

1989
ID
15

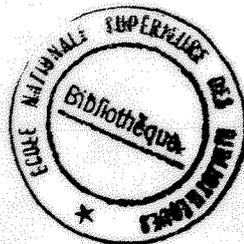
Université Claude Bernard - Lyon 1
Villeurbanne
1989

Construction des langages
d'indexation :
aspects théoriques

note de synthèse

Présentée par Alexis RIVIER

Sous la direction du Professeur Richard BOUCHE



DESS d'informatique documentaire

1355
33
15

Université Claude Bernard - Lyon 1
Villeurbanne
1989

Construction des langages
d'indexation :
aspects théoriques

note de synthèse

Présentée par Alexis RIVIER

Sous la direction du Professeur Richard BOUCHE



DESS d'informatique documentaire

1989

FD
15

SOMMAIRE

1. Méthode de recherche bibliographique	3
<i>1.1. Délimitation du sujet</i>	3
1.1.1. Terminologie	3
1.1.2. Autres critères de sélection	5
<i>1.2. Instruments de la recherche bibliographique</i>	5
1.2.1. Recherche manuelle	5
1.2.1.1. Bibliographies courantes	5
1.2.1.2. Bibliographies rétrospectives	6
1.2.1.3. Bibliographie thématique	6
1.2.1.4. Synthèses bibliographiques	7
1.2.1.5. Résultats	7
1.2.2. Recherche automatisée	7
1.2.2.1. Equations de recherche	8
1.2.2.2. Analyse des résultats	9
<i>1.3. Présentation de la bibliographie - Localisation des documents primaires</i>	11
2. Synthèse	12
<i>2.1. Le renouvellement des classifications</i>	12
2.1.1. Le Classification research group (C.R.G.)	12
2.1.2. Classifications a priori ou a posteriori	13
2.1.3. La revue <i>International classification</i>	14
2.1.4. Recherches en Europe de l'Est	15
<i>2.2. Evolution des langages d'indexation en chaîne</i>	16
2.2.1. Ordre des termes	17
2.2.2. Permutation des termes	18
<i>2.3. Développement des thésaurus</i>	19
2.3.1. Le thésaurus en Amérique du Nord	19
2.3.2. Etudes critiques des théaurus en Europe	20

2.4. Langages d'indexation et théories linguistiques	23
2.4.1. Langages d'indexation et langues naturelles	23
2.4.2. Les deux axes de référence	24
2.4.3. Le triangle sémantique	25
2.4.4. les universaux du langage	26
2.4.5. La terminologie	26
2.4.6. Le problème du sujet	27
2.5. Langages d'indexation et théories mathématiques	28
2.5.1. Modèles généraux	28
2.5.2. Modèles spécifiques	29
2.6. Conclusions	31
2.6. Aperçu chronologique	32
2.7. Liste des abréviations	34
3. Bibliographie	35
3.1. Théories et modèles	35
3.1.1. Généralités	35
3.1.2. Théories et modèles linguistiques	37
3.1.3. Théories et modèles mathématiques	39
3.2. Relations entre les termes d'indexation	41
3.2.1. Généralités	41
3.2.2. Relations sémantiques	42
3.2.3. Relations syntaxiques	43
3.3. Langages et systèmes d'indexation	43
3.3.1. Indexation en chaîne	43
3.3.1.1. Généralités	43
3.3.1.2. PRECIS et POPSI	44
3.3.1.3. Autres systèmes	45
3.3.2. Théorie des graphes : une application	46

1. METHODE DE RECHERCHE BIBLIOGRAPHIQUE

1.1. DELIMITATION DU SUJET

Des spécialistes se sont plaints du manque de recherches théoriques sur les langages d'indexation (1). On pense habituellement que leur efficacité en est la meilleure justification. La littérature est très abondante en manuels pratiques sur la construction des LI ainsi qu'en évaluation de leurs performances, alors que les textes en exposant les principes théoriques sont moins nombreux et moins connus. C'est déjà une raison suffisante pour chercher à les rassembler. De plus des exemples ont montré qu'une meilleure connaissance interne des LI peut également conduire à améliorer leur conception.

1.1.1. Terminologie

Notre recherche porte sur les aspects théoriques des langages d'indexation. Cet intitulé contient deux mots appartenant à des réalités différentes. Le terme de "langage d'indexation" a un sens très précis dans le domaine des sciences de l'information (2). "Théorie", au contraire, relève du langage courant et est commun à toutes les disciplines scientifiques, d'où sa plurivocité.

Les "langages d'indexation" (3) sont des langages artificiels - c.-à-d. construits à l'aide d'un ensemble de règles données - servant à la représentation abrégée du contenu d'un document. Il y a des degrés : les LI utilisés sont plus ou moins codifiés. En nous inspirant de Maniez [17], nous représenterons les LI selon eux axes : celui de la coordination et celui du contrôle :

1. Par exemple J. Maniez [65 : 133].

2. "Science qui étudie la communication de l'information" selon Yves Le Coadic, Bulletin des bibliothèques de France. v. 29, n. 2, p. 168-171. Cet article situe la science de l'information comme une "interdiscipline" entre les sciences humaines et les sciences pures.

3. Abrégé LI. Nous choisissons ce terme parmi beaucoup de synonymes ("langages documentaires", etc.). Il recouvre également les nombreux équivalents anglais : "indexing languages", "documentary languages", "information retrieval languages", etc.

	Précoordination	Postcoordination
+	classifications hiérarchiques	classifications à facettes
Contrôle	langage en chaîne (PRECIS, POPSI...)	thésaurus
		terminologies
-	indexation matière libre	descripteurs libres
		mots-clés (unitermes)

A vrai dire les LI peu contrôlés (comme les descripteurs choisis librement pour représenter le contenu d'un document) se rapprochent des langages naturels. Nous nous intéressons au contraire aux LI les plus contrôlés, car c'est précisément un trait qui les différencie nettement de la langue courante. Il s'agit des LI *post-coordonnés* (la combinaison des descripteurs se fait au moment de la recherche documentaire) comme les *thésaurus* ; ainsi que les LI *pré-coordonnés* (la combinaison des termes est fixée au moment de l'indexation) comme les *classifications*, et surtout les *langages en chaîne* (PRECIS, SYNTOL, POPSI, etc.), doté d'une syntaxe contraignante.

Le terme *théorie* a, positivement, le sens de construction intellectuelle ou spéculative, ce qui est vague. Négativement, il s'oppose bien sûr à *pratique*. Nous emploierons également la notion plus technique de *modèle* qui représente un type de démarche théorique. Le *modèle* est un outil intellectuel servant de principe explicatif d'un domaine de la connaissance. Ainsi des modèles linguistiques ou mathématiques ont été employés pour représenter la structure des LI.

Nous écartons donc les nombreux manuels de construction des LI (surtout des *thésaurus*), les évaluations statistiques de leur efficacité, de leurs avantages ou inconvénients ; leur portée étant essentiellement pratique.

1.1.2. Autres critères de sélection

Notre travail ne vise pas l'exhaustivité, mais se veut une *synthèse*. La littérature est parfois redondante. On veillera à sélectionner les textes pour leur valeur novatrice, leur importance pour la recherche théorique, ou pour leur qualité de synthèse précisément.

Notre recherche portant sur des aspects théoriques, une sélection selon la *date* des documents est secondaire. Au contraire des pratiques et des techniques (nous pensons bien sûr au bouleversement apporté par l'informatique dans la documentation), les théories des LI peuvent garder leur intérêt bien des années plus tard. La bibliographie de notre sujet aura donc une dimension *rétrospective*.

Nous limitons notre recherche aux langues suivantes : allemand, anglais, français. Elles regroupent l'essentiel de la littérature occidentale. L'importante production des pays de l'Est nous est donc connue dans la mesure où elle est traduite (4).

1.2. INSTRUMENTS DE LA RECHERCHE BIBLIOGRAPHIQUE

1.2.1. Recherche manuelle

Nous la présentons selon l'ordre d'élaboration intellectuelle des types de sources.

1.2.1.1. Bibliographies courantes

- **Bulletin signalétique du C.N.R.S. 101.** Science de l'information, documentation. 1970-1983. Mensuel.

Devient :

- **PASCAL Thema. T 205.** Sciences de l'information, documentation. 1983 →. Mensuel.

Cette bibliographie compilée par le C.D.S.T.-C.N.R.S. couvre bien la littérature européenne. Les mises-à-jour sont malheureusement assez longues. Plan de classement hiérarchisé. Rubrique d'accès : 04 B "langages documentaires".

4. Certains titres anglo-saxons sont même des traductions de revues soviétiques, comme *Automatic documentation and mathematical linguistics* ou *International forum on information and documentation*.

- **LISA : Library information science abstracts.** 1969 →. Mensuel.

Produite par la "Library Association" et "Aslib". Malgré son origine britannique, LISA s'efforce d'indexer les revues non anglo-saxonnes. Plan de classement à facettes élaborées par le Classification Research Group (C.R.G.). Les différentes facettes ou aspects exprimant un sujet sont toujours citées dans le même ordre. Entrée par la facette principale V "Subject indexing" ou X "classification".

- **ISA : Information science abstracts.** 1966 →. Mensuel.

Éditée par plusieurs associations américaines, dont l'"American society for information science", ISA est, sauf très rares exceptions, principalement anglo-saxonne. Pas une seule revue française n'était analysée en 1988 ! Plan de classement hiérarchisé. Rubrique d'accès : 5.7. "Indexing, thesauri, cataloging, classification".

Les références sont accompagnées d'un résumé, facilitant la sélection. Ces trois bibliographies ont été surtout consultées sur leur forme de base de données.

1.2.1.2. *Bibliographies rétrospectives*

- **Library literature.** 1921 →. Vol. cumulatifs pluriennaux, puis annuels.

Éditée par l'American Library Association. Indexe aussi un certain nombre de revues européennes. Le classement est de type dictionnaire, ce qui permet une recherche précise, mais disperse les entrées. Points d'accès : "index terms", "indexing", "classification", "thesauri". Les références ne sont pas accompagnées d'indications de contenu.

1.2.1.3. *Bibliographies thématiques*

Par leur aspect rétrospectif et sélectif selon le sujet abordé, elles nous sont très précieuses. Les références sont classées méthodiquement dans des rubriques détaillées.

- **Dahlberg, Ingetraut (ed.).** *Classification and indexing systems : theory, structure, methodology : 1950-1982.* Frankfurt : Indeks Verlag. 1985. (International classification and indexing bibliography, ICIB 3).

Très complète : 5600 références, généralement non commentées. Mises-à-jour dans la revue *International classification*.

- **Wellisch, Hans H. (ed.).** *Indexing and abstracting : an international bibliography.* Santa Barbara, Cal. ; Oxford : ABC-Clio, 1980.

Sélection de 2400 références résumées couvrant la littérature jusqu'en 1976.
Mises-à-jour dans la revue The indexer.

- Maccafferty, Maxine. *Thesauri and thesauri construction*. London : Aslib, 1977. (Aslib bibliography, 7).

800 références commentées. Pas de mises-à-jour.

1.2.1.4. Synthèses bibliographiques

- **Annual review of information science and technology**. Washington : American society for information science, 1966 →. Annuel.

Chaque volume dresse pour l'année écoulée le tableau de la recherche en science de l'information. Sous des noms divers, un des chapitres est régulièrement consacré à l'analyse du sujet ("subject analysis"). Une brève présentation introduit une liste sélective des travaux les plus importants (une centaine), presque toujours en anglais.

- Des articles de synthèse ("surveys") font le point sur des sujets ponctuels (comme POPSI ou PRECIS). Ils sont souvent accompagnés d'une liste très complète de références. On trouvera les plus récents dans notre bibliographie.

1.2.1.5. Résultats

Les synthèses bibliographiques et les bibliographies thématiques, ainsi que les références citées dans les articles, nous ont fourni l'essentiel des documents nécessaires.

1.2.2. Recherche automatisée

Quatre bases de données ont été interrogées le 20-3-1989 : PASCAL, INSPEC (Questel+), LISA et ISA (Dialog). INSPEC et PASCAL sont des bases pluridisciplinaires, mais il est possible de limiter sa recherche aux sous-bases concernant les sciences de l'information.

Bases de données	Nb. références	Couverture (5)
PASCAL (PASCAL science de l'information)	6'500'000 + 430'000/an (42'000)	1973 → (1973 →)
INSPEC (INSPEC info science)	3'000'000 + 220'000/an	1969 → (1971-1979 →)
LISA	85'000 + 6'000/an	1969 →
ISA	90'000 + 9'000/an	1966 →

1.2.2.1. Equations de recherche

Le principe général : croiser "langages d'indexation" et "théorie", puis trier sur les langues retenues. Comme il n'est pas possible d'"éclater" la recherche à partir d'un descripteur générique, il faut enrichir l'équation avec des termes spécifiques : "thésaurus", "classification",...

- PASCAL :

1. .LIM SCIENCE INFORMATION/FG
- 2 (LANGAGE DOCUMENTAIRE OU INDEXATION OU THESAUR+/t
OU CLASSIFICATION) ET THEORI+/DE/T
- 3 2 ET (FRE/LA OR GER/LA OR ENG/LA)

- INSPEC :

- 1..LIM CT " ^{C72+T} +/CC/T
- 2 (INFORMATION ANALYSIS OU INDEXING/DE OU
VOCABULARY/DE OU THESAUR+/DE/T) ET (THEORY OU
THEORIES OU MODEL?/T)
- 3 2 ET (ENG/LA OU GER/LA OU FRE/LA)
- 4 3 ET /DP >1980

- LISA et ISA :

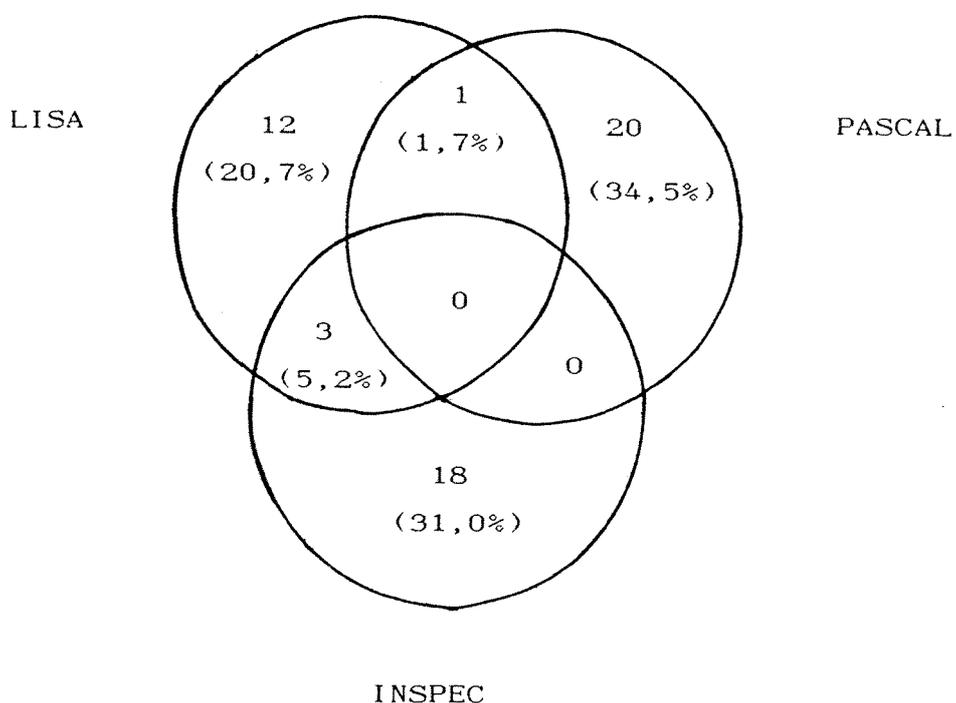
S1 INDEXING (W) LANGUAGE OR THESAUR?
 S2 S1 AND (THEORY OR THEORIES OR MODEL?)
 S3 S2/ENG OR (S2 AND LA=FRENCH) OR (S2 AND LA=GERMAN)

1.2.2.2. Analyse des résultats

Les 270 références de ISA n'ont pas été visualisées, et nous avons limité la recherche dans INSPEC à partir de 1980.

	PASCAL	LISA	INSPEC
Réponses			
-total	102	66	73
-pertinentes	38 (37 %)	24 (36 %)	21 (29 %)
Langue			
-anglais	79 (77,5 %)	52 (78,8 %)	69 (94,5 %)
-allemand	14 (13,7 %)	10 (15,1 %)	4 (5,5 %)
-français	9 (8.8 %)	4 (6,1 %)	0

- Recouvrement et apports spécifiques (réponses pertinentes >1980)



Le taux de pertinence (même s'il n'est pas toujours aisé à évaluer lorsqu'on n'a pas les documents primaires entre les mains) est peu élevé. La plurivocité du terme "théorie" en est certainement la cause. Même lorsqu'il est admis dans le champ descripteur (comme pour PASCAL), les résultats sont à peine supérieurs. Il était difficile pourtant de se passer de ce terme ; "modèle" ne relevant que d'un aspect de la question.

Le taux de recouvrement est très faible : aucune notice n'est commune aux trois bases. PASCAL et INSPEC sont complètement disjointes. Pourtant elles dépouillent toutes les mêmes grandes revues. Par contre, leur politique d'indexation est différente. PASCAL analyse complètement les périodiques retenus. INSPEC au contraire, dépouille les articles en fonction de son orientation technique et informatique. Elle nous donne beaucoup de documents d'automatisation documentaire, aspect hors de notre sujet.

De plus les références récentes sont rares dans LISA et PASCAL. Les années 1976-1981 concentrent plus de 50 % des références dans LISA. Dans PASCAL, la recherche n'a donné aucune référence de 1988 et une seule pour 1987 !

La recherche documentaire souffre d'une double imprécision, cause de bruit. Comme Vickery l'a souligné (inform. syst : 218), le processus d'indexation ou de condensation appauvrit le contenu du texte original. De son côté, le chercheur est contraint, à l'interrogation, de traduire sa question dans une formule logique simplifiée agréée par le logiciel d'interrogation.

Pour notre sujet, on aurait peut-être obtenu des résultats meilleurs mais plus limités par des questions très précises comme ("TOSAR" et "théorie des graphes") ou ("PRECIS" et "indexation" et "théorie"). Vu les bons résultats de la recherche manuelle, nous n'avons pas jugé nécessaire d'entreprendre de nouvelles interrogations.

2.3. PRESENTATION DE LA BIBLIOGRAPHIE - LOCALISATION DES DOCUMENTS PRIMAIRES

Les quelques cent références retenues sont classées méthodiquement et sous-classées alphabétiquement par noms d'auteurs. Nous retenons trois grandes divisions :

1. Les principales théories et modèles, tels que ces termes ont été définis plus haut.
2. L'expression des relations entre les termes d'indexation. Comme tout langage, les LI possèdent des liens sémantiques et syntaxiques.
3. Quelques systèmes d'indexation originaux considérés à partir des réflexions théoriques qui ont été à leur base.

La plupart des périodiques et plusieurs monographies de notre bibliographie se trouvent à la bibliothèque de l'E.N.S.B. On remarque que plus de la moitié des articles recensés sont concentrés dans 4 revues : *Journal of the american society for information science*, *International classification*, *Information processing and management* (suite de : *Information storage and retrieval*) et *Journal of documentation*.

Les autres périodiques non reçus par l'E.N.S.B. doivent être localisés par le C.C.N. (*Catalogue collectif national des publications en série*) accessible sur microfiches, CD-ROM ou en ligne. Les monographies seront localisées par le C.C.O.E. (*Catalogue collectif des ouvrages étrangers*, 4 rue de Louvois, 75002 Paris) consultable sur place ou sur demande.

2. SYNTHESE

2.1. LE RENOUVELLEMENT DES CLASSIFICATIONS

2.1.1. Le *Classification research group* (C.R.G.)

Après 1945 surtout, les insuffisances des classifications traditionnelles deviennent de plus en plus évidentes. Même s'il n'est pas encore question d'ordinateurs, une gestion plus rationnelle de fonds documentaires croissants s'impose dans les pays industrialisés. En Amérique du Nord, les recherches se tournent résolument vers d'autres méthodes, comme l'*indexation coordonnée* [23] qui conduira aux thésaurus. En Europe, on essaie plutôt de fonder les classifications sur de nouvelles bases.

En 1948 est fondé à Londres le *Classification research group* (C.R.G.). Un de ses membres, D.J. Foskett [5] en dresse le panorama des thèmes abordés jusqu'en 1970. La classification n'est plus considérée comme le seul rangement physique de livres sur des rayons, mais comme un authentique LI dont le but est la recherche documentaire.

Les théories de S.R. Ranganathan en Inde ont eu une influence considérable sur le C.R.G. et de manière générale sur la plupart des chercheurs en sciences de l'information jusqu'à nos jours. Sa méthode de classification *analytico-synthétique* [18 : 109] consiste à analyser le sujet en différents points de vue ou *facettes*, puis à le resynthétiser par l'arrangement des facettes obtenues selon une séquence canonique. Par rapport aux classifications énumératives rigides, la possibilité de coordonner les facettes présente une plus grande souplesse d'indexation, et les relations polyhiérarchiques qu'elles engendrent entre elles sont plus proches de la multiplicité des traits composant un sujet. Le C.R.G. a produit plusieurs classifications sectorielles à facettes, dont la "classification of library and information science de 1971, utilisée pour le service bibliographique LISA.

Le C.R.G. a aussi pour but de dresser l'esquisse d'une nouvelle classification universelle sur une base naturaliste reflétant l'ordre du monde réel. Il adopte la thèse biologique des *niveaux d'intégration* de J.K. Feiblemann.

L'univers s'est développé en niveaux successifs de complexité croissante, chaque niveau intégrant le précédent (particules fondamentales, atomes, molécules, etc.) Tous les domaines de la connaissance peuvent être ordonnés selon ce principe.

Il s'oppose ainsi à la tendance pragmatique américaine qui répond au principe de la "caution des textes" (*literary warrant*) : une rubrique n'a de raison d'être que si elle est potentiellement utile au classement d'au moins un document réel [F.W. Lancaster, 14 : 86].

La théorie des niveaux d'intégration a inspiré la notion de *contextualisation* appliquée par D. Austin à PRECIS [89, 90]. D.J. Foskett [6] généralisera cette thèse par la *théorie générale des systèmes*. Un système est un tout qui est plus que la somme des parties qu'il contient, mais qui n'est lui-même qu'une partie d'un système plus important, et ainsi de suite.

2.1.2. Classifications *a priori* ou *a posteriori*.

La construction des classifications doit-elle se faire *a priori* en imposant un ordre théorique ou *a posteriori* en partant des relations concrètes ?

Dès 1950, J.E.L. Farradane [100, 101] soutient qu'une classification doit être élaborée sur des considérations épistémologiques sur le processus même d'acquisition des connaissances. Une démarche scientifique est une alternance de phases *déductives* et *inductives*. Les classifications traditionnelles ne sont pas scientifiques car purement *déductives*. Il décrit une base inductive : des *isolats* (concepts) sont reliés entre eux par des termes relationnels (opérateurs) pour former un *analet*, représentant un sujet. De proche en proche, en combinant les *analets* on obtient une structure ordonnée, coextensive à l'ensemble du savoir. Pourtant Farradane s'est intéressé aux propriétés des *analets* isolés. Par des règles d'arrangement et de permutation, ils permettent de servir à l'établissement d'index alphabétiques. Il parlera de sa méthode comme une *indexation relationnelle* [102] apparentée aux *LI en chaîne*.

Dans les années soixante les calculateurs font leur apparition. R.M. Needham et K. Sparck Jones [66] essaient de construire automatiquement une classification en partant des documents. Ils élargissent la notion de classe en celle d'*amas* (*clump*) : si deux mots-clés figurent dans l'indexation d'un document, alors ils sont équivalents du point de vue de la recherche de ce document. En mesurant le *degré de cooccurrence* de toutes les paires possibles de mots-clés on pourra déterminer des ensembles (*amas*) dont chaque

mot-clé a un degré de cooccurrence plus élevés avec les autres membres qu'avec les non-membres.

A partir des années quatre-vingts, un examen critique des travaux du C.R.G. conduit certains auteurs à une position *aprioriste*. D. Austin [60] postule l'existence d'une part de classes de *concepts de base* (choses, actions, propriétés) et, d'autre part, de classes de *relations de base* (logiques, grammaticales).

B.P. Frohmann [70] critique le choix du C.R.G. de structurer *a posteriori* les relations sémantiques et de rejeter la formule rangianthienne prédéfinie PMEST. Chaque domaine donne lieu à ses propres relations et il n'est alors pas possible d'atteindre un système général de catégories. Une *base sémantique* doit être dégagée *a priori* une fois pour toutes selon les indications de D. Austin.

2.1.3. La revue *International classification*

En Allemagne, une impulsion a été déclenchée dans l'étude des classifications par I. Dahlberg, fondatrice en 1974 de la revue *International classification*, editrice de la collection *Studien zur Klassifikation* publiée par la Gesellschaft für Dokumentation.

I. Dahlberg [4] a mené une étude très vaste (près de 600 référence bibliographiques !) à la recherche des "fondements universels de l'ordre du savoir" (*Grundlagen universaler Wissensordnung*). L'ordre du savoir est vital à une époque où les disciplines se fragmentent toujours plus. La notion de classification déborde le domaine de la bibliothéconomie. Une recherche interdisciplinaire est conduite en passant par la philosophie, l'épistémologie, la linguistique, les théories scientifiques, etc. Le résultat se présente comme une esquisse formelle d'une classification universelle.

T.T. Ballmer [26, 27] suit le même but mais en restreignant son attention sur les expressions linguistiques. Le verbe joue un rôle important dans la classification du savoir. Par exemple "manger" détermine une classe d'objets "mangeables" : fruits, légumes, viandes...

2.1.4. Recherches en Europe de l'Est

Y.A. Shreider [53, 54], pose une question épistémologique fondamentale : l'ordre est-il une forme imprimée par l'homme au chaos ou est-il une propriété des choses même ? La logique et la classification aristotéliennes offrent une base de réflexion. Une notion peut être décrite en *extension* ou en *intension*. L'*extension* est la classe des objets déterminés par la notion (appelée *taxon* par l'auteur). L'*intension* (ou *compréhension*) est l'ensemble des caractéristiques de la notion (appelées *mérons*). La classification repose sur une dualité d'approches : *taxonomie*, où les objets sont représentés sur la base de leurs similarités, et *méronomie*, soit l'identification des caractéristiques essentielles communes d'un taxon.

2.2. EVOLUTION DES LANGAGES D'INDEXATION EN CHAÎNE

L'indexation en chaîne consiste à analyser un sujet en une suite linéaire et rigide de termes dans le but de construire des index alphabétiques manuels.

J.O. Kaiser, S.R. Ranganathan et E.J. Coates sont à différents titres les précurseurs d'une technique qui s'est développée surtout à la fin des années soixante. En 1960 le système KWIC de H.P. Luhn est le premier index entièrement mécanisé

Les anglo-saxons [B.W. Mineur, 86] distinguent volontiers en les opposant *chain indexing* basée sur la *procédure en chaîne* de S.R. Ranganathan à *string indexing* qui désigne les systèmes modernes.

L'ouvrage de T.C. Craven [84] est une synthèse actuelle décrivant un grand nombre de systèmes.

En 1968, la *British national library* (B.N.B.) décide de remplacer la production d'index au moyen de la procédure en chaîne par une nouvelle technique, en liaison avec le projet UK-MARC. Dès 1971, un prototype de PRECIS est testé.

A la même époque l'"école indienne" avec G. Bhattacharrya lance POPSI. Dans les années soixante-dix, d'autres langages apparurent : NEPHIS et LIPHIS de T.C. Craven.

Récemment, des comparaisons de ces LI ont été menées [A. Cheti, 94]. J.E.L. Farradane [85] compare PRECIS, POPSI et NEPHIS avec son *indexation relationnelle*. Des "surveys" font le point des études consacrées à PRECIS [M. Mahapatra ; S.C. Biswas, 96] ou à POPSI [S.C. Biswas, 93].

Ces LI sont de type *précoordonné*. Comme ils servent de moyen de recherche manuel ("index" au sens traditionnel), chaque terme doit apparaître en tête de chaîne pour être accessible. Deux problèmes doivent être résolus : celui de l'ordre de succession des termes (une "grammaire") et celui des règles de *permutation* des termes.

2.2.1. Ordre des termes

L'*indexation systématique* de J.O. Kaiser [103 ; E. Svenonius, 104] consiste à analyser tous les sujets selon deux grandes catégories : les *concrets* (ou objets) et les *processus* (ou actions). Syntactiquement, les concrets précèdent toujours les processus.

S.R. Ranganathan formule des règles syntaxiques très précises pour disposer les facettes d'un sujet en une séquence fixe, comme par exemple le fameux "wall-picture principle" [18 : 425].

Plus tard, E.J. Coates [61], reprenant Kaiser, cherche à fonder psychologiquement la préséance de la notion la plus "claire" afin de mieux suivre l'attente de l'utilisateur : les choses (stables et déterminées) précèdent les *matériaux* (stables mais non déterminés) et les *actions* (dynamiques, instables).

G. Bhattacharyya reprend les principes de S.R. Ranganathan [91, 92] qu'il appelle *postulats* (d'où le nom de "POPSI"). La séquence fondamentale est celle de la "base" (catégorie élémentaire ou discipline) et du "noyau" (*core*) des concepts se rapportant à cette base.

Dans PRECIS, l'ordre est déterminé par le principe de *dépendance contextuelle* (*context dependency*) [90]. Chaque terme suit le terme qui représente son concept et qu'il précise ("Asie - Inde" par ex.). La chaîne syntagmatique est renforcée par l'attribution d'opérateurs de rôle à chaque terme. Le système est conçu astucieusement de telle sorte que l'ordre de citation des opérateurs associés aux termes conduit automatiquement à une chaîne de dépendances contextuelles.

B.W. Mineur [86] cependant critique PRECIS parce qu'il privilégie un ordre de base unique en occultant toutes les relations secondaires entre les termes non contigus dans la chaîne. En ce sens l'*indexation relationnelle* de J.E.L. Farradane, bien qu'apparentée, n'est pas un LI en chaîne puisque les *analets* peuvent avoir une structure bi-dimensionnelle.

Face à des systèmes formalisés comme POPSI et PRECIS, NEPHIS (*Nested phrase indexing system*) [98] apparaît très empirique. Le sujet est analysé en expressions simples et complexes "emboîtées" (*nested*) à plusieurs niveaux les unes dans les autres. Le système structure donc ces éléments dans une arborescence. Pour certains sujets une telle simplification n'est pas judicieuse. LIPHIS (*Linked phrase indexing system*) améliore ce système en donnant justement la possibilité de "lier" (*link*) latéralement les éléments d'un sujet dont la structure n'est pas en arbre [99].

2.2.2. Permutation des termes

A partir d'une chaîne source il faut déterminer les chaînes dérivées qui mettront en position d'accès chacun des termes qui la composent. Plusieurs solutions formelles sont possibles [W.J. Hutchins, 34 : ch. 6].

La *procédure en chaîne* de S.R. Ranganathan utilise la *truncation*. A partir d'une chaîne ABCD, on dérivera BCD, CD, et D. Chaque terme est bien en position d'accès, mais il n'est suivi que d'une partie des termes représentant le sujet. Le défaut du système est qu'un sujet spécifique (ABCD) figurera sous des entrées trop généraux (CD, D). Ce problème est celui de la "chaîne disparaissante" (*disappearing chain*) [S.C. Biswas, 93] et une des raisons de l'abandon de cette technique utilisée de 1950 jusque vers 1970 par la BNB.

De même J.E. Armitage et M.F. Lynch [81] formalise la production d'"index matière articulés" par permutation d'éléments syntaxiquement liés par des prépositions en LN, sans introduire d'ambiguïtés.

La *permutation* des termes permet de respecter le principe de spécificité. Mais il multiplie les entrées et disperse l'ordre et les relations de la chaîne source.

Les systèmes mécanisés de type KWIC sont des *rotations* de mots-clés dans une phrase. PRECIS utilise la rotation et dispose la chaîne dérivée sur deux lignes, de façon que la chaîne source et ses relations puissent toujours être virtuellement reconstruits.

2.3. DEVELOPPEMENT DES THESAURUS

Le terme de *thesaurus* choisi pour désigner un type de LI contrôlé fait référence au *Thesaurus of english word and phrases* (1852) de P.M. Roget [J. Maniez, 65]. A l'inverse des dictionnaires de langue habituels où l'on part d'un *terme* pour découvrir sa définition (donc son *sens*), Roget part du *sens* (les concepts) pour en donner tous les termes qui le représentent, sans définitions. Le but des *thesaurus* modernes est également de représenter d'abord des concepts, mais aussi de préciser leurs relations.

2.3.1. Le *thesaurus* en Amérique du Nord

Aux Etats-Unis, les chercheurs ont rapidement développé de nouvelles techniques documentaires visant à remplacer les méthodes classificatoires traditionnelles.

Dès les années cinquante, M. Taube [23, 58] élabore son *indexation coordonnée* (aujourd'hui nous dirions plutôt "indexation post-coordonnée"). Chaque mot-clé ou *uniterme* d'un sujet est la vedette d'une fiche sur laquelle sont inscrits les numéros de référence de tous les documents pertinents. A la recherche l'utilisateur tire les fiches correspondant à sa demande et détecte les numéros communs. Par ce moyen, le nombre de combinaisons possibles est très grand, l'utilisateur n'a pas besoin de connaître l'ordre des éléments d'un sujet complexe et les renvois deviennent superflus.

L'avènement des ordinateurs a permis de faciliter la manipulation de l'*indexation post-coordonnée*, avec le succès que l'on sait. Pourtant les *unitermes* de M. Taube étaient choisis de manière empirique. Le contrôle du vocabulaire, le traitement des relations paradigmatiques (surtout la synonymie) ont vite été ressentis comme une nécessité.

Selon F.W. Lancaster [14], auteur d'une étude générale sur les LI contrôlés, le premier *thesaurus* moderne a été publié en 1959 par la société Dupont de Nemours. A la même époque, les notions de "précision" et de "rappel" apparaissent en relation avec les performances de la recherche.

Durant les années soixante des améliorations ont constamment été apportées aux *thesaurus*.

Les relations syntaxiques étant absentes des *thesaurus*, il en résulte à la recherche des mauvaises coordinations. L'Engineers Joint Council présente en 1964 le premier

thésaurus avec *indicateurs de rôle* permettant d'indiquer à chaque descripteur une fonction.

J.C. Costello [78] compare ce procédé avec les LN à déclinaisons (latin, russe, allemand) comme moyen de liaison syntaxique. Mais de même que d'autres LN utilisent des moyens différents de représenter la syntaxe (ordre des termes, prépositions, etc.), les indicateurs de rôle ne sont bien sûr pas les seuls contrôles syntaxiques possibles des LI.

J. Aitchison présente en 1969 un thésaurus d'un genre nouveau : le *thesaurofacet* [77]. Ce "nouveau concept" intègre en un seul système les avantages de la présentation systématique des termes par l'analyse d'un domaine en facettes, et ceux des thésaurus alphabétiques. Le thesaurofacet est un outil universel qui peut aussi bien servir de LI pré-coordonné dans la production de chaînes pour un index, ou de système post-coordonné de recherche. Deux traditions d'indexation très différentes (américaine et "indo-européenne") se rejoignent ainsi.

2.3.2. Etudes critique des thésaurus en Europe

Après 1970, les spécialistes prennent conscience d'inconvénients plus fondamentaux du thésaurus. L'accroissement des fonds documentaires, la parcellisation des savoirs gonflent le lexique et posent le problème des mises à jour. L'arrangement alphabétique est incapable de bien représenter les relations utiles entre les termes. Le thésaurus et la nature du processus d'indexation est repensé.

Dans le cadre d'une recherche fondamentale sur les LI, R. Fugmann [7] définit de nouvelles notions en rapport avec la tâche même de l'indexeur. La "précision" avec laquelle le terme le plus approprié sera choisi dépend de l'"ordre" dans lequel les termes sont disponibles. L'alphabétisation disperse les termes de sens proche qui ne sont que virtuellement reliés par les renvois.

Plus le terme approprié est difficile à trouver (d'autant plus que la liste s'allonge), plus sa "prévisibilité" (predictability) est faible. En s'inspirant de la méthode de Ranganathan, un LI doit analyser les termes (afin de limiter le lexique) mais aussi les *re-synthétiser* afin de préserver les liens entre eux.

Sur ces bases théoriques, R. Fugmann a développé un système d'indexation bi-dimensionnel appelé TOSAR (*Topological representation of synthetic and analytical relations of concepts*), utilisé pour la documentation automatisée en chimie [105].

Comme aide à l'indexation, les thésaurus ont proposé des "schémas fléchés (*arrowgraphs*) représentant les descripteurs et leurs relations sémantiques. TOSAR va beaucoup plus loin. La structure et les propriétés de la *théorie des graphes* sont exploitées : un *graphe* est un ensemble de points appelés *sommets* et reliés par des arcs. On peut alors définir des notions d'orientation, de distance entre sommets, etc.

En documentation, un sujet se présente par un tel graphe. Les termes sont les sommets et les relations (analytiques ou synthétiques) sont les arcs. Les paramètres de la demande sont aussi formulés par un graphe dont sommets et arcs sont plus ou moins précisés. A la recherche, un programme compare le graphe de la demande avec les graphes des documents enregistrés en mémoire.

La méthode de TOSAR a été décrite formellement au moyen de la logique des prédicats par H. et I. Nickelsen [106].

Récemment R. Fugmann [10] a synthétisé sa théorie sous la forme de cinq axiomes ou postulats de l'indexation. Une dimension éthique est apportée avec le concept d'"indexation impérative" (*mandatory indexing*) : l'indexeur est contraint à choisir le terme le plus approprié, contrairement à l'indexation simplement contrôlée qui laisse toujours la possibilité de choisir plusieurs termes diversement adéquats.

D. Soergel [74] définit généralement un LI comme un ensemble de descripteurs, de relations et de règles pour la formation d'expressions. Le thésaurus contient en plus un vocabulaire de non-descripteurs qui conduisent (*lead-in vocabulary*) vers les termes reconnus. Les utilisateurs étant différents, l'indexation doit être multi-modèle : des contextes, des approches différents doivent amener au même descripteur. L'analyse des sujets par *facettes* - ou *points de vue* précisément - est un exemple d'indexation multi-modèle.

De nombreuses études considèrent des aspects particuliers des thésaurus.

M. Willets [67] compare l'usage de différents types de relations dans plusieurs thésaurus. En prenant pour modèle les catégories générales du C.R.G. et les neuf opérateurs de J.E.L. Farradane, elle étudie tout particulièrement les relations dites "affinitives". Ce type regroupe habituellement toutes sortes de relations que l'on renonce à distinguer, faute d'analyse suffisante.

R. Jansen [72] montre l'utilité d'un nouveau type de relation appelée "relation d'appartenance" (*Zugehörigkeitsrelation*) à ne pas confondre avec les *relations hiérarchiques*. Elle met en rapport un terme avec tous les descripteurs pré-coordonnés dans lequel il entre en composition.

A. Ghose et A.J. Dawle en Inde [71] proposent une technique pragmatique (dérivée de la méthode des amas) de structuration d'un lexique en sciences humaines au moyen de questions posées à des experts dans une discipline.

Pour C. Kim [36], le processus du *dictionnaire* (des mots vers les concepts exprimés) et celui du *thésaurus* de P.M. Roget (des concepts vers les mots) sont utiles dans l'*élaboration* et la *mise à jour* des thésaurus. En effet, la *signification* d'un mot ne dépend pas seulement de la notion ou de l'objet qu'il représente (*théorie référentielle*), mais aussi des relations qu'il entretient avec les autres mots (*théorie opérationnaliste ou contextuelle*).

2.4. LANGAGES D'INDEXATION ET THEORIES LINGUISTIQUES

Les raisons qui ont poussé les documentalistes à s'intéresser aux théories linguistiques sont nombreuses. Les propriétés des LI ressemblent beaucoup à celles des LN. Certains dérivent plus ou moins profondément des LN. On peut se demander si la linguistique ne serait pas adéquate à les décrire, voire à les formaliser. Comme les LI sont appliqués principalement - mais non exclusivement - à des textes exprimés en LN, le passage de l'un à l'autre se pose.

Dans les années soixante, certains chercheurs fondèrent beaucoup d'espoir dans l'automatisation des traitements documentaires. Pour cela ils durent analyser sérieusement les entités qu'ils manipulaient : mots, phrases, résumés, descripteurs, etc. Cette période est marquée par la multiplication des études linguistiques des LI.

K. Sparck Jones et M. Kay [41] font le point des travaux menés entre 1965 et 1973 orientés particulièrement vers les procédures d'automatisation.

En France la thèse de M. Coyaud [30] est une description et comparaison de nombreux LI particuliers de l'époque. L'effort théorique se situe dans l'élaboration d'une grille d'analyse uniforme dont les constituants sont empruntés à la terminologie linguistique : les "monèmes" (plus petites unités décomposables) sont les "lexèmes" (termes) et "morphèmes" (relations). La syntaxe comporte le "mot" (un "lexème" et un "morphème"), le "syntagme" (plusieurs "mots" en relation), etc. L'intérêt de cette méthode est de permettre le rapprochement structurel de LI apparemment différents.

L'ouvrage de W.J. Hutchins [34] est une bonne introduction aux structures linguistiques générales des LI en les comparant sous tous les aspects (formels, sémantiques, pragmatiques, ...) avec les LN.

2.4.1. Langages d'indexation et langues naturelles

J. Maniez soulève dans la première partie de sa thèse [80] une question générale : quelles sont les différences et les ressemblances nécessaires entre LN et LI ? Pour lui, l'élément commun est l'universalité de la *fonction référentielle*, c'est-à-dire la nature symbolique du signe. L'utilisateur ne cherche pas des termes d'indexation pour eux-mêmes mais pour les documents dont ils représentent le sujet. La structure des LI est d'ailleurs calquée (mais "appauvrie") sur celle des LN.

J.-Cl. Gardin [32] nomme *analyse documentaire* l'extraction de la signification de documents. Elle regroupe plusieurs traitements possibles (résumé, indexation, etc.) plus ou moins élaborés. Par rapport aux LN, les LI sont qualifiés de *métalangage*, soit un langage ou système de symboles adéquats utilisés pour exprimer le contenu d'un document en LN.

R. Fugmann [31] considère les LI et les LN comme complémentaires, utiles aussi bien les uns que les autres dans l'indexation, suivant le genre de *concepts* auxquels on a affaire. Les *concepts individuels* référant à un seul objet sont exprimés en LN par une seule expression lexicale. Par contre les *concepts généraux* référant à une multitude d'objets sont souvent rendus en LN par plusieurs expressions lexicales (synonymes), voire par des expressions non lexicales (périphrase), surtout dans le domaine scientifique où la terminologie est en retard sur les notions. Dans ce cas un LI contrôlé est plus avantageux que les LN, car il a précisément pour tâche de ne faire correspondre qu'une seule expression à un concept.

A contrario, J.E.L. Farradane [102] estime que l'approche linguistique n'offre aucun intérêt pour les théories de l'indexation. La linguistique n'est qu'un outil pour exprimer des idées et ne peut, à l'inverse de la psychologie, nous éclairer sur les relations effectivement en jeu dans le processus même de la connaissance.

4.2. Les deux axes de référence

Au début de ce siècle, le linguiste F. de Saussure distingue l'axe *syntagmatique* qui met en relation *in praesentia* deux termes de l'énoncé, alors que l'axe *associatif* (nous dirions aujourd'hui "paradigmatique") lie *in absentia* un terme de l'énoncé avec une "série virtuelle" de synonymes, termes génériques, etc.

J. Maniez [80] précise que les relations syntagmatiques appartiennent au *discours* et jouent dans le processus de la création d'une phrase, alors que les relations paradigmatiques appartiennent à la *langue* et sont hors de tout contexte.

Dès le début des années soixante, J.-Cl. Gardin [3] a clairement intégré cette opposition essentielle dans la construction d'un nouveau LI : le SYNTOL (Syntagmatic organizing language). Le SYNTOL est un "modèle général", car applicable à n'importe quel domaine scientifique. Le terme de "langage" est justifié par l'importance qu'il donne à la syntaxe dont l'élément de base est le "syntagme" : couple de mots-clés liés par une relation et noté $R_i(x, y)$. Les relations sont limitées au nombre de trois. Des règles formelles permettent de les manipuler et de les composer.

Ces deux axes permettent encore à J.-Cl. Gardin [32] de classer les LI (ou "métalangages") selon qu'ils comportent une *structure syntagmatique* (les relations de J.E.L. Farradane), une *structure paradigmatique* (les thésaurus), ou ni l'une ni l'autre, soit complètement "inorganisés" (les unitermes). Il montre le caractère fondamental du syntagme $R_i(x, y)$ qui permet de montrer l'identité structurelle des procédés syntaxiques tels que les opérateurs de rôle ou les relateurs de Farradane. Par combinaison de syntagmes, des relations à 3, 4, ..., n places sont représentées.

Dans le cas du thésaurus, K.T. Bivins [29] insiste sur l'influence de l'axe paradigmatique matérialisé par la *structure syndétique* (le jeu des renvois) dans le résultat de la recherche.

Les thésaurus contiennent pourtant des relations qui ne sont pas linguistiques, ni paradigmatiques, ni syntagmatiques et que J. Maniez [65] appelle *extra-sémantiques* : elles relient des termes et des faits réels. et correspondent à certaines relations dites habituellement "associatives".

4.3. Le triangle sémantique

Le *triangle sémantique* des linguistes C.K. Ogden et I.A. Richards, est constitué des *expressions* (ou "signifiants"), des *concepts* (ou "signifiés"), des *objets* (ou "référénts"), ainsi que de leurs relations.

Ce modèle est très important en documentation. Pour R. Fugmann [10], le pôle central de la recherche est le *concept*, non l'objet ou l'expression. Deux documents peuvent être pertinents à une demande même si leurs titres sont distincts. Il rappelle que S.R. Ranganathan distinguait déjà le *niveau des idées (idea plane)*, du *niveau verbal (verbal plane)*.

B. Long [37] modifie un peu le modèle en distinguant encore le *signifiant graphique* et le *signifiant phonologique*. Son étude consiste alors à passer en revue systématiquement toutes les relations à partir de ces quatre pôles. La face sémantique est prédominante en documentation.

4.4. Les universaux du langage

Dans les années soixante, la *linguistique structuraliste* (N. Chomsky, C.J. Fillmore, B. Pottiers, J. Lyons, etc.) a cherché à modéliser une "structure profonde" qui permet de rendre compte des "structures de surface" aussi diverses que sont les langues naturelles que nous connaissons. L'intérêt pour les sciences de l'information consiste à se demander si ces modèles ne seraient pas suffisamment fondamentaux pour expliquer aussi la structure des langages artificiels comme les LI.

Plus spécialement, des documentalistes ont compté sur ces travaux - par exemple la grammaire des cas de C.J. Fillmore - afin de leur fournir une liste des catégories fondamentales utilisables pour la syntaxe des LI (J. Maniez, [80] ; W.J. Hutchins, [34]).

J. Sorensen et D. Austin [97] s'intéressent aux structures profondes des linguistes et les comparent avec l'indexation dans PRECIS. En démontrant que celui-ci se situe déjà à un niveau de généralité élevé, il devient logiquement possible de l'employer dans un "contexte multilingue" pour l'indexation en chaîne avec d'autres LN que l'anglais.

Les documentalistes ont entrepris de leur côté, la recherche des universaux des LI. Les *facettes* de Ranganathan en sont un exemple. G. Bhattacharrya [91, 92] décrit une structure de base qui permet de dériver plusieurs SIL (subject indexing language) réels : indexation systématique, CC, DDC, etc. C'est à partir de ce modèle qu'il tire les "structures profondes" de POPSI.

C'est pourquoi F.J. Devadason [95] peut envisager l'utilisation de POPSI comme un *métalangage* permettant de générer automatiquement des thésaurus. Des chaînes formées en POPSI et enrichies de codes spéciaux sont introduites, et l'ordinateur produit le lexique du thésaurus avec sa structure syndétique.

2.4.5. La terminologie

W. Nedobity [39] oppose *linguistique* et *terminologie*. La linguistique est réduite à l'étude des LN qui sont des langues floues et dont la signification des termes dépend beaucoup du contexte. La terminologie au contraire attribue au terme un concept distinct et défini. Elle est ainsi plus utile pour les sciences de l'information qui s'occupent de langages spécialisés.

Le but du terminologue (G. Wersig [25]) est de rationaliser le processus de la communication. Son travail ressemble à celui de l'élaboration d'un thésaurus : dans un champ du savoir, il doit identifier d'abord les concepts, puis leur attribuer un terme (dans une ou plusieurs langues) tout en contrôlant leurs relations de synonymie, homonymies, etc. En plus selon une méthode déductive il donne une définition rigoureuse de chaque concept, tel celui de "thésaurus".

2.4.6. Le problème du sujet

Les tendances les plus récentes semblent tourner autour de la notion que les anglo-saxons nomment *aboutness*, très littéralement le "de quoi" un texte parle, c'est-à-dire le *sujet*.

J. Maniez [17] dans un ouvrage de synthèse récent part de l'opposition que les linguistes structuralistes font entre "thème" et "rhème" (ou "commentaire"), soit ce dont on parle et ce que l'on en dit. La *thématisation* est une activité essentielle en documentation au moyen des LI. Par opposition à l'énonciation, elle n'a pas de valeur informative, ce qui se remarque par l'absence de verbe.

De même pour K.P. Jones [13] l'indexation c'est identifier l'*aboutness* du document en question. Ce sujet peut être appréhendé en plusieurs niveaux d'analyse (selon les besoins), dont la structure textuelle.

C. Beghtol [28] traite en détail les implications linguistiques de l'analyse de l'*aboutness*. Traditionnellement, la linguistique a considéré la phrase comme l'unité de base. Mais pour élucider les mécanismes de compréhension du sujet, une *linguistique textuelle* considérant la totalité des phrases d'un document est nécessaire. Dans le cas d'une documentation idéale, un texte, son résumé et son indexation devraient représenter le même *aboutness*.

2.5. LANGAGES D'INDEXATION ET THEORIES MATHÉMATIQUES

L'intérêt de la modélisation mathématique n'est pas seulement de fournir des bases solides aux LI, mais ce travail exige de ceux qui l'accomplissent une définition rigoureuse, pas à pas des termes employés : descripteurs, classes, relations.

Le genre de mathématiques employé est très souvent la théorie des ensembles, dont certains termes ressemblent à ceux des LI. Le niveau de connaissance requis pour suivre ces études est variable. Celles de G. Beling [45], de V. Diodato [46] et de C.N. Mooers [48] sont très accessibles. De même la formalisation est plus ou moins importante.

Le domaine modélisé peut être très général et recouvrir tous les LI voire le système général d'information, ou bien il représente un type de LI, voire un aspect précis de ceux-ci. B.C. Vickery [24] en recense quelques-uns.

2.5.1. Modèles généraux

Un des premiers modèles, celui de J.W. Perry et A. Kent [49], est la partie théorique d'une étude consacrée à la réalisation pratique d'un système documentaire. Il s'agit en fait d'un "modèle réduit". Un lexique limité de termes répartis en catégories ou facettes. est donné au départ. Par *simulation* les propriétés de ce système fermé facile à manipuler sont observées. Par exemple il est facile de montrer que n facettes permettent de définir $n!$ classifications hiérarchiques différentes. Des définitions sont données : le *degré de caractère générique* d'une classe (rapport du nb. de descripteurs subordonnés sur le nb. total de descripteurs), *degré de synonymie*, etc.

A la même époque, C.N. Mooers [48] définit un modèle pouvant représenter un grand nombre de "systèmes d'indexation" et permettant même de déduire logiquement des systèmes possibles meilleurs mais non existant. Le problème de l'indexation est posé au niveau de l'utilisateur : comment fournir à une demande le "bon" sous-ensemble de documents parmi tous les documents du système. Mathématiquement, le système d'indexation s'exprime par une *transformation* entre l'ensemble des termes de la demande et l'ensemble des documents disponibles. Le modèle distingue alors plusieurs familles différentes.

D. Soergel [55] fournit un modèle pour tous les LI. Parallèlement à la recherche des universaux linguistiques, il décrit une base logico-mathématique générale. L'étude est la construction d'un *système formel*. Le *métalangage* mathématique et le *langage* du champ de la documentation qu'il représente sont bien distingués. Le lien entre l'un et l'autre est l'*interprétation* ou la *signification* du modèle. Au niveau du métalangage on se donne un lexique, et des règles de formation d'expressions. Des théorèmes en sont tirés. En restreignant le modèle à des cas particuliers, on peut décrire les LI réels.

S. Reball [50] constate que les LN sont *expressifs* par leur richesse sémantique mais peu *formalisés* ; alors que les LI sont au contraire *formalisés*, mais peu *expressifs*. Au moyen de la *logique des prédicats*, il construit un modèle de LI cumulant ces deux qualités.

L'étude de V. Diodato [46] présente de manière didactique les principaux concepts généralement utilisés pour représenter la *structure syndétique* des LI. La relation de synonymie entre deux descripteurs est mathématiquement une *relation d'équivalence*. Tous les synonymes d'un concept forment alors une *classe d'équivalence*. Les renvois d'un terme à l'autre sont représentés par des *applications* et les relations hiérarchiques par des *relations d'ordre partiel*. De la même façon G. Beling [45] étudie la structure des classifications. M.S. Sridhar [55] détaille les caractéristiques mathématiques de ces relations dans les thésaurus.

2.5.2. Modèles spécifiques

En Europe de l'Est surtout, des études plus ponctuelles ont été menées, particulièrement sur les thésaurus.

V.D. Sydorchenko [57, 75] définit le descripteur d'un thésaurus comme la *classe d'équivalence* de termes synonymes en LN. L'efficacité d'un LI dépend de sa "puissance sémantique" (*semantic power*) qui est le nombre d'expressions bien formées non synonymes qu'il permet de construire. Par la souplesse de leur combinatoire, les LI post-coordonnés ont une puissance sémantique supérieure aux LI pré-coordonnés.

W.M. Turski [59] donne un modèle de thésaurus compris comme un ensemble de "descripteurs" et d'"ascripteurs" liés par des relations de synonymie.

N.A. Stokolova [42] approfondit les aspects sémantiques. La pertinence d'un système d'information se mesure en comparant la *signification* de deux textes : celle de la demande et celle du document. L'objet des LI est de représenter cette signification. La *pertinence stricte* est une *relation d'équivalence* entre deux textes. Des *degrés de pertinence* sont aussi définis. Dans un autre travail [43] elle prend en considération les différents outils syntaxiques (liens, rôles,...) et mesure mathématiquement leur influence sur la puissance sémantique du LI.

En France, A. Dewèze [69] formalise la représentation des *relations sémantiques*. Dans la perspective de construire des thésaurus multilingues il adopte une théorie sémantique *extra-lexicale* qui se situe à un niveau supérieur de celui des LN. Un signifié est défini comme un ensemble de sèmes (obtenus par une analyse dont S.R. Ranganathan a montré la voie), et auxquels on peut ultérieurement attribuer des "étiquettes lexicales" en plusieurs langues. Les relations d'un LI sont décrites en étudiant, au moyen de *graphes*, les rapports entre ces "configurations sémiques". Par exemple, un niveau dans une monohiérarchie aura dans sa configuration sémique un sème de plus que le niveau qui lui est immédiatement supérieur.

Récemment, R. Schaüble [52] a proposé une nouvelle structure de l'information : l'"espace conceptuel" (*concept space*). Celui-ci permet de construire une théorie du thésaurus exposée comme un système formel.

2.6. CONCLUSIONS

Ce compte rendu montre la vitalité des recherches théoriques sur les LI menés depuis une quarantaine d'années. L'augmentation du rythme d'accroissement des publications - fait qu'il devient banal de souligner - a stimulé les façons de considérer l'indexation, clé d'accès à la documentation.

Il est significatif de constater que tous les LI traditionnels (classifications, vedettes-matière...) ont été continuellement remis en question durant cette période, et que même les nouveaux-venus (thésaurus...) n'ont pas échappé aux critiques.

2.7. APERCU CHRONOLOGIQUE

- 1852 *Grande-Bretagne*. P.M. Roget : *Thesaurus of english words and phrases*.
- ca. 1900 *Etats-Unis*. J.O. Kaiser : *Indexation systématique*.
- 1911 *Etats-Unis*. E.W. Hulme formule la notion de *caution des textes*.
- 1915 *France*. F. de Saussure : *Théorie des deux axes de référence : Rapports syntagmatiques et rapports associatifs (paradigmatiques)*.
- 1923 *Grande-Bretagne*. C.K. Ogden et I.A. Richards, *Triangle sémantique (expressions - concepts - objets)*.
- Années 30 *Inde*. S.R. Ranganathan expose la *Procédure en chaîne*. et la *méthode analytico-synthétique de classification*.
- 1948 *Grande-Bretagne*. Fondation du *Classification research group (C.R.G.)*, à Londres.
- 1950 *Grande-Bretagne*. J.E.L. Farradane propose une *théorie scientifique pour les classifications*, point de départ de *l'indexation relationnelle*.
- ca. 1951 *Etats-Unis*. M. Taube conçoit un système d'*indexation coordonnée au moyen d'unitermes*.
- 1954 *Grande-Bretagne*. Le biologiste J.K. Feibleman décrit la *théorie des niveaux d'intégration*.
- 1955 *Etats-Unis*. Fondation du *Centre for documentation and communication research* de la *Western reserve university (W.R.U.)* à Cleveland.
- 1957-1958 *Etats-Unis*. Premiers modèles mathématiques des LI de J.W. Perry - A. Kent et C.N. Mooers.
- 1959 *Etats-Unis*. Premier *thésaurus* au sens moderne du terme de la société Dupont de Nemours.
- Fin des années 50. Notions de *précision* et de *rappel*.
- Années 60. Travaux des *linguistes structuralistes*, N. Chomsky, C.J. Fillmore, J. Lyons.

- 1960 *Etats-Unis*. H.P. Luhn développe le système KWIC entièrement mécanisé, précurseur des *LI en chaîne* .
- 1962 *Inde* : Fondation du *Documentation research and training centre* à Bangalore, centre de l' "école indienne" initiée par S.R. Ranganathan.
- 1963 *Etats-Unis*. Thésaurus de l'Engineers Joint council, avec *indicateurs de rôles*.
- 1964 *France*. J.C. Gardin : le modèle général SYNTOL.
- 1964 *Grande-Bretagne*. R.M. Needham et K. Sparck Jones : approche automatique des classifications et notion d'*amas*
- 1967 *R.F.A.* D. Soergel : *formalisation mathématique des LI*.
- 1969 *Grande-Bretagne*. J. Aitchison présente le *Thesaurofacet* thésaurus disposant les concepts analysés en facettes.
- ca. 1969 *Inde*. G. Bhattacharrya, développement du LI en chaîne *POPSI*.
- 1971 *Grande-Bretagne*. D. Austin teste un prototype de *PRECIS* à la BNB.
- 1974 *R.F.A.* I. Dahlberg : recherche interdisciplinaire des *fondements universels de l'ordre du savoir* et création de la revue *International classification*.
- 1974 *R.F.A.* R. Fugmann remet en cause les principes des thésaurus et développe un système d'indexation au moyen de graphes *TOSAR*.
- 1977-1978 *Canada*. T.C. Craven : LI en chaîne *NEPHIS* et *LIPHIS*.
- 1979 *Inde* : G. Bhattacharrya. Recherche les fondements communs des LI avec le *Subject indexing language (S.I.L.)*.
- 1981 *France*. A. Dewèze, *réseaux sémantiques extra-lexicaux* appliqués à la constitution des thésaurus.
- 1986 *Grande-Bretagne*. C. Beghtol analyse la notion de *sujet (aboutness)* au moyen de la *linguistique textuelle*.
-

2.8. LISTE DES ABREVIATIONS

BNB	British national bibliography
CC	Colon classification
C.R.G.	Classification research group
DDC	Decimal Dewey classification
KWIC	Keyword in context
LCC	Library of Congress classification
LI	Langage d'indexation
LIPHIS	Linked phrase indexing system
LN	Langage naturel
NEPHIS	Nested phrase indexing system
POPSI	Postulate-based permuted subject indexing
PRECIS	Preserved context indexing system
S.I.L	Subject indexing language
SYNTOL	Syntagmatic organizing language
TOSAR	Topological representation of synthetic and analytical relations of concepts

3. BIBLIOGRAPHIE

3.1. THEORIES ET MODELES

3.1.1. Généralités

1. **Bhattacharyya, G.** Fundamentals of subject indexing languages. In : **Neelameghan, A.** (ed.). *Ordering systems for global information network : proceedings of the third international study conference on classification research, 1975, jan. 6-11, Bombay, India.* Bangalore : International federation for documentation ; Committee on classification research (FID/CR) ; Sarada Ranganathan endowment for library science, 1979. 511 p. (Sarada Ranganathan endowment for library science series, n. 13 ; FID Publication, n. 533), p. 83-89.
2. **Cooper, Williams S.** Indexing documents by Gedanken Experimentation. *Journal of the american society for information science.* V. 29, n. 3, 1978, p. 107-119.
3. **Cros, René-Charles ; Gardin, Jean-Claude, Lévy, Francis.** *L'automatisation des recherches documentaires : un modèle général "le SYNTOL".* Paris : Gauthier-Villars, 1964.
4. **Dahlberg, Ingeraut.** *Grundlagen universaler Wissensordnung.* München : Verlag Dokumentation, 1974. (DGD-Schriftenreihe, v. 3).
5. **Foskett, Douglas J.** *Classification for a general index language.* London : Library association, 1970. 48 p. (Research pamphlet, 2).
6. **Foskett, Douglas J.** Systems theory and its relevance to documentary classification. *International classification.* V. 7, n. 1, 1980, p. 2-5.
7. **Fugmann, Robert.** The glamour and the misery of the thesaurus approach : treatise IV in information retrieval theory. *International classification.* V. 1, n. 2, 1974. p. 3-15.

8. **Fugmann, Robert.** Toward a theory of information supply and indexing : treatise VI on retrieval system theory. *International classification*. V. 6, n. 1, 1979, p. 3-15.
9. **Fugmann, Robert.** On the practical of indexing and its theoretical foundations. *International classification*. V. 7, n. 1, 1980, p. 13-20.
10. **Fugmann, Robert.** The five-axiom theory of indexing and information supply. *Journal of the american society for information science*. V. 36, n. 2, 1985, p. 116-129.
11. **Gardin, Jean-Claude.** Eléments d'un modèle pour la description des lexiques documentaires. *Bulletin des bibliothèques de France*. V. 11, n. 5, 1966, p. 171-182.
12. **Grolier, Eric de.-** *Etude sur les catégories générales applicables aux classifications et codifications documentaires*. Paris : UNESCO, 1962. 262 p. (Documentation et terminologie scientifiques).
13. **Jones, Kevins P.** Towards a theory of indexing. *Journal of documentation*. V. 32, 1976, p. 118-123.
14. **Lancaster, F. W.** *Vocabulary control for information retrieval*. 2nd ed. Arlington, Virg. : Information Resources Press, 1986. XVII-270 p.
15. **Landry, Bertrand C. ; Rush, James E.** Toward a theory of indexing. *American society for information science, proceedings vol. 5. 31st Annual meeting, october 20-24, 19olumbus, Ohio*. New York : Greenwood Press, 1968, p. 59-64.
16. **Landry, Bertrand C. ; Rush, James E.** Toward a theory of indexing. II. *Journal of the american society for information science*. V. 21, 1970, p. 358-367.
17. **Maniez, Jacques.** *Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires*. Paris : Ed. d'organisation, 1987. 291 p. IIe partie : Etude générale des langages documentaires. p. 235-286.
18. **Ranganathan, Shiyali Ramamrita.** *Prolegomena to library classification*. 3rd. ed. Bombay, Calcutta : Asia Publishing House, 1967. 640 p. (Ranganathan series in library science, 20).
19. **Rothmann, John.** Index, indexer, indexing. *Encyclopedia of library and information science*. New York : Dekker, 1974. V. 11, p. 286-299.

20. **Soergel, Dagobert.** Some remarks on information languages, their analysis and comparison. *Information storage and retrieval*. V. 3, 1967, p. 219-291.
21. **Soergel, Dagobert.** Theoretical problems of thesaurus building with particular reference to concept formation. In : **Petőfi, J. S. ; Podlech, A. ; Savigny, E. V.** (eds). *Fachsprache-Umgangssprache*. Kronberg/Taunus : Scriptor-Verlag, 1975. p. 355-382.
22. **Svenonius, Elaine F.-** Directions for research in indexing, classification and cataloging. *Library resources and technical services*. V. 25, n. 1, 1981, p. 88-103.
23. **Taube, Mortimer ; Gull, C. D. ; Wachtel, Irma S..** Unit terms in coordinate indexing. *American documentation*. V. 3, n. 4, 1952, p. 213-218.
24. **Vickery, B. C.** *Information systems*. London : Butterworth, 1973. Ch. 8 : Retrieval language models, p. 203-222.
25. **Wersig, Gernot.** Eine neue Definition von "Thesaurus". *Nachrichten für Dokumentation*. V. 20, n. 2, 1969, p. 53-61.

3.1.2. Théories et modèles linguistiques

26. **Ballmer, Thomas T.** Zur Gewinnung einer Fundamental Klassifikation des menschlichen Wissens. II. *International classification*. V. 10, n. 3, 1983, p.127-134.
27. **Ballmer, Thomas T.** Zur Gewinnung einer Fundamental Klassifikation des menschlichen Wissens. IV. *International classification*. V. 11, n. 1, 1984, p. 21-27.
28. **Beghtol, Clare.** Bibliographic classification theory and text linguistics : aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of documentation*. V. 42, n. 2, 1986, p. 84-113.
29. **Bivins, Kathleen T.** An approach to an atomic indexing. In Taylor, Peter J. (ed.).- *New trends in documentation and information : proceedings of the 39th FID congress. University of Edinburgh, 25-28 sept. 1978*. London : Aslib, 1980. p. 55-50. (FID publication, 566).
30. **Coyaud, Maurice.** *Introduction à l'étude des langages documentaires*. Paris : Klincksieck, 1966. 148 p. (T.A. Documents, 1).

31. **Fugmann, Robert.** The complementarity of natural and indexing languages. *International classification*. V. 9, n. 3, 1982, p. 140-144.

32a. **Gardin, Jean-Claude.** Document analysis and linguistic theory. *International classification*. V. 9, n. 2, 1973, p. 137-168.

Traduit dans :

32b **Gardin, Jean-claude.** *Les analyses de discours*. Neuchâtel : Delachaux et Niestlé, 1974. 178 p. Ch. IV : "Analyse documentaire et théorie linguistique", p. 120-168.

33. **Henriksen, Tor.** A linguistic description model for indexing languages. In : In : **Neelameghan, A.** (ed.).- *Ordering systems for global information network : proceedings of the third international study conference on classification research, 1975, jan. 6-11, Bombay, India*. Bangalore : International federation for documentation ; Committee on classification research (FID/CR) ; Sarada Ranganathan endowment for library science, 1979. 511 p. (Sarada Ranganathan endowment for library science series, n° 13 ; FID Publication, n° 533). p. 3-8.

34. **Hutchins, W. J.** *Languages of indexing and classification : a linguistic study of structures and functions*. Stevenage, Herts. : Peregrinus, 1975. 148 p. (Librarianship and information studies, v. 3).

35. **Jones, Kevin P.** The environment of classification. II. How we classify. *Journal of the american society for information science*. V. 25, n. 1, 1974, p. 44-51.

36. **Kim, Chai.** Theoretical foundations of thesaurus-construction and some methodological considerations for thesaurus-updating. *Journal of the american society for information science*. V. 24, 1973, p. 148-156.

37. **Long, B.** Linguistique et indexation. *Documentaliste*. V. 17, 1980, p. 99-106.

38a. **Melton, Jessica S.** *A use for the techniques of structural linguistics in documentation research*. Cleveland, Ohio : Western Reserve University, Center for documentation and communication research, 1964. 20 p.

Egalement dans :

38b. **Atherton, Pauline** (ed.). *Classification research. Proceedings of the second international study conference. Elsinore, Denmark, sept. 14-18, 1964*. Copenhagen : Munksgaard, 1965. p. 466-480.

39. **Nedobity, Wolfgang.** Terminology and its application to classification indexing and abstracting. *UNESCO Journal of information sciences, librarianship and archives administration*. V. 5, n. 4, 1983, p. 227-234.
40. **Shreider, Yu. A.** The thesaurus in information science and theoretical semantics. *Automatic documentation and mathematical linguistics*. V. 5, n. 1, 1971, p. 52-57.
41. **Sparck Jones, Karen ; Kay, Martin.** *Linguistics and information science*. New York ; London : Academic Press, 1973. 244 p. (FID publication, 492).
42. **Stokolova, N. A.** Elements of a semantic theory of information language. I. The concepts of relevance and information language. *Information processing and management*. V. 13, 1977, p. 227-234.
43. **Stokolova, N. A.** Syntactic tools and semantic power of information languages. (Pt. II of "elements of a semantic theory of information retrieval"). *International classification*. V. 3, n. 2, 1976, p. 75-81.

3.1.3. Théories et modèles mathématiques

44. **Barnes, R. F.** *Mathematico-logical foundations of retrieval theory : general concepts and methods*. Bethlehem, Pa : Lehigh University center for the information science, 1965. 29 p.
45. **Beling, Gerd.** Towards a set theoretical foundation of classification. In : **Neelameghan, A.** (ed.)- *Ordering systems for global information network : proceedings of the third international study conference on classification research, 1975, jan. 6-11, Bombay, India*. Bangalore : International federation for documentation ; Committee on classification research (FID/CR) ; Sarada Ranganathan endowment for library science, 1979. 511 p. (Sarada Ranganathan endowment for library science series, n. 13 ; FID Publication, n. 533). P. 77-82.
46. **Diodato, Virgil.** Modern algebra and information science. *Journal of the american society for information science*. V. 34, n. 4, 1983, p. 257-261.
47. **Heilprin, Laurence B.** *Mathematical model of indexing*. Washington : Documentation incorporation, 1957. 26 p.
48. **Mooers, Calvin N.** A mathematical theory of language symbols in retrieval. *Proceedings of international conference on scientific information*. Washington, 1958. p. 1327-1364.

49. Perry, James W. ; Kent, Allen. *Documentation and information retrieval ; an introduction to basic principles and cost analysis*. Cleveland, Ohio : The Press of Western Reserve University and interscience publishers, 1957. Ch. II : "A mathematical model system", p. 4-25.
50. Reball, Siegfried. On the application of predicate calculus in information indexing and the degree of indexing languages formalisation. *International forum on information and documentation*. V. 3, n. 3, 1978, p. 14-17.
51. Richmond, Phyllis A. Classification in syndetic structures. In : Dahlberg, Ingetraut (ed.).- *Universal classification 1 : Subject analysis and ordering systems. Proceedings 4th international study conference on classification research ; 6th Annual conference on research and development in information retrieval*. New York : ACM Press, 1987. p. 254-262.
52. Schauble, P. Thesaurus based concept spaces. *Proceedings of the tenth Annual international ACM SIGIR conference on research and development in information retrieval*. New York : ACM Press, 1987. p. 254-262.
53. Schreider, Yu. A. The logic of classification. *Automatic documentation and mathematical linguistics*. V. 7, n. 2, 1973, p. 30-36.
54. Schreider, Yu. A. The duality of classification : taxonomy and meronomy. *International forum on information and documentation*. V. 6, n. 1, 1981, p. 3-10.
55. Soergel, Dagobert. Mathematical analysis of documentation systems : an attempt to a theory of classification and search request formulation. *Information storage and retrieval*. V. 3, 1967, p. 129-173.
56. Sridhar, M. S. A mathematical approach to relations in thesaurus. *Journal of library and information science (India)*. V. 5, n. 1, 1980, p. 76-87.
57. Sydorchenko, V. D. On the construction of a descriptor model. *Annals of library science and documentation*. V. 21, n. 4, 1974, p. 142-145.
58. Taube, Mortimer ; Wachtel, Irma S. The logical structure of coordinate indexing. *American documentation*. V. 4, 1953, p. 67-68.
59. Turski, W. M. On a model of information retrieval based on thesaurus. *Information storage and retrieval*. V. 7, 1971, p. 89-94.

3.2. RELATIONS ENTRE LES TERMES D'INDEXATION

3.2.1. Généralités

60. Austin, Derek. Basic concept classes and primitives relations. In : Dahlberg, Ingetraut (ed.).- *Universal classification 1 : Subject analysis and ordering systems. Proceedings 4th international study conference on classification research ; 6th Annual conference of Gesellschaft für Klassifikation e. V., Augsburg, 28 june-2 july 1982.* Frankfurt : Indeks Verlag, 1982. p. 86-94.

61a. Coates, E. J. Significance and term relationship in compound headings. In : *Subject catalogues.* London : Library association, 1960. p. 51-64.

Repris dans :

61b. Chan, Lois Mai ; Richmond, Phyllis A. ; Svenonius, Elaine (eds). *Theory of subject analysis. A source book.* Littleton, Col. : libraries unlimited, 1985. p. 183-195.

62. Coates, E. J. Some properties of relationships in the structure of indexing languages. *Journal of documentation.* V. 29, n. 4, 1973, p. 390-404.

63. Gull, C. D. Structure of indexing authority lists. *Library resources and technical services.* V. 10, fall 1966, p, 507-511.

64. Kervégant, Désiré. Classification et analyse des relations. *Bulletin des bibliothèques de France.* N. 11, 1959, p. 495-511.

65. Maniez, Jacques. Relationships in thesauri : some critical remarks. *International classification.* V. 15, n. 3, 1988. p. 133-138.

66. Needham, R. M. ; Sparck Jones, Karen. Keywords and clumps. *Journal of documentation.* V. 20, 1964, p. 5-15.

67. Willets, M. An investigation of the nature of the relation between terms in thesauri. *Journal of documentation.* V. 31, n. 3, 1975, p. 158-184.

3.2.2. Relations sémantiques

68. **Beghtol, Clare.** Semantic validity : concepts of warrant in bibliographic classification systems. *Library resources and technical services*. V. 30, n. 2, 1986, p. 109-125.
69. **Dewèze, André.** Réseaux sémantiques : essai de modélisation - Application à l'indexation et à la recherche de l'information documentaire. Lyon, 1981. (Th. univ. : Lyon 1 : 1981).
70. **Frohmann, Bernard P.** An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG. *Cataloging and classification quarterly*. V. 4, n. 1, 1983, p. 11-27.
71. **Ghose, A. ; Dhawle, A. S.** Problems of thesaurus construction. *Journal of the american society for information science*. V. 28, n. 4, 1977, p. 211-217.
72. **Jansen, R.** Die Bedeutung der Zugehörigkeitsrelation für Thesaurusstruktur, Indexierungsmethode und Recherche. In : *Deutsche Gesellschaft für Dokumentation (ed.). Deutscher Dokumentartag, 1974. Band I. Bonn, 7-11 Okt. 1974. München : Verlag Dokumentation. p. 354-374.*
73. **Popovska, H.** Synonymic and hierarchical relations in descriptor languages. *Automatic documentation and mathematical linguistics*. V. 19, n. 2, 1985, p. 8-15.
74. **Soergel, Dagobert.** *Organizing information : principles of data base and retrieval systems*. Orlando, Fl. : Academic Press, 1985. Ch. XX "Index language structure. I : conceptual", p. 251-287.
75. **Sydorchenko, V. D.** Typological classification of thesaurus semantic structures. *Automatic documentation and mathematical linguistics*. V. 9, n. 2, 1975, p. 16-23.
76. **Yu, Clement T. ; Raghavan, Vijay V.** Single-pass method for determining the semantic relationships between terms. *Journal of the american society for information science*. V. 28, n. 6, 1977, p. 345-354.

3.2.3. Relations syntaxiques

77. Aitchinson, J. Thesaurofacet : a new concept in subject retrieval schemes. In : Wellisch, H. ; Wilson, T. D. (eds). *Subject retrieval in the seventies : new directions. Proceedings of an international symposium. University of Maryland, 1971.* Westport, Ct. : Greenwood Press, 1972. p. 72-98.
78. Costello, J. C. A basic theory of roles as syntactical control devices in coordinate indexes. *Journal of chemical documentation.* V. 4, 1964, p. 116-124.
79. Hutchins, W. J. Facets, roles and cases. In : informatics 1. *Proceedings of a conference held by the Aslib co-ordinate indexing group on 11-13 april 1973 at Durham University.* London : Aslib, 1974. p. 89-97.
80. Maniez, Jacques. *Le rôle de la syntaxe dans les systèmes de recherche documentaire.* Dijon : I.U.T. de Dijon, Département carrières de l'information, 1976. 184 p. Tome I : "Aspects linguistiques". Tome II : "Etude critique de quelques S.R.D.".

3.3. LANGAGES ET SYSTEMES D'INDEXATION

3.3.1. Indexation en chaîne

3.3.1.1. Généralités

81. Armitage, Janet E. ; Lynch, Michael F. Some structural characteristics of articulated subject indexes. *Information storage and retrieval.* V. 4, 1968, p. 101-111..
82. Batty, C. D. Chain indexing. *Encyclopedia of library and information science.* New York : Dekker, 1970. V. 4, p. 423-434.
83. Bhattacharyya, G. Chain procedure and structuring of a subject. *Library science with a slant to documentation.* V. 9, 1972, p. 585-635.
84. Craven, Timothy C. *String indexing.* Orlando, Fl. : Academic Press, 1985. 270 p.
85. Farradane, Jason E. L. A comparison of some computerproduced permuted alphabetical subject indexes. *International classification.* V. 4, n. 2, 1977, p. 94-101

86. Mineur, B. W. Relations in chains. *Journal of librarianship*. V. 5, n. 3, 1973, p. 175-202.
87. Rajan, T. N. From chain indexing to string indexing. *Library herald*. V. 20, n. 2-4, 1977, p. 105-110.
88. Rajan, T. N. (ed.). *Indexing systems : concepts models and techniques*. Calcutta : Indian association of special libraries and information centres (IASLIC), 1981. 270 p.

3.3.1.2. PRECIS et POPSI

89. Austin, Derek. An information retrieval language for MARC. *Aslib proceedings*. V. 22, n. 10, 1970, p. 481-491.
- 90a. Austin, Derek ; Digger, Jeremy A. PRECIS : the preserved context index system. *Library resources and technical services*. V. 21, n. 1, 1977, p. 13-30.

Repris dans :

- 90b. Chan, Lois Mai ; Richmond, Phyllis A. ; Svenonius, Elaine (eds.). *Theory of subject analysis : a source book* Littleton, Col. : Libraries Unlimited, 1985, p. 372-389.
91. Bhattacharyya, G. POPSI : its fundamentals and procedure based on a general theory of subject indexing languages. *Library science with a slant to documentation*. V. 16, n. 1, 1979, p. 1-34.
92. Bhattacharyya, G. Some significant results of current classification research in India. *International forum on information and documentation*. V. 6, n. 1, 1981, p. 11-18.
93. Biswas, Subal C. Two decades of POPSI, 1969-1988 : a literature review. *Libri*. V. 38, n. 4, 1988, p. 279-296.
94. Cheti, A. Sistemi d'indicizzazione preordinati : indicizzazione a catena, PRECIS, POPSI, NEPHIS. *Bolletino d'informazioni (Associazione italiana biblioteche)*. V. 24, n. 1, 1984, p. 19-44.
95. Devadason, F. J. Postulate-based permuted subject indexing language as a metalanguage for computer-aided generation of information retrieval thesaurus. *International forum on information and documentation*. V. 8, n. 1, 1983, p. 22-29.

96. **Mahapatra, M. ; Biswas, S. C.** PRECIS : its theory and application. An extended state-of-the-art review from the beginning up to 1982. *Libri*. V. 33, n. 4, 1983, p. 316-330.
97. **Sorensen, Jutta ; Austin, Derek.** PRECIS in a multi-lingual context. Part 2. A linguistic and logical explanation of the syntax. *Libri*. V. 26, n. 2, 1976, p. 108-139.

3.3.1.3. Autres systèmes

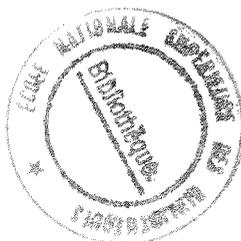
98. **Craven, Timothy C.** NEPHIS : a nested phrase indexing system. *Journal of the american society for information science*. V. 28, n. 2, 1977, p. 107-114.
99. **Craven, Timothy C.** Linked phrase indexing. *Information processing and management*. V. 14, n. 6, 1978, p. 469-476.
100. **Farradane, Jason E. L.** A scientific theory of classification and indexing and its practical applications. *Journal of documentation*. V. 6, n. 2, 1950. p. 83-99.
101. **Farradane, Jason E. L.** A scientific theory of classification and indexing : further considerations. *Journal of documentation*. V. 8, n. 2, 1952. p. 73-92.
102. **Farradane, Jason E. L.** Relational indexing and classification in the light of recent experimental work in psychology. *Information storage and retrieval*. V. 1, 1963, p. 3-11.
- 103a. **Kaiser, Julius Otto.** Systematic indexing. In : *Aslib. report of proceedings of the 3rd conference. Oxford, sept. 24-27, 1926.* London : Aslib, 1927. p. 20-33.

Repris dans :

- 103b. **Olding, R. K. (ed.).** *Readings in library cataloguing.* London : Crosby Lockwood, 1966. p. 145-162.
104. **Svenonius, Elaine.** Facet definition : a case study. *International classification*. V. 5, n. 3, 1978, p. 134-141.

3.3.2. Théorie des graphes : une application

105. Fugmann, Robert ; Nickelsen, Herbert ; Nickelsen Ingeborg [et al.]. Representation of concept relations using the TOSAR system of the IDC. Treatise III on information retrieval theory. *Journal of the american society for information science*. V. 25, n. 5, 1974, p. 287-307.
106. Nickelsen, Herbert ; Nickelsen Ingeborg. Mathematische Analyse des TOSAR Verfahrens. *Information storage and retrieval*. V.9, n. 2, 1973, p. 95-119.





* 9 5 4 4 2 0 6 *