogation sur un terme assez général (par exemple : OUVRAGE D'ART), les documents également pertinents indexés par des termes qui lui sont spécifiques (par exemple PONT). La manière dont est pratiqué l'autopostage (manuelle ou automatique, vers le haut ou vers le bas) affecte le processus d'indexation (1). Quant aux liens et aux indicateurs de rôle qui, introduisant entre les descripteurs des relations syntaxiques, évitent que ne se produisent entre ceux-ci des combinaisons erronées, ils sont d'utilisation plus délicate : il est conseillé de ne les adopter pour un système documentaire qu'en cas de réelle nécessité.

Notre politique d'indexation sera enfin déterminée par les rapports que nous concevrons entre les trois bibliothèques à ce niveau. Il serait intéressant de pouvoir mettre en commun non seulement l'outil d'indexation et de recherche documentaire qu'est le thésaurus, mais aussi le contenu même de la mémoire documentaire. D'une part on permettrait ainsi aux utilisateurs d'avoir en même temps accès lors de la recherche documentaire aux fonds des trois bibliothèques. D'autre part, on pratiquerait alors une indexation partagée (les caractéristiques citées ci-dessus auraient donc été définies d'un commun accord). Le plus simple serait sans doute, puisque les trois bibliothèques disposent (ou disposeront) d'un terminal, d'utiliser les services du S.U.N.I.S.T. (2). Sur l'ensemble des titres de

(1) Sur l'autopostage, voir CHAUMIER (Jacques) Analyse et langages documentaires, p. 48-49.

(2) Serveur Universitaire National pour l'Information Scientifique et Technique.

périodiques dont le dépouillement est jugé utile, chaque bibliothèque dépouillerait les périodiques qu'elle est seule à recevoir et une répartition aurait lieu pour les titres reçus à la fois par les trois bibliothèques ou par deux d'entre elles (1). Les bibliothèques procèderaient après indexation à la saisie des notices sur leur terminal respectif. En ce qui concerne les livres, on peut imaginer un système analogue à celui pratiqué par certaines banques de données. Dans chaque bibliothèque, l'indexeur, avant de se mettre à l'indexation d'une série d'acquisitions récentes, interrogerait l'ordinateur pour savoir si certains de ces documents ont déjà été enregistrés par une autre bibliothèque. Pour chaque réponse négative, il devrait lui-même indexer le document et en enregistrer la notice. Dans le cas des livres déjà entrés en mémoire, il lui suffirait d'ajouter le code de sa bibliothèque. (Pour la localisation des périodiques, on peut se contenter des listes des périodiques reçus par telle bibliothèque à condition qu'elles mentionnent l'état de collection et qu'elles soient tenues à jour).

Ce système pose cependant quelques problèmes. Selon le mode actuel d'organisation du S.U.N.I.S.T., lorsque l'on accède grâce au mot de passe à ses propres fichiers, on peut tout aussi bien les interroger que les modifier. Cela veut dire que dans les trois bibliothèques, qui posséderaient nécessairement la clef d'accès aux fichiers communs, n'importe quel utilisateur pourrait s'il le voulait et s'il en était capable, effacer ou modifier une partie de la mémoire. Une solution consisterait à disposer comme les banques de données de deux modules distincts : l'un pour l'interrogation, l'autre pour la mise à jour. Mais il n'est pas certain que cette formule puisse être réalisée par le S.U.N.I.S.T. . Ensuite, se poserait le problème de la facturation. Il faudrait demander au S.U.N.I.S.T. s'il lui est possible de déterminer la provenance de chaque appel afin que la facturation puisse être partagée entre les trois bibliothèques.

(1) Les trois bibliothèques ensemble reçoivent plus de trente périodiques identiques. Si on les considère deux par deux, le nombre de titres communs est encore plus élevé.

Au cours de cette analyse sont apparus quelques-uns des critères qui devront être pris en compte dans le choix du logiciel destiné à l'exploitation informatique du thésaurus. Cette question du choix du logiciel de recherche documentaire ne sera pas traitée ici. Je voudrais seulement souligner les avantages offerts par le logiciel TEXTO exploité à l'aide du système LOGOTEL. Ce logiciel fait partie de ceux qui sont actuellement employés au S.U.N.I.S.T.. Il offre en matière de Thésaurus de nombreuses possibilités, (opérations de manipulation du vocabulaire, autopostage, pondération, interrogation possible sur tous les champs de la notice etc ...), actuellement étudiées au centre de calcul du C.N.R.S. à Strasbourg et dont certaines sont mises en application à la Maison de l'Orient à Lyon. Enfin, le système LOGOTEL se caractérise par sa facilité d'utilisation : une fois les paramètres fixés à l'initialisation, n'importe quel non-spécialiste peut s'en servir.

C O N C L U S I O N

Il est évidemment plus simple de décrire sur le papier le fonctionnement idéal d'un système que d'en mener la réalisation concrète. Les possibilités et les orientations présentées dans ce mémoire seront -on peut du moins l'espérer- mises à l'épreuve des faits. De faits parfois difficiles à maîtriser. Les trois bibliothèques devront organiser leur collaboration d'une manière harmonieuse et équitable. L'élaboration d'un thésaurus représente un travail assez considérable : le premier problème qui se posera sera peut-être celui de la disponibilité du personnel. On peut certes envisager, à l'E.N.S.B. et à l'I.N.T.D., de faire appel aux élèves qu'intéresserait un tel travail, qui pourrait tenir lieu de mémoire de fin d'études. Mais il serait de toute façon souhaitable que les professionnels des trois établissements eux-mêmes participent, au moins à certains niveaux, à la construction du thésaurus : celle-ci exige une connaissance sûre du domaine considéré et une expérience confirmée de l'indexation. Quoi qu'il en soit, le projet mérite que l'on s'y intéresse. Les difficultés liées à sa réalisation seront à mon avis compensées par les nombreux avantages qu'il présente : gain en efficacité et en possibilités de recherche documentaire, coopération plus étroite entre les bibliothèques, intérêt pédagogique, expérimentation éventuelle d'un logiciel encore mal connu comme TEXTO...

B I B L I O G R A P H I E

Ne sont signalés ici que les documents qui ont été utiles à la rédaction du mémoire ou auxquels il est fait référence dans le cours du texte. On trouvera des références bibliographiques plus nombreuses dans les ouvrages de M. T. LAUREILHE et de J. CHAUMIER mentionnés ci-dessous (réf. n) 10 et réf. n°6).

Sous chaque rubrique les documents sont classés par ordre alphabétique des noms d'auteurs et sous-classés par ordre alphabétique des titres.

NORMES

- 1 - ASSOCIATION FRANCAISE DE NORMALISATION. Documentation : Principes généraux pour l'indexation des documents : norme française enregistrée NF Z 47-102. Paris : AFNOR, 1978.
- 2 - ASSOCIATION FRANCAISE DE NORMALISATION. Documentation : Règles d'établissement des thésaurus monolingues : norme française homologuée NF Z 47-100. Paris : AFNOR, 1981.
- 3 - ASSOCIATION FRANCAISE DE NORMALISATION. Documentation : Thésaurus monolingues et multilingues : symbolisation des relations : norme expérimentale Z 47-103. Paris : AFNOR, 1980.
- 4 - ORGANISATION INTERNATIONALE DE NORMALISATION. Documentation : Principes directeurs pour l'établissement et le développement de thésaurus monolingues : norme internationale ISO 2788-1974 (F). Genève : ISO, 1974.

ETUDES DIVERSES

- 5 - BRETON (Jacques). Bibliologie, communicologie et bricolage idéologique. In : Schéma et schématisation, 1983, n° 19, p. 75-84.
- 6 - CHAUMIER (Jacques). Analyse et langages documentaires : le traitement linguistique de l'information documentaire. Paris : Entreprise moderne d'édition, 1982.

- 7 - CHAUMIER (Jacques). Les Techniques documentaires. 2° éd. Paris : Presses universitaires de France, 1974.
- 8 - ESTIVALS (Robert). Les sciences bibliologiques. In : Schéma et schématisation, 1981, n°15, p. 21-35.
- 9 - GUINCHAT (Claire) et MENOUE (Michel). Introduction générale aux sciences et techniques de l'information et de la documentation. Paris : Presses de l'Unesco, 1981.
- 10 - LAUREILHE (Marie-Thérèse). Le Thésaurus : son rôle, sa structure, son élaboration. 2° éd. Villeurbanne : Presses de l'E.N.S.B., 1981.
- 11 - MEYRIAT (Jean). De la science de l'information aux métiers de l'information. In : Schéma et schématisation, 1983, n°19, p. 65-74.
- 12 - MEYRIAT (Jean). L'informatologie science soeur de la bibliologie. In : Schéma et schématisation, 1981, n°15, p. 9-19.
- 13 - MEYRIAT (Jean). Pour une classification des sciences de l'information et de la communication. In : Schéma et schématisation, 1983, n°19, p. 61-64.

BIBLIOGRAPHIES DE THESAURUS

- 14 - BUREAU INTERORGANISATIONS POUR LES SYSTEMES D'INFORMATION. Bibliography of United Nations : Thesauri, classifications, nomenclatures. Geneva, 1979.
- 15 - BUREAU NATIONAL D'INFORMATION SCIENTIFIQUE ET TECHNIQUE. Annuaire des thésaurus : suivi d'un index des domaines et sous-domaines couverts par les thésaurus et d'une liste des organismes auteurs. 3° éd. Paris : B.N.I.S.T., 1976.
- 16 - International classification and indexing bibliography : ICIB 1 : Classification systems and thesauri, 1950-1982. Frankfurt : Indeks Verlag, 1982.
- 17 - LAUREILHE (Marie-Thérèse). Bibliographie des thesauri et index par matières parus depuis 1960 ... Paris : Bibliothèque nationale, 1975. Extr. du Bull. Bibl. France, mai 1969, janv. 1970, janv. 1971, fév. 1972, mars 1973, mai 1974.
+ compléments :
 - Bull. Bibl. France, 20° année, n°3, mars 1975, p. 119-127
 - Bull. Bibl. France, 21° année, n°3, mars 1976, p. 107-113
 - Bull. Bibl. France, 22° année, n°3, mars 1977, p. 161-170
 - Bull. Bibl. France, 23° année, n°7, juil. 1978, p. 407-418
 - Bull. Bibl. France, 25° année, n°3, mars 1980, p. 109-114
 - Bull. Bibl. France, 27° année, n°3, mars 1982, p. 151-155

THESAURUS

- 18 - ASSOCIATION OF SPECIAL LIBRARIES AND INFORMATION BUREAUX. Classification Research Group. A Classification of library science. London ; Aberystwyth : ASLIB, 1965 .
- 19 - BANQUE D'INFORMATION POLITIQUE ET D'ACTUALITE DE LA DOCUMENTATION FRANCAISE. Thésaurus : actualité politique, économique et sociale. Paris : La Documentation française, 1983.
- 20 - BUREAU MARCEL VAN DIJK et SANDEAU (Georges). Thésaurus du management et de l'économie ... 2° éd. Paris ; Bruxelles : Bureau Marcel van Dijk, 1975. 2 vol.
- 21 - CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE. Centre de documentation scientifique et technique. Informascience. Thésaurus sciences de l'information. Paris : C.N.R.S., 1977.
- 22 - DAHLBERG (I.) et DEUTSCHE GESELLSCHAFT FUR DOKUMENTATION (et) KOMITEE THESAURUS FORSCHUNG. Entwurf für ein Deskriptorsystem der Informationswissenschaften als Basis eines Bereichsthesaurus. Frankfurt, 1971.
- 23 - DANIEL (R.) et MILLS (J.) et LIBRARY ASSOCIATION. A classification of library and information science, LONDON, 1975.
- 24 - FEDERATION INTERNATIONALE DE DOCUMENTATION. Subject descriptors to be used in ISORID. 's-Gravenhage, 1972.
- 25 - GILCHRIST (A.) et al. A compressed term index language for library and information science. London : ASLIB, 1973.
- 26 - GREAVES (M.A.) et UNIVERSITY OF IBADAN. Institut of librarianship. Librarianship : a faceted classification scheme as a teaching aid. Ibadan, 1970.
- 27 - INSTITUT FUR BIBLIOTHEKSWISSENSCHAFT UND WISSENSCHAFTLICHE INFORMATION DER HUMBOLDT-UNIVERSITÄT BERLIN. Abteilung Information und Dokumentation. Systematik der Information und Dokumentation : Entwurf. Berlin, 1967.
- 28 - NORTH LONDON POLYTECHNIC. A Classification and thesaurus of library and information science. London : North London Polytechnic, 1973.
- 29 - SCHULTZ (Claire K.). Thesaurus of information science terminology. Metuchen (N.J.) ; London : The Scarecrow Press, 1978.
- 30 - UNESCO. Thesaurus : a structured list of descriptors for indexing and retrieving literature in the fields of education, science, social science, culture and communication/compiled by Jean Aitchison. Paris : Unesco, 1977. 2 vol.
- 31 - UNESCO. Thésaurus de l'Unesco : liste structurée de descripteurs pour l'indexation et la recherche bibliographiques dans les domaines de l'éducation, de la science, des sciences sociales, de la culture et de la communication/ établi par Jean Aitchison ; trad. de Jean Viet. Paris : Unesco, 1983.

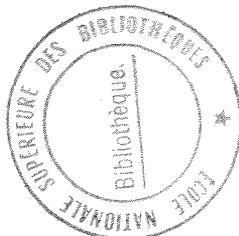
- 32 - VIET (Jean). Thésaurus : mass communication : communication de masse : comunicacion colectiva. Paris : Unesco, 1974.

LISTES DE VEDETTES-MATIERES

- 33 - BIBLIOTHEQUE NATIONALE (France). Liste des vedettes-matières. Paris : Bibliothèque nationale, 1982. Microfiches.
- 34 - BIBLIOTHEQUE PUBLIQUE D'INFORMATION. Autorité matière : noms communs. 5° éd. Paris : Centre National d'art et de culture Georges Pompidou, 1981. 4 vol.
- 35 - BIBLIOTHEQUE PUBLIQUE D'INFORMATION. Autorité matière : noms propres. 4° éd. Paris : Centre national d'art et de culture Georges Pompidou, 1981. 2 vol.

TERMINOLOGIE

- 36 - WERSIG (G.) et NEVELING (U.). Terminologie de la documentation = Terminology of documentation : a selection of 1200 basic terms published in English, French, German, Russian and Spanish. Paris : Unesco, 1976.



On peut également interroger PASCAL (C.N.R.S./C.D.S.I.). Du fait de la coopération qui existe entre le C.D.S.I. et l'I.N.T.D., les notices produites par ce dernier (signalement, indexation et résumé des articles de périodiques et des mémoires d'élèves) sont en effet utilisées par le C.D.S.I. pour la constitution des bases PASCAL dans le domaine des sciences de l'information (On les retrouve donc également dans le bulletin bibliographique T205 Sciences de l'information. Documentation, anciennement section 101 du Bulletin signalétique du C.N.R.S.). Dans une recherche par sujets, pour l'indexation des questions, il faut évidemment utiliser le lexique du C.D.S.I., qui diffère très légèrement de celui qu'emploie l'I.N.T.D.. Une fois qu'ont été obtenues, par la recherche directe sur lexiques, les références répondant à la question posée, on a deux possibilités pour sélectionner les documents analysés par l'I.N.T.D. parmi l'ensemble des documents signalés par le C.D.S.I. : on peut, soit faire défiler les notices et retenir soi-même celles qui comportent la mention I.N.T.D., donnée entre parenthèses à la suite du résumé ; soit -mais seulement si l'on interroge PASCAL sur TELESYSTEMES-QUESTEL- faire une recherche sur texte au moyen du terme I.N.T.D.. Il faut toutefois tenir compte du fait que la répartition de l'analyse des documents entre les deux organismes n'est pas définie de manière rigide : il peut arriver que des documents ne portent pas la mention (INTD) bien qu'ils se trouvent à la bibliothèque de l'I.N.T.D., par exemple parce que celle-ci n'a pu en communiquer les notices à temps au C.D.S.I..

On peut signaler ici ce qu'il en est de la coopération de l'E.N.S.B. et du Centre de documentation de la D.B.M.I.S.T. à la constitution des bases de données PASCAL dans le domaine des sciences de l'information. La participation de l'E.N.S.B. se limite pour le moment au traitement des mémoires de fin d'études des élèves. Quant au Centre de documentation de la D.B.M.I.S.T., il ne participe pas actuellement au système. Toutefois le Centre de documentation de la DICA, devenue AUDIST puis DIST, a coopéré régulièrement jusqu'en 1981 environ. Il était chargé de l'analyse des monographies françaises et étrangères, principalement dans le domaine de l'informatique documentaire et de l'automatisation des bibliothèques.



Centre Régional des COOPératives Etudiantes

CERCOOPE

8 Rue Volney 69 008 LYON - Tél. 874-89-08