

PRECIS et le système de vedettes-matière de la Library of Congress : vers une étude comparative globale

Jean-Pierre Côté

Volume 25, numéro 1, mars 1979

URI : <https://id.erudit.org/iderudit/1054367ar>

DOI : <https://doi.org/10.7202/1054367ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

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

ISSN

0315-2340 (imprimé)

2291-8949 (numérique)

[Découvrir la revue](#)

Citer cet article

Côté, J.-P. (1979). PRECIS et le système de vedettes-matière de la Library of Congress : vers une étude comparative globale. *Documentation et bibliothèques*, 25(1), 11–21. <https://doi.org/10.7202/1054367ar>

Résumé de l'article

Après avoir rappelé les étapes constitutives de l'analyse documentaire et les problèmes qui y sont rattachés, l'auteur expose ce qui caractérise PRECIS, en superposant ses principaux éléments sémantiques et syntaxiques à ceux du système de vedettes-matière de la Library of Congress. Du même coup est projetée une image plus précise de ce dernier, avec ses limites, mais aussi avec ses capacités évolutives. Cette confrontation doit assurer, grâce à l'éclairage mutuel qu'elle fournit, une plus juste comparaison.

PRECIS et le système de vedettes-matière de la Library of Congress: vers une étude comparative globale

Jean-Pierre Côté

Bibliothèque nationale du Québec
Montréal

Après avoir rappelé les étapes constitutives de l'analyse documentaire et les problèmes qui y sont rattachés, l'auteur expose ce qui caractérise PRECIS, en superposant ses principaux éléments sémantiques et syntaxiques à ceux du système de vedettes-matière de la Library of Congress. Du même coup est projetée une image plus précise de ce dernier, avec ses limites, mais aussi avec ses capacités évolutives. Cette confrontation doit assurer, grâce à l'éclairage mutuel qu'elle fournit, une plus juste comparaison.

After having recalled the different steps which constitute documentary analysis and its problems, the author explains the main characteristics of PRECIS, through a comparison of its main semantic and syntactic elements with those of the Library of Congress system of subject headings. This gives a very precise image of the limits of the latter, as well as of its possibilities of adaptation and evolution. This confrontation, because of the light it sheds on these two systems, insures a more accurate comparison.

Después de recordar las etapas constitutivas del análisis de documentos y sus problemas, el autor presenta las características de PRECIS, superponiendo sus elementos semánticos y sintácticos principales con los del sistema de encabezamientos de materias de la Library of Congress. Al mismo tiempo, el autor hace una descripción más precisa de este sistema, con sus limitaciones, pero también con sus capacidades evolutivas. Gracias a la perspectiva más clara que da sobre ambos sistemas, el artículo permite y asegura una comparación mas justa.

Dans toutes les disciplines, il y a des oeuvres qui marquent par la synthèse qu'elles proposent aux praticiens de leur époque. A chaque fois, comme dans toute synthèse, c'est le résultat de l'intégration d'éléments qui existaient déjà mais qui, en vertu de cette intégration (fondée sur une différenciation préalable) décuplent leurs possibilités d'application. Ce fut le cas en 1876 lorsque Charles Ami Cutter publia ses *Rules for a Dictionary Catalog*¹ et, près d'un siècle plus tard, lorsque Derek Austin publia en 1974 son *PRECIS: A Manual of Concept Analysis and Subject Indexing*².

Le but de cet article n'est pas de faire une comparaison systématique des deux systèmes qui en sont nés, en confrontant chacun de leurs éléments afin d'aboutir à une évaluation globale de l'efficacité avec laquelle, quoique de manière et à des degrés différents, ils solutionnent des problèmes communs liés à l'analyse et à la recherche documentaire. Cet article se propose plus modestement de situer le mieux possible le débat qui prévaut actuellement chez les spécialistes du domaine. Cet objectif suppose une connaissance égale des fondements respectifs des deux systèmes, connaissance qui doit par ailleurs être exprimée en des termes susceptibles de favoriser la comparaison. Dans ce but, l'étude préalable des thesaurus servira à présenter les aspects sémantiques particuliers à PRECIS là où l'étude des vedettes-matière de la Library

1. Charles A. Cutter, *Rules for a Dictionary Catalog*, Washington, Government Printing Office, 1891, 140 p.

2. Derek Austin, *PRECIS: A Manual of Concept Analysis and Subject Indexing*, London, Council of the British National Bibliography, 1974, x, 551 p.

of Congress contribuera à exposer l'essentiel de ses aspects syntaxiques.

Nous remonterons donc à la source et nous nous interrogerons sur le pourquoi de l'analyse documentaire. Cet éclairage nous aidera à saisir le comment du fonctionnement général des langages documentaires et les problèmes qui y sont rattachés, problèmes de nature linguistique et documentaire. A regarder les résultats d'indexation par PRECIS, certains se sont demandé s'il n'y avait pas là quelque exagération injustifiable par rapport aux besoins réels d'information. C'est dans ce contexte que nous tenterons de montrer la nouveauté de PRECIS: au lieu d'en démontrer toute la mécanique, nous essaierons plutôt de trouver sa place parmi les autres types de langages documentaires, de le classer.

L'automatisation des fonctions des bibliothèques (ici pris dans un sens générique) et particulièrement de la recherche documentaire alliée à des besoins d'information conditionnés par les orientations contemporaines de la recherche (qui est de plus en plus inter-, trans- et pluri-disciplinaire) sont des facteurs plus récents qui soulignent avec force le problème des critères de choix d'un langage documentaire. Ce problème est perçu peut-être avec une acuité nouvelle depuis l'existence de PRECIS qui a justement été conçu en vue d'exploiter les possibilités de l'ordinateur. Le seul fait d'exister est déjà un de ses apports les plus concrets. Il est depuis sa naissance un concurrent virtuel des vedettes-matière de la Library of Congress: tous deux sont des langages documentaires à vocation encyclopédique, dont l'indexation est précoordonnée à partir d'un vocabulaire contrôlé. C'est un peu parce qu'il est situé sur ce même plan mais offrant des possibilités nouvelles que PRECIS a suscité une remise en question des vedettes-matière. Cette constatation suffit à justifier notre propos.

Résumons notre projet: plutôt que d'expliquer le fonctionnement de PRECIS en utilisant le cadre pédagogique fourni par le système lui-même, nous emprunterons un long détour pour nous permettre d'identifier à la limite ce qui le caractérise.

De ces observations, nous serons en mesure de déduire quelques conditions essentielles à respecter dans un projet d'évaluation de l'efficacité respective des deux systèmes.

L'analyse documentaire

La première question qu'on doit se poser en toute logique porte sur la nature du document: qu'est-il ce document qui sert à définir la plupart de nos concepts et de nos activités? Indiscutablement, il participe de deux natures différentes intrinsèquement liées, l'une physique, l'autre mentale. La première créera des problèmes d'organisation spatiale, la deuxième d'organisation intellectuelle. Il s'agira alors, dans ce dernier cas, de processus se produisant dans le temps (croissance, évolution, transfert, etc. des connaissances). Nous ajoutons que c'est la finalité du document qui doit primer, donc sa nature mentale. Il n'y aurait en effet pas de documents (imprimés, sonores, visuels et toutes les combinaisons possibles) s'il n'y avait eu d'abord et avant tout le désir d'un auteur (écrivain, compositeur, artiste et toutes les combinaisons possibles) de communiquer ses perceptions et ses conceptions au-delà des limites que la nature lui impose. Ces connaissances matérialisées par et dans le document ne sont par conséquent plus sous le contrôle direct de leur auteur, à tel point qu'elles lui survivent et continuent d'être communicables après sa mort.

Ce document, une fois publié, se retrouvera possiblement dans une ou plusieurs collections. Dans chacune il devra être intégré de manière telle que sera facilitée sa communication à tout usager qui en aura besoin, soit pour connaître toutes les oeuvres de cet auteur, rendre intime et approfondir sa pensée, soit pour consulter les documents que possède la bibliothèque sur tel sujet. La fonction ultime de l'indexation sera de créer des classes de documents (nous rappelons sa nature mentale) partant de la possession de caractères communs, par exemple le même auteur, la même collection, le même sujet.

L'indexation par sujets rattachera ce document à une classe (qui existe déjà ou

qu'on créera pour la circonstance) dans laquelle on ne retrouvera que les documents qui traitent ce même sujet. La justification est économique: il est plus facile et plus rapide de retrouver une information ou un document pertinent (dont le chercheur établira, évaluera la pertinence) dans ce sous-ensemble de la collection désigné par son appartenance à telle classe que dans l'ensemble de la collection. Si l'on peut dire qu'un document existe par le désir de communication de son auteur, l'analyse documentaire, elle, existe pour satisfaire le désir de communication du chercheur. Ce désir du chercheur se manifeste par l'expression d'un besoin d'informations qu'il ne peut maintenant plus satisfaire lui-même devant le trop grand nombre de documents.

L'analyse documentaire comprend tout un ensemble d'opérations, chacune traînant son bagage de problèmes: interprétation du sujet d'un document, son message essentiel; analyse conceptuelle de l'énoncé du sujet; création de classes à partir des résultats de cette analyse; attribution d'un nom, d'une étiquette. Nous ne les aborderons pas toutes également en profondeur, mais seulement celles plus utiles à notre propos.

A l'étape de l'interprétation du sujet, les principaux problèmes rencontrés sont ceux de la profondeur et de l'homogénéité de l'analyse (nous faisons abstraction de l'exactitude de l'analyse, ce problème ayant la même teneur dans tous les systèmes, étant davantage lié aux individus qu'aux systèmes) dont le niveau retenu déterminera le contenu des classes. Ainsi, plus les documents sur un sujet appartiendront à de grandes classes, plus il sera facile de les retrouver tous dans ces sous-ensembles, mais en contrepartie plus difficile de retrouver seulement ceux qui seront pertinents, la discrimination n'étant pas assez forte. A l'inverse, plus l'on distribue les documents dans un grand nombre de petits sous-ensembles, plus il sera facile d'obtenir uniquement des documents pertinents, mais partant plus difficile de les retrouver tous dans l'optique d'une recherche exhaustive (il faudra alors jeter des ponts entre les sous-ensembles qui auront des contenus sémantiquement

proches). On aura reconnu les deux mesures traditionnelles d'évaluation de l'efficacité des systèmes de communication documentaire: le rappel et la précision.

On peut déjà constater que la qualité de l'analyse est un facteur déterminant le degré de réussite de toute recherche documentaire, car en admettant que la stratégie de recherche ait permis de retenir les classes pertinentes à considérer, seul ce qu'on y trouvera en dernier ressort permettra d'évaluer la réussite de la recherche. Ce qu'il faut juger ce sont les classes qui sont formées à partir des résultats de l'analyse du sujet: leur contenu, mais aussi leur nom.

Les aspects sémantiques

Dans un système verbal d'indexation par sujets, les classes sont désignées par des termes tirés d'une langue naturelle. Trois choix se présentent alors comme trois niveaux de complexité croissante sur un même continuum: l'indexation du sujet énoncé sous la forme d'une phrase peut être faite à partir des mots ou unitermes; à partir des concepts ou descripteurs; à partir des sujets ou vedettes-matière. Peu importe l'élément de base retenu, l'ensemble des unitermes, des descripteurs ou des vedettes-matière qui serviront à l'indexation constituent un langage documentaire.

Etant donné le contexte encyclopédique dans lequel nous nous sommes situés, les unitermes présentent peu d'attrait. Qu'il nous suffise de rappeler qu'ils sont nés en réaction au contrôle excessif exercé à cette époque par les vedettes-matière et qu'ils ont rapidement évolué vers une certaine précoordination de plusieurs mots à mesure que des liaisons douteuses résultaient de la postcoordination. Dans ce contexte, le descripteur est apparu comme la solution intermédiaire.

Les descripteurs représentent les concepts avec des termes composés d'un ou de plusieurs mots précoordonnés. Quand ces descripteurs s'intègrent dans un thesaurus, on parle d'un langage contrôlé. Il y a alors contrôle formel de la représentation des concepts pour assurer avec consistance la désignation des clas-

ses de documents. Mais devant l'évolution rapide de la terminologie de différentes disciplines, il est tentant de renoncer à tout contrôle et de préconiser le choix des termes mêmes utilisés par les auteurs. Après un certain temps, on pourrait effectuer un contrôle a posteriori. Le thesaurus, afin d'éviter la dispersion des sujets qu'une telle démarche occasionne, effectue ce contrôle a priori et choisit un terme qui a pratiquement valeur de symbole. La forme du terme choisi est déterminée avec le plus grand discernement et doit tenir compte de l'usage le plus répandu chez les spécialistes de la discipline; les mêmes exigences s'appliquent par ricochet au choix des termes équivalents. Ces derniers serviront à guider la recherche vers le descripteur retenu. De sorte qu'on peut être assuré de trouver dans la classe ainsi formée tous les documents qui traitent le même sujet peu importe les différents noms que celui-ci peut porter.

La principale difficulté dans l'indexation à partir des concepts, c'est la notion même de concept, difficile à cerner, comme la notion de liberté. Où se termine un concept? Où débute-t-il? Une fois qu'on a répondu, le principal avantage réside dans la possibilité d'organiser sémantiquement les termes qui les représentent en vertu de la signification propre qu'on peut leur donner.

Le premier objectif d'un thesaurus, nous le répétons, c'est d'assurer avec consistance la représentation verbale de concepts, ces derniers étant eux-mêmes, dans un système de communication documentaire, les représentations mentales de classes de documents. Un thesaurus est un outil de communication. Il doit pouvoir aider tout usager de système dont les besoins d'informations ne peuvent être satisfaits que par une recherche exhaustive. Il exposera pour ce faire les relations hiérarchiques et associatives qu'entretiennent entre eux les termes du thesaurus, pour orienter l'usager vers d'autres classes de documents dont le contenu sémantique est étroitement associé. Ce sont là des relations paradigmatiques, selon la terminologie de Gardin, en ce sens qu'elles sont permanentes, peu importe le contexte dans lequel ces termes sont utilisés, par

opposition aux relations syntagmatiques qui ne sont que circonstancielles, liées au contexte, spécifiques aux documents.

Le thesaurus de PRECIS

PRECIS, sur le plan sémantique, emploie un thesaurus dont les principes d'élaboration sont ceux contenus dans la norme ISO 2788 de l'Organisation internationale de normalisation, *Guidelines for the Establishment and Development of Monolingual Thesauri*. Première observation: PRECIS a un vocabulaire contrôlé, c'est-à-dire que chaque concept est représenté par un seul terme «préféré», comme dans tout thesaurus. Cependant, ce vocabulaire est ouvert et permet d'accueillir dans le système un nouveau terme rencontré dans un document dès qu'il représente un concept différent. Il est introduit dans la structure par l'établissement de relations hiérarchiques, de bas en haut, du plus simple au plus complexe, et seulement dans ce sens.

Cette norme reconnaît les trois types de relations sémantiques fondamentales que nous avons vus plus tôt: les relations d'équivalence, les relations hiérarchiques et les relations associatives. Cette classification, on le reconnaît, n'est pas nouvelle. Elle a fait partie depuis longtemps de tous les langages documentaires structurés. Si dans ceux-ci, l'établissement des deux premiers types (d'équivalence, hiérarchique) ne faisait généralement pas de difficultés, il n'en était pas de même pour le troisième type. Dans PRECIS, une règle permet d'en faire un type plus rigoureux de relations syntagmatiques: on établira une relation associative entre deux termes s'ils sont proches sémantiquement, s'ils ne sont liés ni par une relation d'équivalence, ni par une relation hiérarchique et si un des deux sert à définir ou à expliquer l'autre.

Une question plus cruciale encore doit logiquement être la première à recevoir une réponse puisqu'il s'agit de la représentation mentale du matériau constitutif des langages documentaires; cette question porte sur la notion de concept. Nous l'avons déjà signalée plus haut. PRECIS propose une solution pratique à ce pro-

blème d'identification. Rappelons qu'un terme y est la représentation verbale d'un concept. Aussi, tout terme capable de jouer une fonction syntaxique telle que définie par la grammaire de PRECIS sera un concept. Nous le verrons plus tard, la fonction de chaque terme dans l'expression du sujet indexé est identifiée par un indicateur. Il n'y a donc pas seulement analyse conceptuelle de l'énoncé du sujet, mais aussi analyse fonctionnelle. L'ensemble de ces indicateurs constitue la «grammaire» de PRECIS. Nous abordons déjà les relations syntagmatiques. Ces dernières surgissent en effet dès qu'il s'agit d'indexation à partir des sujets.

Les aspects syntaxiques

Nous avons exposé la nécessité de l'analyse et l'importance d'un langage documentaire. Nous avons reconnu les avantages d'un thesaurus et distingué celui de PRECIS. Le tracé que nous suivons ne nous a permis d'effleurer que les aspects sémantiques des langages documentaires. L'indexation à partir des sujets se particularise par l'établissement ou plutôt le maintien des relations fonctionnelles entre les termes de l'énoncé du sujet.

Les descripteurs peuvent être pré ou postcoordonnés selon que le système possède ou non une syntaxe, c'est-à-dire selon qu'il emploie ou non un ensemble de règles qui président à l'ordre de citation des mots et donc à la construction de vedettes-matière. Cependant, dans un cas comme dans l'autre, ils peuvent être combinés à volonté pour exprimer un sujet, peu importe son niveau de complexité. Ce n'est pas tout à fait le cas, comme nous le verrons, pour les vedettes-matière de la Library of Congress. Les différents mots qui composent ces dernières sont totalement précoordonnés mais d'une seule manière admissible. On y contrôle donc et le vocabulaire et l'expression du sujet. C'est malgré tout un énorme avantage pour la recherche documentaire. Créer et nommer des classes à partir des sujets rend encore plus spécifique le contenu de celles-ci. Cependant, ceux qui ont l'habitude de faire des recherches documentaires à partir d'une liste de vedettes-matière reconnaîtront

qu'une des principales difficultés est de prévoir comment sera entrée dans la liste la forme acceptée d'une vedette composée de plusieurs mots: vedette d'entrée directe, vedette inversée ou avec subdivision. C'est un problème de nature syntaxique.

L'entrée directe des vedettes-matière a été la règle fondamentale qui a accompagné la création de ces dernières. Elle fut énoncée par Cutter dans ses *Rules for a Dictionary Catalog*. Dans un catalogue alphabétique, elle permettait un accès plus rapide car elle évitait l'étape de consultation de l'index alphabétique d'un catalogue systématique. Il était maintenant possible de regrouper sous un seul terme les différents aspects d'un sujet, ce qui était impossible dans un système de classification basé sur les grandes disciplines. C'était là la justification d'une telle innovation. Pourtant, on a rapidement constaté que cet arrangement alphabétique faisait perdre un avantage évident: la collocation des sujets étudiés sous un même aspect permise par la structure logique des systèmes de classification. Cette dispersion était partiellement corrigée par le réseau de renvois qui accompagnait chaque vedette introduite dans la liste, réseau qui pouvait exprimer à la manière des thesaurus des relations d'équivalence, hiérarchiques et associatives. Mais dans le même but, on a introduit des formes de vedettes qui redonneraient, espérait-on, une partie des avantages perdus par l'abandon du catalogue systématique.

On le comprend plus clairement, il y a deux choix à faire dans l'établissement de vedettes-matière: choix des termes autorisés et choix de l'ordre de citation de leurs éléments s'ils sont composés de plusieurs mots. Il s'agissait à la fois d'un problème terminologique et d'un problème de nature syntaxique. C'est ainsi qu'on a recouru aux inversions de vedettes et à l'emploi de sous-vedettes. Les vedettes-matière qui en résultaient ne ressemblaient pas vraiment aux rubriques habituelles d'un catalogue systématique, car elles exprimaient rarement des relations de genre à espèce ou de tout à partie, mais bien plus souvent des relations fonctionnelles, des relations syntagmatiques, quand elles ne représentaient pas qu'un seul concept dont le terme

aurait dû être entré de manière constante selon la règle générale, soit directement. Encore eût-il fallu définir la notion de concept dans le contexte des vedettes-matière. En fait, on n'a pas formulé suffisamment de règles précises prévoyant l'introduction de telles vedettes et s'il y en avait quelques-unes, elles ne furent pas appliquées uniformément. Le résultat en fut une imprévisibilité dans la forme que prend une vedette-matière lors de son intégration dans la liste.

Une constatation s'est imposée depuis longtemps: en dehors de tout besoin terminologique nouveau, la complexité du sujet d'un document réside souvent dans les relations particulières que l'auteur a placées entre plusieurs concepts dont les termes qui les représentent sont déjà disponibles dans le système. Prenons pour exemple un rapport sur les possibilités d'utilisation des salles de classe des écoles secondaires d'une région déterminée comme bureaux de scrutin. L'«originalité» du sujet réside dans la relation particulière qui est établie par l'auteur, qui n'existe pas a priori, entre les salles de classe et les bureaux de scrutin, deux concepts simples, déjà disponibles. Les difficultés surgissent quand il s'agit d'exprimer cette relation particulière.

Dans le système des vedettes-matière de la Library of Congress, certains sujets complexes peuvent être exprimés par l'emploi de sous-vedettes. Ces dernières exposent à la fois un aspect du sujet et une relation fonctionnelle. Parmi les sous-vedettes, seules les subdivisions de sujet ont un arrangement véritablement systématique; mais elles sont finalement très rares. Il y a par exemple: Automobiles-Motors-Oil Filters (Automobiles-Moteurs-Filtres à huile). Dans la majorité des cas, il s'agit d'exprimer une relation fonctionnelle entre les termes composant la vedette là où il n'y a pas d'expression consacrée par les spécialistes du domaine: Automobiles-Corrosion (Automobiles-Corrosion); Automobiles-Registration (Automobiles-Immatriculation); Automobiles-Motors-Cooling (Automobiles-Moteurs-Refroidissement).

Depuis la publication du *Cataloging Service*, bulletin 111, (Fall 1974), on assiste à

une évolution importante des principes d'élaboration de la liste des vedettes-matière de la Library of Congress. On y permettait, en effet, l'emploi d'un grand nombre de «subdivisions affranchies» (free-floating subdivisions). Ce concept signifie qu'on peut les employer à volonté sans qu'il soit nécessaire de les trouver sous chaque sujet ni de les intégrer dans la liste au moment de leur premier usage. Elles se présentent sous deux formules: une liste de subdivisions de forme et de sujet dont il faut scrupuleusement respecter le mode d'emploi donné sous chacune dans l'introduction de la 8th ed. de *Library of Congress Subject Headings*; les subdivisions de vedettes-modèles dont le nombre a été considérablement augmenté et qu'on peut employer sous les vedettes appartenant à la même catégorie.

Malgré une plus grande souplesse introduite par ce concept, une certaine rigidité reste encore la marque des vedettes-matière de la Library of Congress. Toute vedette combinant de manière particulière des termes qui existent néanmoins déjà dans la liste, mais qui entretiennent des relations fonctionnelles non exprimables par l'emploi des sous-vedettes, doit être établie par les éditeurs: Socialism and youth (Socialisme et jeunesse); Women and socialism (Femmes et socialisme); Plants, Effect of wind on (Plantes, Effets du vent sur les).

Ainsi, pour tous les sujets complexes dont l'indexation ne peut pas être totalement précoordonnée, la stratégie de la recherche documentaire devra penser à compléter la coordination manquante à l'aide des opérateurs booléens. C'est donc d'un système mixte qu'il s'agit. L'indexation est dans un premier temps précoordonnée jusqu'aux limites permises par le système. Ces limites atteintes, des vedettes-matière doivent être ajoutées de manière indépendante pour exprimer des aspects qui jouent un rôle dans l'expression du sujet.

La syntaxe de PRECIS

PRECIS a réalisé ce qui peut paraître impossible à ce stade de notre réflexion: un système d'indexation à partir des sujets

totalelement précoordonnée, constituant en quelque sorte des vedettes-matière dont chaque terme ou mot d'un terme composé choisi par l'indexeur sert de point d'accès. Chacun est accompagné de tous les autres termes, exprimant ainsi dans chaque rubrique (on peut lire vedette-matière) le sujet complet du document. Le terme d'accès est l'élément à partir duquel s'organisent tous les autres termes suivant le principe fondamental de la dépendance du contexte, principe qui détermine dans chaque rubrique produite l'ordre de citation des termes. Cet ordre de citation est constant et valable pour tous les domaines du savoir et pour tous les niveaux de complexité; il assure ainsi l'univocité de la rubrique. Plus concrètement, chaque élément de l'énoncé du sujet, terme ou mot d'un terme composé, est affecté d'un indicateur qui identifie sa fonction syntaxique et dont la valeur ordinale fixe sa position dans la séquence de termes qu'on doit au préalable rédiger. Selon les indicateurs en présence, trois formats de manipulation détermineront la place de ces éléments dans chaque rubrique. De plus, chaque terme composant une séquence, qui est tiré du thesaurus ou qui y sera intégré, commande automatiquement au moyen d'un numéro qui lui est attribué (le RIN, Reference Indicator Number) un réseau de renvois exprimant les relations d'équivalence, hiérarchiques et associatives du terme. Pour arriver à ce résultat, il fallait qu'une étape soit franchie: il fallait que soient identifiés les éléments constitutifs de toute classification documentaire.

Les classifications documentaires traditionnelles du type LC et Dewey ont été essentiellement basées sur un processus de division appliqué à l'univers des connaissances. Les éléments qu'elles organisaient étaient regroupés pour ce faire selon leur degré de ressemblance et intégrés dans une structure hiérarchique arangée du complexe au simple. On y reconnaît les relations de genre à espèce et de tout à partie. Dans une telle structure, seuls les concepts trouvent facilement leur place. Pourtant, il a fallu aménager cette structure pour y incorporer des sujets complexes, dont la complexité implique des relations fonctionnelles entre plusieurs concepts. On espérait ainsi pouvoir énu-

mérer tous les sujets qui se présenteraient et leur attribuer une place précise dans un système dont les principales classes étaient les disciplines. Devant l'arbitraire et les difficultés de mise à jour de tels systèmes, plusieurs en vinrent à remettre en question leurs structures.

On peut voir l'influence des classifications philosophiques des sciences de l'époque sur la construction de ces classifications documentaires. Toutes s'appuyaient sur la science du 19^e siècle qui avait des systèmes complexes de la nature une vision statique, dont le comportement était prévisible, reproductible et réversible. On croyait que ces classifications des connaissances reflétaient et continuaient de refléter la croissance des connaissances par de simples additions aux schèmes déjà existants sans prévoir les restructurations que la croissance à un certain seuil nécessite. La science contemporaine a de la nature une vision beaucoup plus dynamique et déclare son comportement, dans les systèmes complexes, imprévisible, irréproductible et irréversible. On tente plutôt d'identifier leurs éléments, mais surtout on se concentre sur leurs interactions.

Cette étape allait être franchie quand Ranganathan, fortement impressionné et influencé par Bliss, proposa une approche complètement nouvelle, la classification à facettes, sorte de mathématique qualitative: «Possibly it is in the way in which Ranganathan tried to analyze subjects into their constituent elements, and to subsequently formulate and formalize through his «facet formulas» the statements about subjects found in the books and other kinds of documents that we find this new kind of qualitative mathematics expressed that Leibniz had been looking for»³.

Cette approche analytique exige que nous soyons conscients de notre propre activité classificatoire, que nous la contrôlions et la maîtrisions à l'aide d'une manipulation réfléchie de concepts classificatoires, facettes appartenant à des catégories, et que nous reconnaissons *les nom-*

3. Ingefraut Dahlberg, «Classification theory, yesterday and today», *International Classification*, vol. 3, no. 2 (1976), 86.

breuses façons possibles de classifier les sujets complexes. C'est dans cette lignée que se situe PRECIS et c'est en cela qu'on ne peut oublier qu'il est né à la suite de recherches sur les fondements d'un nouveau système général de classification.

PRECIS est essentiellement à la fois une technique d'analyse documentaire et un système d'indexation; orienté vers le contrôle bibliographique, il a été conçu d'abord pour produire l'index alphabétique par sujets à la *British national bibliography* et pour fournir un accès par sujets aux bandes UK/MARC. Nous présenterons les principales caractéristiques du système en excluant cependant celles qui sont relatives uniquement à la production d'un index périodique.

En tout premier lieu, il nous faut dire que PRECIS partage l'ensemble des tâches impliquées dans la génération de rubriques entre l'ordinateur et l'indexeur. Celui-ci doit d'abord formuler une phrase-sujet, puis faire l'analyse conceptuelle de cet énoncé. Il rédige ensuite une séquence de termes ordonnés selon la valeur des indicateurs qui leur furent attribués à la suite de l'analyse fonctionnelle de chacun, puis choisit ceux qui doivent faire l'objet d'une entrée. L'ordinateur, à partir de cette séquence et des codes qu'elle comprend, entérine ces décisions et produit les rubriques conséquentes selon les règles du système.

Les fondements de PRECIS s'éclairent à la lumière de deux objectifs préétablis: chaque rubrique devait être intelligible et significative à la lecture, ressemblant autant que possible au discours des langues naturelles et chacune devait exprimer le sujet complet du document et donc contenir l'ensemble des concepts indexables. L'indexation devait donc être totalement précoordonnée. Pour rencontrer ces deux objectifs avec constance, on a élaboré un ensemble de règles qui permettraient de respecter dans tous les cas le principe de la dépendance du contexte. Ce principe a été la clef de voûte du système. Chaque terme dans la rubrique établit le suivant dans son contexte logique, chacun entretenant avec celui qui le précède et celui qui le suit une relation biunivoque et avec les deux une relation simultanée. On reprodui-

sait de la sorte les conditions du discours où le sens des mots utilisés dépend du contexte et où le sens complet de la phrase n'est connu qu'avec le dernier mot.

La syntaxe du système, c'est l'ensemble des indicateurs de fonction constitués en schéma, plus les démarches intellectuelles qui leur sont associées dans le choix de l'indicateur approprié. Les indicateurs de fonction forment une partie des codes attachés aux termes de la séquence mais ne font évidemment pas partie intégrante des rubriques. Le but de cette syntaxe est l'expression des relations fonctionnelles entre les concepts indexables du sujet du document analysé. Les indicateurs numériques, de 1 à 6, commandent la typographie et le format des rubriques; les indicateurs alphabétiques règlent la signification des rubriques. Nous n'introduirons pas ici la mécanique intime de la syntaxe de PRECIS. Le *Manual* est seul habilité à le faire. Nous prenons pour acquis comme plusieurs avant nous que cette syntaxe fonctionne, et très bien. Plutôt qu'à son fonctionnement, attardons-nous à sa fonction en l'illustrant par un exemple.

Les indicateurs de fonction sont continuellement présents à l'esprit de l'indexeur d'expérience et sont donc intimement liés à l'analyse documentaire; ils la précèdent même. De sorte que réciproquement et parfois par approximations successives, les termes seront marqués selon leur fonction par les indicateurs appropriés, mais aussi ces derniers feront rechercher des concepts qui leur correspondent. Ainsi, après que l'énoncé du sujet aura été rédigé ou intériorisé, la première étape consistera à identifier un terme qui dénote une idée d'action. Prenons un exemple: pollution des mers par les hydrocarbures. Ici, l'action, c'est la pollution. Ce terme sera affecté d'un indicateur (2). Il faut ensuite identifier le «système-clé» qui sera l'objet de l'action si le mode est transitif et l'agent si le mode est intransitif. Dans le cas présent, les mers seront l'objet de l'action. Ce terme sera affecté d'un indicateur (1). Si l'agent de l'action n'avait pas été initialement mentionné dans la phrase-sujet, notre connaissance de l'existence de l'indicateur (3) nous l'aurait fait rechercher. Si l'auteur a conduit des recherches sur ce sujet en

général, mais qu'il se trouvait à un endroit précis où il a cueilli les données dont il a déduit ses conclusions générales, ce lieu ne peut pas être considéré comme l'environnement général du sujet, auquel cas il recevrait l'indicateur (0), mais plutôt comme une région étudiée. Supposons qu'il s'agisse de la mer Rouge. Deux indicateurs sont nécessaires: un indicateur (q) introduisant ce lieu et dépendant de l'indicateur (5) dont le contenu servira à préciser le contexte dans lequel ce lieu est utilisé. Le résultat final donnera la séquence suivante:

- (1) mers
- (2) pollution \$v par les \$w des
- (3) hydrocarbures
- (5) régions étudiées
- (q) mer Rouge

et produira les rubriques suivantes:

Mers

Pollution par les hydrocarbures —
Régions étudiées: mer Rouge

Pollution. Mers

Par les hydrocarbures — Régions étudiées: mer Rouge

Hydrocarbures

Pollution des mers — Régions étudiées: mer Rouge

Mer Rouge. Régions étudiées

Mers. Pollution par les hydrocarbures

Les deux premières rubriques sont manipulées selon le format normal, la troisième selon le format dit de la transformation prédicative et la quatrième selon le format inversé. Bien sûr, l'indexeur n'a pas à se soucier de tous ces types de manipulation, même s'il doit connaître le résultat des transformations qu'ils entraînent. L'ordinateur détectant un indicateur (3) précédé d'un indicateur (2) utilisera la transformation prédicative quand le terme «hy-

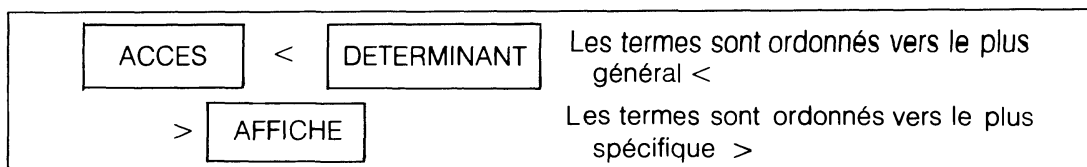
drocarbures» occupera la position d'accès. Il utilisera également le format inversé dès qu'il détectera l'indicateur (5) et que le terme «mer Rouge», précédé de l'indicateur (q), occupera la position d'accès. On aura remarqué également l'emploi des connecteurs \$v et \$w qui assurent une formulation en un langage plus naturel.

Arrêtons-nous là. Continuer ainsi nous entraînerait trop loin, exigerait trop d'explications. Pourtant, on pourrait déjà rétorquer avec raison que plusieurs rubriques peuvent être produites à partir d'une même séquence avec des procédures beaucoup moins complexes tout en conservant à chaque fois le contexte complet du sujet. Ainsi, nous pourrions avoir:

Mers. Pollution par les hydrocarbures
Hydrocarbures. Pollution des mers par les hydrocarbures
Pollution des mers par les hydrocarbures

Ce que nous perdons ici, c'est la collocation des sujets semblables (et le bouquillage intellectuel) réalisée par la disposition des rubriques sur deux lignes et leur composition en trois positions fondamentales (accès, déterminant, affiche) que peuvent occuper successivement les termes de la séquence originale et qui rend possible l'expression des relations simultanées qu'entretient le terme d'accès avec les termes du déterminant et ceux de l'affiche.

L'accès, seul ou avec le déterminant (s'il y en a), constitue la vedette. Dans un index imprimé (nous incluons sous ce vocable les supports qui reproduisent une page, par affichage sur terminal, sortie COM, etc.), pour des sujets ayant la même vedette, nous retrouverons sous celle-ci, dans l'affiche, les aspects plus particuliers de chacun. A ce point, on peut également prétendre que les contenus des différentes précisions affichées sous la vedette constituent autant de sujets étudiés sous



l'aspect exposé par la vedette. Cette caractéristique en ferait un système de classification basé sur les objets d'étude par opposition à ceux basés sur les disciplines.

Vers une étude comparative

Dès le départ, nous postulons que toute étude voulant comparer l'efficacité respective des deux systèmes en présence, vedettes-matière de la Library of Congress et celles produites par PRECIS, devra respecter quelques conditions essentielles à l'intégrité des résultats qu'on obtiendra. Et d'abord, elle devra être indépendante de toute application antérieure à l'étude, car à ce moment c'est autre chose qu'on évaluerait: la compétence des agences qui les utilisent. Avant d'être entreprise, elle devra, au préalable, pousser aux limites les possibilités d'évolution de l'un et les possibilités d'exploitation de l'autre. Ce qu'il faudra évaluer après cela, c'est le taux de réussite des recherches documentaires effectuées dans l'un et l'autre, limitées ou exhaustives, conditionnées qu'elles sont par le contenu des classes consultées, elles-mêmes déterminées par la plus ou moins grande spécificité des termes qui les représentent (en ce sens que les représentations verbales sont coextensives aux sujets des documents) et par leur plus ou moins exacte coordination.

Refaisons rapidement une revue de leurs propriétés respectives. Sur le plan terminologique, tous deux développent un vocabulaire contrôlé, structuré et ouvert. Nous avons déjà reconnu cette dernière propriété dans le vocabulaire utilisé par PRECIS. Voici, si on en doutait pour les vedettes-matière de la Library of Congress, une citation intéressante qui confirme notre assertion: «In general, subject catalogers are required to establish each new topic as a subject heading, when the topic is first encountered in a work being cataloged. This practice is a corollary to Haykin's rules of specificity: headings assigned to a work *must* [c'est nous qui soulignons] represent its contents exactly, even if one must establish new headings to do so. Catalogers are not to wait for a concept to appear in several works before establishing it if it

represents a discrete, identifiable concept»⁴.

Sur le plan syntaxique, l'existence d'une définition opérationnelle de la notion de concept, à l'aide des indicateurs de fonction, a contribué à éliminer dans PRECIS toute nécessité d'inversion des termes composés. En effet, on a d'une part reconnu qu'il y a plusieurs façons également valables de classer les termes composés: l'élément général de ceux-ci, désignant une plus grande classe, permet tout aussi bien le regroupement de classes plus précises définies par l'élément plus spécifique; d'autre part, les indicateurs de fonctions associés au «differencing» permettent de rendre la rubrique également accessible par tous les mots désirés composant le terme d'accès.

Une semblable définition pourrait être établie pour le système de vedettes-matière de la Library of Congress. Tous les termes composés ne représentant qu'un seul concept seraient dans tous les cas entrés directement, abandonnant la pratique des inversions. Elles devraient cependant continuer d'être utilisées en renvois, exprimant des relations d'équivalence d'un type particulier, s'il n'est pas possible d'y substituer des relations hiérarchiques appropriées (ce qui doit être rare), car alors, la vedette ne serait accessible que par un des éléments du terme composé. On réserverait l'emploi des sous-vedettes aux termes comprenant plus d'un concept tel que défini pour le système. Nous l'avons déjà dit, les sous-vedettes expriment à la fois un aspect particulier du sujet, propriété, partie, processus, etc. et cette relation fonctionnelle. Pensons aux avantages retirés de la prévisibilité de la représentation formelle d'une classe.

Le même sujet peut être exprimé par des unitermes, mots isolés, tels que trouvés dans les documents et sans relations: c'est le minimum de traitement à l'entrée. Le maximum est représenté par des descripteurs totalement précoordonnés syntaxiquement et structurés sémantiquement: c'est la formule proposée par PRECIS.

4. «Headings for new concepts», *Cataloging Service*, bulletin 121 (Spring 1977), 16.

Quelque part entre les deux, mais plus près du dernier, nous avons trouvé les vedettes-matière de la Library of Congress.

Elles nous sont apparues en pleine évolution également sur le plan syntaxique, mais rencontrant aussi certaines limites, inhérentes celles-là au système. On a déverrouillé les séquences enchaînées que constituaient certaines vedettes-matière en affranchissant un grand nombre de «subdivisions». Malgré cela, l'ordre de citation est unidirectionnel: une seule représentation formelle est acceptée. On accède à cette vedette composée par un réseau de renvois parfois fort compliqué. Une séquence de PRECIS rend accessibles plusieurs rubriques, par tous les accès qu'on a choisis, chaque vedette ainsi produite étant recomposée pour respecter dans chacune le principe de la dépendance du contexte: plusieurs représentations formelles équivalentes sont acceptées.

La Library of Congress, après s'être interrogée sur le meilleur type de système d'indexation dans l'optique d'une recherche documentaire automatisée et avoir étudié les possibilités d'adoption de PRECIS (ce dernier ayant été conçu dans cette optique)⁵ concluait que les limites à la précoordination totale seraient corrigées par la logique booléenne. A l'inverse, il est évident que celle-ci ne peut être d'un grand intérêt dans un système d'indexation totalement précoordonnée. Chaque terme qui sert d'accès à une rubrique est placé dans un réseau de relations sémantiques et syntaxiques. Il faudra donc que soient conçues des stratégies de recherche exploitant au maximum ces caractéristiques.

Conclusion

Tout n'a pas été dit sur PRECIS, bien sûr. En plus, nous avons employé une démarche inhabituelle, avec tous les risques qu'elle pouvait comporter. Sur un arrière-plan esquissé à grands traits et constitué par les différentes facettes de l'analyse documentaire, nous avons su-

perposé les aspects correspondants de PRECIS pour mieux démarquer ceux qui le caractérisent. D'aucuns reprocheront à cet article son caractère trop théorique. Nous répétons, en d'autres mots, qu'une étude quantitative comparant les coûts, sans étude qualitative préalable, présente souvent des résultats fort douteux qui démontrent de dangereuses tendances au réductionnisme, car elle ne connaît pas bien ce qu'il faut mesurer. Il nous faut savoir compter avec les phénomènes d'évolution, de croissance, etc. qui sont tous liés à la durée de l'existence des systèmes.

Les connaissances de l'homme sont le reflet d'un univers partiellement accessible à ses sens et à son esprit et font habituellement l'objet d'un consensus. Lorsque s'agrandit cette partie de l'univers accessible, un nouveau paradigme prend place qui oblige à restructurer les connaissances. Les langages documentaires ont pour fonction d'organiser ces connaissances quand elles sont matérialisées dans des documents. Il faut donc dans cette perspective se demander dans quelle mesure et à quelles conditions ils remplissent bien cette fonction. Le système des vedettes-matière y a réussi depuis longtemps déjà tant bien que mal, malgré la détérioration et l'usure qu'apporte le temps, avec un minimum de règles. Faut-il conclure que PRECIS, avec une technique d'analyse documentaire et une philosophie d'indexation exposée dans un manuel de plus de 500 pages, est mieux préparé pour l'affronter?

5. «Policy for subject heading changes», *Cataloging Service*, bulletin 123 (Fall 1977), 10-11.