

CONSTRUCTION DES LANGAGES D'INDEXATION

Aspects théoriques

Alexis RIVIER

Cette contribution est l'aboutissement d'une recherche portant sur les études consacrées, dans la littérature spécialisée, aux fondements théoriques des langages d'indexation. Plus d'une centaine de références ont été rassemblées, que l'on trouvera toutes à la fin de cet article. Après avoir présenté sa méthode d'investigation, l'auteur examine ici, en une synthèse des travaux recensés, l'évolution depuis une quarantaine d'années des principes théoriques présidant à la conception des classifications, des langages d'indexation en chaîne et des thésaurus, ainsi que l'apport des théories linguistiques et mathématiques aux langages d'indexation.

D'après certains auteurs (1), les études théoriques sur les fondements des langages d'indexation (LI) n'ont pas retenu suffisamment l'attention des chercheurs. Les considérant comme de simples outils, on pense habituellement que de tels langages se justifient par leur seule efficacité. La littérature est très abondante en manuels pratiques sur la construction des langages d'indexation ainsi qu'en évaluations de leurs performances, alors que les textes en exposant les principes théoriques sont moins nombreux et moins connus. Ce serait déjà une raison suffisante pour chercher à les rassembler. De plus, il apparaît clairement à la lecture de ces textes qu'une meilleure connaissance interne des langages d'indexation conduit souvent à améliorer leur conception.

Cet article fait suite à un DESS en informatique documentaire préparé et soutenu à l'Université de Lyon 1 sous la direction de Richard Bouché.

Alexis Rivier est responsable du projet d'informatisation de la Bibliothèque cantonale du Valais, 9 rue des Vergers, CH - 1951 Sion, tél. 027 / 21 63 21.

Méthode de recherche

Terminologie. Préciser les termes de notre étude a son sens dans la mesure où cela permet de justifier et de réitérer la recherche bibliographique effectuée. Celle-ci 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

(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*, 1984, vol. 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) 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.*

toutes les disciplines scientifiques, d'où sa plurivocité.

Les *langages d'indexation* (3) sont des langages artificiels – c'est-à-dire construits à l'aide d'un ensemble de règles données – servant à la représentation abrégée du contenu d'un document. Mais tous les langages d'indexation ne sont pas aussi rigoureusement codifiés : il y a des degrés. En nous inspirant de J. Maniez [17], nous pouvons les représenter selon deux axes : celui de la coordination et celui du contrôle (voir encadré page suivante).

A vrai dire, les LI peu contrôlés (comme les descripteurs choisis librement pour représenter le contenu d'un document) se rapprochent des langages naturels (LN). Nous nous intéressons au contraire aux LI les plus contrôlés, car c'est précisément un trait qui les différencie nettement de la langue naturelle. Il s'agit des LI postcoordonnés (la combinaison des descripteurs se fait au moment de la recherche documentaire) comme les thésaurus ; ainsi que des LI précoordonnés (la combinaison des termes est fixée au moment de l'indexation)

comme les classifications, et surtout les langages en chaîne (PRECIS, SYNTOL, POPSI, etc.), dotés d'une syntaxe contraignante.

Quant au terme *théorie*, il a le sens de construction intellectuelle ou spéculative. Il s'oppose bien sûr à *pratique*. Un type important de démarche théorique qui a guidé notre recherche est la modélisation. Le *modèle* est un outil intellectuel servant de principe explicatif d'un domaine de la connaissance. Ainsi beaucoup de modèles linguistiques ou mathématiques ont été employés pour représenter la structure des langages d'indexation.

Délimitations. Notre travail ne vise pas l'exhaustivité, mais se veut une synthèse. La littérature est parfois redondante. Les textes ont été sélectionnés pour leur valeur novatrice, leur importance pour la recherche théorique, ou pour leur qualité de synthèse précisément.

ABRÉVIATIONS UTILISÉES DANS LE TEXTE

BNB	British national bibliography
CC	Colon classification
CRG	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
SIL	Subject indexing language
SYNTOL	Syntagmatic organization language
TOSAR	Topological representation of synthetic and analytical relations of concepts

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), l'intérêt des théories des LI n'est pas mise en cause avec le temps. La bibliographie qui est

(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*.

(5) Pour les indications pratiques, cf. le *Répertoire des banques de données professionnelles*. 11^e éd. Paris : ADBS-ANRT, 1989.

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

Représentation des langages d'indexation en fonction de leurs degrés de coordination et de contrôle

présentée ici 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).

Instruments. Cette recherche bibliographique a été menée à la fois par consultation de répertoires imprimés et de banques de données.

Les instruments que nous avons utilisés pour la recherche manuelle sont présentés en encadré, selon l'ordre d'élaboration intellectuelle des types de sources : bibliographies courantes et rétrospectives, bibliographies thématiques, synthèses bibliographiques.

Pour ce qui concerne la recherche automatisée, quatre banques de données ont été interrogées : PASCAL, INSPEC, LISA et ISA (5). La grande difficulté avec l'interrogation des banques de données réside dans la

manière d'établir les équations de recherche. Ainsi le concept *aspect théorique* se traduit très mal en descripteurs et le terme *théorie* est utilisé de manière très extensive par les indexeurs. Les banques de données consultées ont donné pour cette raison des taux de pertinence assez faibles. Nous indiquons ci-après les équations de recherche utilisées.

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

INSPEC (Questel+) : 1. LIM CT72+/CC/T, 2. (INFORMATION ANALYSIS OU THESAUR+/DE/T) ET (THEORY OU THEORIES OU MODEL?/T).

LISA et ISA (Dialog) : 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).

LE RENOUVELLEMENT DES CLASSIFICATIONS

Le Classification research group

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 (CRG). Un

de ses membres, D.J. Foskett [5], dresse le panorama des thèmes abordés par ce groupe jusqu'en 1970. La classification n'est plus considérée comme le seul rangement physique de livres sur des rayons, mais comme un authentique langage dont le but est la recherche documentaire.

Les théories de S.R. Ranganathan en Inde ont eu une influence considérable sur le CRG 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 sé-

Du côté de l'Europe de l'Est

Y.A. Schreider [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 mêmes ? La logique et la classification aristotéliennes offrent une base de réflexion. Une notion peut être décrite en *extension* ou en *intension*. L'exten-

sion 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 : la *taxonomie*, où les objets sont représentés sur la base de leurs similarités, et la *méronomie*, soit l'identification des caractéristiques essentielles communes d'un taxon.

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 langage d'indexation en chaîne puisque les analets peuvent avoir une structure bidimensionnelle.

Face à des systèmes formalisés comme POPSI et PRECIS, NEPHIS [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. C'est pourquoi LIPHIS, une version améliorée du système, donne justement la possibilité de *lier* (*link*) latéralement les éléments d'un sujet dont la structure ne peut se réduire à une pure hiérarchie [99].

ÉVOLUTION 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* qui se fonde sur la *procédure en chaîne* de S.R. Ranganathan, et *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, qui produit la *British national bibliography* (BNB), 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. Bhattacharyya lance POPSI. Dans les années soixante-dix, d'autres langages apparurent : NEPHIS et LIPHIS de T.C. Craven.

Récemment, des comparaisons de ces langages 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 préordonné. Comme ils servent à l'édition d'index imprimés traditionnels, chaque terme doit apparaître en tête de chaîne pour être accessible. Deux problèmes doivent alors être résolus : celui de l'ordre de succession des termes (une

« grammaire ») et celui des règles de permutation des termes.

Ordre des termes

L'*indexation systématique* de J.O. Kaiser [104 ; E. Svenonius, 106] 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 : ainsi les *choses* (stables et déterminées) précèdent les *matériaux* (stables mais non déterminés) et les *actions* (instables et dynamiques).

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 et précise le terme qui représente son contexte (« Asie - Inde », par exemple). 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

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 consiste à retrancher successivement les termes de gauche. 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 autres termes représentant le sujet. Le défaut du système est qu'un sujet spécifique (ABCD) figurera sous des entrées trop générales (CD, D). Ce problème est celui de la *chaîne disparaissante* (*disappearing chain*) [S.C. Biswas, 93], et c'est 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] formalisent la production d'*index matières articulés* par permutation d'éléments syntaxiquement liés dans la langue naturelle par des prépositions, sans introduire d'ambiguïtés.

La permutation des termes permet de respecter le principe de spécificité. Mais elle 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 reconstruites.

DÉVELOPPEMENT DES THÉSARUS

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 thésaurus modernes est également de représenter d'abord des concepts, mais aussi de préciser leurs relations.

Le thésaurus 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 postcoordonnée »). 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 qui correspondent 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 postcoordonnée, avec le succès que l'on sait. Comme les unitermes de M. Taube étaient choisis de manière encore empirique, le contrôle du vocabulaire et 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 thésaurus 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 thésaurus.

Du fait que les *relations syntaxiques* étaient absentes des thésaurus, il en résultait à la recherche de mauvaises coordinations. L'Engineers joint council présente en 1964 le premier thé-

saurus avec *indicateurs de rôle* permettant d'indiquer à chaque descripteur une fonction. J.C. Costello [78] compare ce procédé avec les langues naturelles 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. Aitchinson 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 précoordonné dans la production de chaînes pour un index, que de système postcoordonné de recherche. Deux traditions d'indexation très différentes (américaine et « indo-européenne ») se rejoignent ainsi.

Etude critique des thésaurus en Europe

Après 1970, les spécialistes prennent conscience d'inconvénients plus fondamentaux. L'accroissement des fonds documentaires, la parcellisation des savoirs conflent 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 sont repensés.

Dans le cadre d'une recherche fondamentale sur les langages d'indexation, 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'ordre alphabétique disperse les termes de sens proche qui ne sont que virtuellement reliés par les renvois. Plus le terme approprié est difficile à trouver (il l'est 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 resynthétiser afin de préserver les liens entre eux.

Sur ces bases théoriques, R. Fugmann a développé un système d'indexation bidimensionnel appelé

TOSAR, utilisé pour la documentation automatisée en chimie [103].

Comme aide à l'indexation, les thésaurus traditionnels proposaient des *schémas fléchés* (*arrowgraphs*) représentant les descripteurs et leurs relations sémantiques. TOSAR va beaucoup plus loin. Le système exploite la structure et les propriétés de la *théorie des graphes*: 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.

Dans les applications documentaires, 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 et retiendra ceux qui ont la même structure.

La méthode TOSAR a été décrite formellement au moyen de la logique des prédicats par H. et I. Nickelsen [105].

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 à peu près adéquats.

D. Soergel [74] définit de façon générale 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 multimodèle: 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 multimodèle.

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 CRG et les neuf opérateurs de J.E.L. Farradane, elle étudie tout particulièrement les *relations dites affirmatives*. Ce type regroupe habituellement toutes sortes de relations que les chercheurs renoncent 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 précoordonnés 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*).

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.

Les deux axes de références

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 sont en jeu dans le processus de la création d'une phrase, alors que les relations paradigmatiques appartiennent à la *langue* et existent 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 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 $Ri(x,y)$. 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 ne comportent ni l'une ni l'autre parce qu'ils sont complètement « inorganisés » (les unitermes). Il montre le caractère fondamental du syntagme $Ri(x,y)$ qui met en évidence 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 syntag-

LANGAGES D'INDEXATION ET THÉORIES 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 langages naturels et certains en dérivent plus ou moins profondément. Il est donc naturel de 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 problème du passage d'un langage 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 manipulent : 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. 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 langages d'indexation : il les compare sous tous les aspects (formels, sé-

mantiques, pragmatiques...) avec les langues naturelles.

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 rédigé 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 (périphrases), 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

matiques et que J. Maniez [65] appelle *relations extra-sémantiques* : elles relient des termes et des faits réels, et correspondent à certaines relations dites habituellement « associatives ».

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) et des *objets* (ou référents), 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.

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 – pour 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 langues que l'anglais.

Les documentalistes ont entrepris, de leur côté, la recherche des universaux des LI. Les facettes de Ranga-

nathan en sont un exemple. G. Bhat-tacharyya [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 d'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.

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, homonymie, etc. En plus, selon une méthode déductive, il donne une définition rigoureuse de chaque concept,

tel celui de *thésaurus* que Wersig prend comme exemple.

Le problème du sujet

Les tendances les plus récentes semblent tourner autour de cette notion que les anglo-saxons nomment *aboutness*, littéralement le « de quoi » un texte parle, c'est-à-dire son *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* au moyen des LI est une activité essentielle en documentation. 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 consiste à identifier l'*aboutness* du document en question. Ce sujet peut être appréhendé en plusieurs niveaux d'analyse (selon les besoins), dont l'un est la structure textuelle.

C. Beghtol [28] traite en détail les implications linguistiques 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 exactement le même *aboutness*.

LANGAGES D'INDEXATION ET THÉORIES MATHÉMATIQUES

L'intérêt de la modélisation mathématique est de fournir des bases solides aux LI, en définissant rigoureusement, pas à pas, les termes employés : descripteurs, classes, relations.

Le genre de mathématiques employé est très souvent la théorie des ensembles, dont la terminologie ressemble parfois à celle 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 le degré de formalisation est plus ou moins poussé.

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

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 et facile à manipuler sont observées. Par exemple, il est aisé de montrer que n facettes permettent de définir $n!$ classifications hiérarchiques différentes. De nouvelles notions sont définies : le *degré de caractère générique* d'une classe (rapport du nombre de descripteurs subordonnés au nombre total de descripteurs), le *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 existants. 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 consiste en 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 déduits. 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 [56] détaille les caractéristiques mathématiques de ces relations dans les thésaurus.

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

APERÇU CHRONOLOGIQUE	
1852, Grande-Bretagne. P.M. Roget publie le <i>Thesaurus of english words and phrases</i> .	mécanisé, précurseur des langages d'indexation en chaîne.
Vers 1900, Etats-Unis. J.O. Kaiser introduit le concept d'indexation systématique.	1962, Inde. Fondation du Documentation research and training centre à Bangalore, centre de l'« école indienne » initiée par S.R. Ranganathan.
1911, Etats-Unis. E.W. Hulme formule la notion de caution des textes.	1963, Etats-Unis. Thésaurus de l'Engineers joint council, avec indicateurs de rôles.
1915, Suisse. F. de Saussure conçoit la théorie des deux axes de référence : rapports syntagmatiques et rapports associatifs (paradigmatiques).	1964, France. J.C. Gardin élabore le modèle général SYNTOL.
1923, Grande-Bretagne. C.K. Ogden et I.A. Richards : triangle sémantique (expressions - concepts - objets).	1964, Grande-Bretagne. R.M. Needham et K. Sparck Jones : approche automatique des classifications et notion d'amas.
Années 30, Inde. S.R. Ranganathan expose la procédure en chaîne et la méthode analytico-synthétique de classification.	1967, RFA. D. Soergel : formalisation mathématique des langages d'indexation.
1948, Grande-Bretagne. Fondation du Classification research group (CRG), à Londres.	1969, Grande-Bretagne. J. Aitchison présente le <i>Thesaurifacet</i> , thésaurus disposant les concepts analysés en facettes.
1950, Grande-Bretagne. J.E.L. Farradane propose une théorie scientifique pour les classifications, point de départ de l'indexation relationnelle.	Vers 1960, Inde. G. Bhattacharyya, développe le langage d'indexation en chaîne POPSI.
Vers 1951, Etats-Unis. M. Taube conçoit un système d'indexation coordonnée au moyen d'unitermes.	1971, Grande-Bretagne. D. Austin teste un prototype de PRECIS à la BNB.
1954, Grande-Bretagne. Le biologiste J.K. Feibleman décrit la théorie des niveaux d'intégration.	1974, RFA. I. Dahlberg : recherche interdisciplinaire des fondements universels de l'ordre du savoir, et création de la revue <i>International classification</i> .
1955, Etats-Unis. Fondation du Centre for documentation and communication research de la Western reserve university (WRU) à Cleveland.	1974, RFA. R. Fugmann remet en cause les principes des thésaurus et développe un système d'indexation au moyen de graphes, TOSAR.
1957-1958, Etats-Unis. Premiers modèles mathématiques des langages d'indexation de J.W. Perry - A. Kent et C.N. Mooers.	1977-1978, Canada. T.C. Craven : langages d'indexation en chaîne NEPHIS et LIPHIS.
1959, Etats-Unis. Premier thésaurus au sens moderne du terme de la société Dupont de Nemours.	1979, Inde. G. Bhattacharyya recherche les fondements communs des langages d'indexation avec le Subject indexing language (SIL).
Fin des années 50. Apparition des notions de précision et de rappel.	1981, France. A. Dewèze : réseaux sémantiques extra-lexicaux appliqués à la constitution des thésaurus.
Années 60. Travaux des linguistes structuralistes : N. Chomsky, C.J. Fillmore, J. Lyons.	1986, Grande-Bretagne. C. Beghtol analyse la notion de sujet (<i>aboutness</i>) au moyen de la linguistique textuelle.
1960, Etats-Unis. H.P. Luhn développe le système KWIC entièrement	

d'expressions bien formées non synonymes qu'il permet de construire. Par la souplesse de leur combinatoire, les LI postcoordonnés ont une puissance sémantique supérieure aux LI précoordonnés.

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], le même auteur 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 sémiqes*. Par exemple, un niveau dans une monohiérarchie aura dans sa configuration sémiqie 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.

*
**

Cette étude montre la vitalité des recherches théoriques sur les langages d'indexation menées depuis une quarantaine d'années dans les pays anglo-saxons, en Europe de l'Est et en Inde, beaucoup plus que dans les pays francophones dont on aura remarqué la modeste contribution d'un point de vue quantitatif. L'augmentation du rythme d'accroissement des publications – fait qu'il devient banal de souligner – a certes stimulé les façons de considérer l'indexation, clé d'accès à la documentation.

Comme preuve, il est significatif de constater que, durant cette période, tous les langages d'indexation traditionnels (classifications, vedettes-matière...) ont été systématiquement remis en question, et que même les nouveaux venus (thésaurus) n'ont pas échappé aux critiques.

Mars 1990

BIBLIOGRAPHIE

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

- les principales théories et modèles, tels que ces termes ont été définis plus haut ;
- l'expression des relations entre les termes d'indexation. Comme tout langage, les langages d'indexation possèdent des liens sémantiques et syntaxiques ;
- quelques systèmes d'indexation originaux considérés à partir des réflexions théoriques qui ont été à leur base.

On remarquera que plus de la moitié des articles recensés sont concentrés dans quatre revues : *Journal of the American society for information science*, *International classification*, *Information processing and management* (qui fait suite à : *Information storage and retrieval*), et *Journal of documentation*.

THÉORIES ET MODÈLES

Généralités

- [1] **Bhattacharyya (G.)**. – Fundamentals of subject indexing languages. *In* : Ordering systems for global information networks : proceedings / of the third international study conference on classification research, 1975, Jan. 6-11, Bombay, India ; A. Neelameghan ed. – Bangalore : International federation for documentation, Committee on classification research (FID/CR) : Sadara Ranganathan endowment for library science, 1979. – p. 83-89. – (Sadara Ranganathan endowment for library sciences series, 13 ; FID publication, 533).
- [2] **Cooper (Williams S.)**. – Indexing documents by Gedanken Experimentation. – *Journal of the American society for information science*, 1978, vol. 29, n° 3, 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. – 260 p.

- [4] **Dahlberg (Ingtraut)**. – Grundlagen universaler Wissensordnung. – München : Verlag Dokumentation, 1974. – 366 p. – (DGD-Schriftenreihe, 3).
- [5] **Foskett (Douglas J.)**. – Classification for a general index language. – London : Library association, 1970. – 48 p. – (Library association research publication, 2).
- [6] **Foskett (Douglas J.)**. – Systems theory and its relevance to documentary classification. – *International classification*, 1980, vol. 7, n° 1, p. 2-5.
- [7] **Fugmann (Robert)**. – The glamour and the misery of the thesaurus approach : treatise VI in information retrieval theory. – *International classification*, 1974, vol. 1, n° 2, p. 76-86.
- [8] **Fugmann (Robert)**. – Toward a theory of information supply and indexing : treatise VI on retrieval system theory. – *International classification*, 1979, vol. 6, n° 1, p. 3-15.
- [9] **Fugmann (Robert)**. – On the practical of indexing and its theoretical foundations. – *International classification*, 1980, vol. 7, n° 1, p. 13-20.
- [10] **Fugmann (Robert)**. – The five-axiom theory of indexing and information supply. – *Journal of the American society for information science*, 1985, vol. 36, n° 2, p. 116-129.
- [11] **Gardin (Jean-Claude)**. – Éléments d'un modèle pour la description des lexiques documentaires. – *Bulletin des bibliothèques de France*, 1966, tome 11, n° 5, 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 (Kevin P.)**. – Towards a theory of indexing. – *Journal of documentation*, 1976, vol. 32, n° 2, 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. *In* : Proceedings, vol. 5 / American society for information science, 31st annual meeting, Oct. 20-24, 1967, Columbus, 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*, 1970, vol. 21, n° 5, p. 358-367.
- [17] **Maniez (Jacques)**. – Etude générale des langages documentaires. *In* : Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires. – Paris : Ed. d'organisation, 1987. – 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. *In* : Encyclopedia of library and information science, vol. 11. – New York : Dekker, 1974. – p. 286-299.
- [20] **Soergel (Dagobert)**. – Some remarks on information languages, their analysis and comparison. – *Information storage and retrieval*, 1967, vol. 3, n° 4, p. 219-291.
- [21] **Soergel (Dagobert)**. – Theoretical problems of thesaurus building with particular reference to concept formation. *In* : Fachsprache, Umgangssprache / J.S. Petöfi, A. Podlech, E.V. Savigny eds. – Kronberg/Taunus : Scriptor-Verlag, 1975. – p. 355-381.
- [22] **Svenonius (Elaine F.)**. – Directions for research in indexing, classification and cataloging. – *Library resources and technical services*, 1981, vol. 25, n° 1, p. 88-103.
- [23] **Taube (Mortimer) ; Gull (C.D.) ; Wachtel (Irma S.)**. – Unit terms in coordinate indexing. – *American documentation*, 1952, vol. 3, n° 4, p. 213-218.
- [24] **Vickery (B.C.)**. – Retrieval language models. *In* : Information systems. – London : Butterworth, 1973. – p. 203-222.
- [25] **Wersig (Gernot)**. – Eine neue Definition von "Thesaurus". – *Nachrichten für Dokumentation*, 1969, Vol. 20, Nr. 2, p. 53-61.
- [26] **Ballmer (Thomas T.)**. – Zur Gewinnung einer Fundamentalklassifikation des menschlichen Wis-

- sens. II. - *International classification*, 1983, vol. 10, n° 3, p. 127-134.
- [27] **Ballmer (Thomas T.)**. - Zur Gewinnung einer Fundamentalklassifikation des menschlichen Wissens. IV. - *International classification*, 1984, vol. 11, n° 1, 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*, 1986, vol. 42, n° 2, p. 84-113.
- [29] **Bivins (Kathleen T.)**. - An approach to an atomic theory of indexing. In : *New trends in documentation and information : proceedings / of the 39th FID congress*, University of Edinburgh, 25-28 Sept. 1978 ; Peter J. Taylor ed. - London : ASLIB, 1980. - p. 50-55. - (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*, 1982, vol. 9, n° 3, p. 140-144.
- [32a] **Gardin (Jean-Claude)**. - Document analysis and linguistic theory. - *Journal of documentation*, 1973, vol. 29, n° 2, p. 137-168.
Traduit dans :
- [32b] **Gardin (Jean-Claude)**. - Analyse documentaire et théorie linguistique. In : *Les analyses de discours*. - Neuchâtel : Delachaux et Niestlé, 1974. - p. 120-168.
- [33] **Henriksen (Tor)**. - A linguistic description model for indexing languages. In : *Ordering systems for global information networks : proceedings / of the third international study conference on classification research*, 1975, Jan. 6-11, Bombay, India ; A. Neelameghan ed. - Bangalore : International federation for documentation, Committee on classification research (FID/CR) : Sadara Ranganathan endowment for library science, 1979. - p. 3-8. - (Sadara Ranganathan endowment for library science series, 13 ; FID publication, 533).
- [34] **Hutchins (W.J.)**. - Languages of indexing and classification : a linguistic study of structures and functions. - Stevenage (Herts.) : Pergamon, 1975. - 148 p. - (Librarianship and information studies, 3).
- [35] **Jones (Kevin P.)**. - The environment of classification. Part II. How we classify. - *Journal of the American society for information science*, 1974, vol. 25, n° 1, p. 44-51.
- [36] **Kim (Chai)**. - Theoretical foundations of thesaurus-updating. - *Journal of the American society for information science*, 1973, vol. 24, n° 2, p. 148-156.
- [37] **Long (B.)**. - Linguistique et indexation. - *Documentaliste - Sciences de l'information*, 1980, vol. 17, n° 3, 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] *Classification research : Proceedings / of the second international study conference*, Elsinore, Denmark, Sept. 14-18, 1964 ; Pauline Atherton ed. - 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*, 1983, vol. 5, n° 4, p. 227-234.
- [40] **Schreider (Yu A.)**. - The thesaurus in information science and theoretical semantics. - *Automatic documentation and mathematical linguistics*, 1971, vol. 5, n° 1, 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 retrieval. I. The concepts of relevance and information language. - *Information processing and management*, 1977, vol. 13, n° 4, 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*, 1976, vol. 3, n° 2, p. 75-81.
- [47] **Heilprin (Laurence B.)**. - Mathematical model of indexing. - Washington : Documentation incorporation, 1957. - 26 p.
- [48] **Moovers (Calvin N.)**. - A mathematical theory of language symbols in retrieval. In : *Proceedings of the international conference on scientific information*. - Washington, 1958. - p. 57-94.
- [49] **Perry (James W.) ; Kent (Allen)**. - A mathematical model system. In : *Documentation and information retrieval : an introduction to basic principles and cost analysis*. - Cleveland (Ohio) : The Press of Western reserve university and interscience publishers, 1957. - 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*, 1978, vol. 3, n° 3, p. 14-17.
- [51] **Richmond (Phyllis A.)**. - Classification in syndetic structures. In : *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 ; Ingetraut Dahlberg ed. - Frankfurt : Indeks Verlag, 1982. - p. 149-156.
- [52] **Schauble (P.)**. - Thesaurus based concept spaces. In : *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*, 1973, vol. 7, n° 2, p. 30-36.
- [54] **Schreider (Yu A.)**. - The duality of classification : taxonomy and meronymy. - *International forum on information and documentation*, 1981, vol. 6, n° 1, 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*, 1967, vol. 3, p. 129-173.
- [56] **Sridhar (M.S.)**. - A mathematical approach to relations in thesauri. - *Journal of library and information science (India)*, 1980, vol. 5, n° 1, p. 76-87.
- [57] **Sydorchenko (V.D.)**. - On the construction of a descriptor model. - *Annals of library science and documentation*, 1974, vol. 21, n° 4, p. 142-145.
- [58] **Taube (Mortimer) ; Wachtel (Irma S.)**. - The logical structure of coordinate indexing. - *American documentation*, 1953, vol. 4, p. 67-68.
- [59] **Turski (W.M.)**. - On a model of information retrieval based on thesaurus. - *Information storage and retrieval*, 1971, vol. 7, n° 2, p. 89-94.

Théories et modèles mathématiques

- [44] **Barnes (R.F.)**. - Mathematical 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 : *Ordering systems for global information networks : proceedings / of the third international study conference on classification research*, 1975, Jan. 6-11, Bombay, India ; A. Neelameghan ed. - Bangalore : International federation for documentation, Committee on classification research (FID/CR) : Sadara Ranganathan endowment for library science, 1979. - p. 67-82. - (Sadara Ranganathan endowment for library science series, 13 ; FID publication, 533).
- [46] **Diodato (Virgil)**. - Modern algebra and information science. - *Journal for the American society for information science*, 1983, vol. 34, n° 4, p. 257-261.

RELATIONS ENTRE LES TERMES D'INDEXATION

Généralités

- [60] **Austin (Derek)**. – Basic concept classes and primitive relations. *In* : 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 ; Ingetraut Dahlberg ed. – 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. 50-64.
Repris dans :
- [61b] Theory of subject analysis : a source book / Lois Mai Chan, Phyllis A. Richmond, Elaine Svenonius eds. – 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*, 1973, vol. 29, n° 4, p. 390-404.
- [63] **Gull (C.D.)**. – Structure of indexing authority lists. – *Library resources and technical services*, 1966, vol. 10, n° 4, p. 507-511.
- [64] **Kervégant (Désiré)**. – Classification et analyse des relations. – *Bulletin des bibliothèques de France*, 1959, vol. 4, n° 11, p. 495-511.
- [65] **Maniez (Jacques)**. – Relationships in thesauri : some critical remarks. – *International classification*, 1988, vol. 15, n° 3, p. 133-138.
- [66] **Needham (R.M.) ; Sparck Jones (Karen)**. – Keywords and clumps. – *Journal of documentation*, 1964, vol. 20, n° 1, p. 5-15.
- [67] **Willets (M.)**. – An investigation of the nature of the relation between terms in thesauri. – *Journal of documentation*, 1975, vol. 31, n° 3, p. 158-184.

Relations sémantiques

- [68] **Beghtol (Clare)**. – Semantic validity : concepts of warrant in bibliographic classification systems. – *Library resources and technical services*, 1986, vol. 30, n° 2, 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*, 1983, vol. 4, n° 1, p. 11-27.
- [71] **Ghose (A.) ; Dhawle (A.S.)**. – Problems of thesaurus construction. – *Journal of the American society for*

information science, 1977, vol. 28, n° 4, p. 211-217.

- [72] **Jansen (R.)**. – Die Bedeutung der Zugehörigkeitsrelation für Thesaurusstruktur, Indexierungsmethode und Recherche. *In* : Deutscher Dokumentartag, 1974. Band 1, Bonn, 7-11 Okt. 1974 / Deutsche Gesellschaft für Dokumentation. – München : Verlag Dokumentation, 1975. – p. 354-374.
- [73] **Popovska (H.)**. – Synonymic and hierarchical relations in descriptor languages. – *Automatic documentation and mathematical linguistics*, 1985, vol. 19, n° 1, p. 74-87.
- [74] **Soergel (Dagobert)**. – Index language structure. I : conceptual. *In* : Organizing information : principles of data base and retrieval systems. – Orlando (Fl.) : Academic press, 1985. – p. 251-287.
- [75] **Syrdorchenko (V.D.)**. – Typological classification of thesaurus semantic structures. – *Automatic documentation and mathematical linguistics*, 1975, vol. 9, n° 2, 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*, 1977, vol. 28, n° 6, p. 345-354.

Relations syntaxiques

- [77] **Aitchinson (J.)**. – Thesaurifacet : a new concept in subject retrieval schemes. *In* : Subjects retrieval in the seventies : new directions : proceedings / of an international symposium, University of Maryland, 1971 ; H. Wellisch, T.D. Wilson eds. – 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*, 1964, vol. 4, 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. Tome I, Aspects linguistiques. Tome II, Etude critique de quelques SRD. – Dijon : IUT de Dijon, Département carrières de l'information, 1976. – VI-184 + 182-XXX p.

LANGAGES ET SYSTÈMES D'INDEXATION

Indexation en chaîne

- [81] **Armitage (Janet E.) ; Lynch (Michael F.)**. – Some structural characteristics of articulated subject indexes. – *Information storage and retrieval*, 1968, vol. 4, n° 2, p. 101-111.
- [82] **Batty (C.D.)**. – Chain indexing. *In* :

Encyclopedia of library and information science, vol. 4. – New York : Dekker, 1970. – p. 423-434.

- [83] **Bhattacharyya (G.)**. – Chain procedure and structuring of a subject. – *Library science with a slant to documentation*, 1972, vol. 9, n° 4, p. 585-635.
- [84] **Craven (Timothy C.)**. – String indexing. – Orlando (Fl.) : Academic press, 1986. – 246 p.
- [85] **Farradane (Jason E.L.)**. – A comparison of some computerproduced permuted alphabetical subject indexes. – *International classification*, 1977, vol. 4, n° 2, p. 94-101.
- [86] **Mineur (B.W.)**. – Relations in chains. – *Journal of librarianship*, 1973, vol. 5, n° 3, p. 175-202.
- [87] **Rajan (T.N.)**. – From chain to string indexing. – *Library herald*, 1981-1982, vol. 20, n° 2-4, 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.

PRECIS et POPSI

- [89] **Austin (Derek)**. – An information retrieval language for MARC. – *Aslib proceedings*, 1970, vol. 22, n° 10, p. 481-491.
- [90a] **Austin (Derek) ; Digger (Jeremy A.)**. – PRECIS : the preserved context index system. – *Library resources and technical services*, 1977, vol. 21, n° 1, p. 13-30.
Repris dans :
- [90b] Theory of subject analysis : a source book / Lois Mai Chan, Phyllis A. Richmond, Elaine Svenonius eds. – 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*, 1979, vol. 16, n° 1, p. 1-34.
- [92] **Bhattacharyya (G.)**. – Some significant results of current classification research in India. – *International forum on information and documentation*, 1981, vol. 6, n° 1, p. 11-18.
- [93] **Biswas (Subal C.)**. – Two decades of POPSI, 1969-1988 : a literature review. – *Libri*, 1988, vol. 38, n° 4, p. 279-296.
- [94] **Cheti (A.)**. – Sistemi d'indicizzazione preordinati : indicizzazione a catena, PRECIS, POPSI, NEPHIS. – *Bollettino d'informazioni / Associazione italiana biblioteche*, 1984, vol. 24, n° 1, 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*, 1983, vol. 8, n° 1, 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*, 1983, vol. 33, n° 4, p. 316-330.
- [97] **Sorensen (Jutta), Austin (Derek)**. – PRECIS in a multilingual context. Part 2. A linguistic and logical explanation of the syntax. – *Libri*, 1976, vol. 26, n° 2, p. 108-139.
- Autres systèmes**
- [98] **Craven (Timothy C.)**. – NEPHIS : a nested phrase indexing system. – *Journal of the American society for information science*, 1977, vol. 28, n° 2, p. 107-114.
- [99] **Craven (Timothy C.)**. – Linked phrase indexing. – *Information processing and management*, 1978, vol. 14, n° 6, p. 469-476.
- [100] **Farradane (Jason E.L.)**. – A scientific theory of classification and indexing and its practical applications. – *Journal of documentation*, 1950, vol. 6, n° 2, p. 83-99.
- [101] **Farradane (Jason E.L.)**. – A scientific theory of classification and indexing : further considerations. – *Journal of documentation*, 1952, vol. 8, n° 2, 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*, 1963, vol. 1, n° 1, p. 3-11.
- [103] **Fugman (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*, 1974, vol. 25, n° 5, p. 287-307.
- [104a] **Kaiser (Julius Otto)**. – Systematic indexing. In : Report of proceedings / ASLIB, 3rd conference, Oxford, Sept. 24-27, 1926. – London : ASLIB, 1927. – p. 20-33.
Repris dans :
- [104b] Readings in library cataloging / R.K. Olding ed. – London : Crosby Lockwood, 1966. – p. 145-162.
- [105] **Nickelsen (Herbert) ; Nickelsen (Ingeborg)**. – Mathematische Analyse des TOSAR-Verfahrens. – *Information storage and retrieval*, 1973, vol. 9, n° 2, p. 95-119.
- [106] **Svenonius (Elaine)**. – Facet definition : a case study. – *International classification*, 1978, vol. 5, n° 3, p. 134-141.

CYCLES DE FORMATION CONTINUE ADBS

Janvier - février 1991

JANVIER 9-11 Orléans	601C – Savoir utiliser les banques de données <i>Animateur</i> : Claude POLLET <i>Prix</i> : 4.000 F T.T.C.
17-18 Chalons/Marne	911C – Concevoir une revue de presse pour la vendre (1^{re} partie) <i>Animatrice</i> : Edwige DARGEGEN <i>Prix</i> : 3.400 F T.T.C.
21-23 Paris	302 – Informatisation d'une unité documentaire (1^{re} partie) <i>Animatrices</i> : Odile GIRAUD, Elisabeth SIMARD <i>Prix</i> : 8.600 F T.T.C.
28-1.02 Paris	33 – Management d'un centre de ressources documentaires <i>Animatrices</i> : Catherine SERENI, Martine SIBERTIN-BLANC <i>Prix</i> : 6.200 F T.T.C.
29-31 Lille	922N – Dialogue documentaliste/utilisateur. Les situations de communication <i>Animateur</i> : Jean-Paul DONCKELE <i>Prix</i> : 3.950 F T.T.C.
FEVRIER 4-5 Paris	901 – Sensibilisation à l'analyse de la valeur <i>Animateur</i> : Jean MICHEL <i>Prix</i> : 2.950 F TTC
5-8 Paris	406 – Informatisation d'une photothèque <i>Prix</i> : 5.300 F TTC
6-8 Paris	920 – Veille informative - De la définition à l'organisation du besoin <i>Animatrice</i> : Evelyne LAUTRE <i>Prix</i> : 4.250 F TTC
11-13 Paris	302 – Informatisation d'une unité documentaire (2^e partie)
25/2-24/6	007 – English for librarians and information officers <i>Animateur</i> : Paul GOULD <i>Prix</i> : 2.600 F TTC
Session supplémentaire :	
MARS 11-12	604 – Savoir utiliser la documentation européenne - Paris <i>Animatrices</i> : Hélène GAUTIER, Anny MAXIMIN <i>Prix</i> : 2.600 F TTC

Programme, renseignements et inscriptions :

ADBS Formation, 25 rue Claude Tillier, 75012 PARIS. Tél. (1) 43.72.25.25