DOCUMENTATION
ET INFORMATION

R. C. CROS
J. C. GARDIN
F. LEVY

L'AUTOMATISATION DES RECHERCHES DOCUMENTAIRES

UN MODÈLE GÉNÉRAL — LE SYNTOL

A[illegible]R-VILLARS | PARIS

L'AUTOMATISATION
DES RECHERCHES DOCUMENTAIRES

DANS LA MÊME COLLECTION

B. C. VICKERY : La classification à facettes. Guide pour la construction et l'utilisation de schémas spéciaux. 1963.

R. DUBUC : La classification décimale universelle. Manuel pratique d'utilisation. 2e éd. revue et augmentée. 1965.

Z. DOBROWOLSKI : Etude sur la construction des systèmes de classification. 1964.

L'organisation de la documentation scientifique. Etudes de : J.-C. GARDIN, E. DE GROLIER, F. LÉVERY, l'Association nationale d'études pour la documentation automatique. Nouveau tirage. 1966.

C. LEGEARD : Guide de recherches documentaires en démographie. 1966.

Economie générale d'une chaîne documentaire mécanisée, par F. ALOUCHE, N. BELY, R.-C. CROS, J.-C. GARDIN, F. LÉVY, J. PERRIAULT. 1967.

DOCUMENTATION ET INFORMATION
Collection dirigée par Paul Poindron,
Directeur des études de l'Institut national
des techniques de la documentation

L'AUTOMATISATION DES RECHERCHES DOCUMENTAIRES

Un modèle général « LE SYNTOL »

par

R.-C. CROS, J.-C. GARDIN, F. LÉVY

(CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
ET
MAISON DES SCIENCES DE L'HOMME)

2e édition revue et augmentée

PARIS
Gauthier-Villars

1968

CROS (R.-C.), GARDIN (J.-C.), LEVY (F.). — L'automatisation des recherches documentaires : un modèle général, le « Syntol ». 2e éd. revue et augmentée. — Paris, Gauthier-Villars, 1968. —
21 cm, XX-262 p., 19 fig.
(Documentation et information)

Documentation, automatisation. — Automatisation, documentation. — Syntol.

CDU : 002 : 65 011.54

TABLE DES MATIÈRES

PRÉFACE
A LA DEUXIÈME ÉDITION

Lorsque parut la première édition de cet ouvrage, en 1964, le SYNTOL avait déjà trois ans. En effet, les grandes lignes du système documentaire qui venait alors de recevoir ce nom furent présentées dès 1961-62 dans une série de rapports adressés à l'organisme initiateur de l'étude, l'EURATOM (Communauté Européenne de l'Energie atomique); le risque est donc grand aujourd'hui de présenter des résultats déjà vieillis, en particulier dans un domaine où les innovations techniques ont la réputation de compter chaque année par dizaines, sans donner aucun signe d'une stabilisation des idées ou des pratiques en matière d'automatique documentaire. Cependant, ce dernier point de vue n'est pas le nôtre; en proposant de construire il y aura bientôt dix ans un système *général* de documentation automatique, nous partions implicitement d'une hypothèse inverse, selon laquelle un grand nombre des méthodes appliquées dans ce domaine étaient en fait réductibles à un modèle abstrait, dont chacune d'elles n'était finalement qu'une interprétation parmi d'autres possibles. Rien n'est venu depuis infirmer ce point de vue : les outils linguistiques utilisés en pratique pour l'automatisation des « recherches documentaires » (1) demeurent essentiellement les mêmes (lexiques de « descripteurs » et « thesaurus », pour le contrôle du vocabulaire scientifique; indicateurs de « rôle » ou relations logiques pour l'expression éventuelle de rapports syntaxiques entre les descripteurs, etc.); quant aux opérations élémentaires mises en jeu dans les processus de recherche, elles continuent à porter les mêmes noms (intersection, réunion, etc. au sens de l'algèbre de Boole), et à se prêter au même genre d'assemblage en macro-opérations indi-

(1) Les termes mentionnés entre guillemets dans cette Préface sont définis dans le Glossaire, *in fine*.

viduelles — désignées plus loin sous le nom de « modulations » — pour la commodité des formulations. L'on est ainsi fondé à présenter à nouveau un ouvrage qui ne visait il y a quatre ans qu'à tirer certaines conséquences de cette constatation, sur le plan des « langages documentaires » et des « métalangages » d'exploitation.

Il n'en reste pas moins que nombre de points appellent aujourd'hui un commentaire, suscité tantôt par les critiques — souvent judicieuses — auxquelles la présentation du SYNTOL a donné lieu, tantôt par les faits eux-mêmes, c'est-à-dire par des recherches ou par des applications nouvelles, en matière d'automatique documentaire, qui conduisent à compléter ou à nuancer certains passages de la première édition.

1. Et d'abord le nom même du SYNTOL : on a regretté que le suffixe « -OL » puisse suggérer qu'il s'agissait ici d'un nouveau langage de programmation (comme ALGOL, COBOL, etc.), alors que l'objet désigné est un « langage documentaire », antérieur en quelque sorte au problème de la programmation. La remarque est juste, mais sans doute moins fondée aujourd'hui qu'elle n'a pu l'être naguère, l'usage ne s'étant pas imposé de baptiser d'un nom en -OL tous les langages de programmation orientés vers des catégories d'applications particulières. Une observation plus sérieuse concerne l'ambiguïté du préfixe, cette fois : SYNT- est ici la forme abrégé de « syntagmatic », lequel terme ne désigne en fait qu'un aspect de l'« organisation » linguistique du SYNTOL, à savoir le mode d'expression syntaxique par chaînes de « syntagmes » élémentaires (*infra*, p. 20). Le second aspect de l'organisation, « paradigmatique », semble relégué au second plan, situation d'autant plus paradoxale que celle-ci est en fait plus couramment attestée dans les langages documentaires existants — sous forme de renvois, classifications, réseaux sémantiques, etc. — que ne l'est l'organisation syntaxique elle-même, parfois complètement absente (cas des langages réduits à des listes de descripteurs, sans procédés de mise en relation logique). Sans doute cette impropriété de l'appellation SYNTOL avait-elle été signalée en son temps (*infra*, p. 21); il restait alors à expliquer pourquoi nous nous y étions néanmoins tenus. La raison a été donnée dans un ouvrage paru un an plus tard en anglais sous ce titre même (2) : l'on y précisait

(2) J. C. GARDIN, *SYNTOL*, Rutgers Series on Systems for the Intellectual Use of Information, ed. S. Artandi, vol. II, Rutgers, The State University, New Brunswick (N. J.), 1965, pp. 42-4.

que les « syntagmes » caractéristiques du SYNTOL, sous leur format général (Ri, *a*, *b*) — où *a* et *b* désignent deux descripteurs, et Ri une relation particulière de l'un à l'autre — servaient à exprimer non seulement la structure syntaxique éventuelle des « représentations » documentaires, mais aussi l'organisation paradigmatique elle-même, le plus souvent réductible à un ensemble de syntagmes binaires de ce genre (ex. : « *a*, voir *b* », ou « *a*, élément de la classe *b* », ou « *a*, propriété de *b* », etc.). Ce n'est pas ici le lieu de reprendre la démonstration (3); contentons-nous de poser qu'il faut élargir dans ce sens l'interprétation de notre acronyme, et comprendre le préfixe SYNT comme se référant à un procédé d'expression très général, par syntagmes (Ri, *a*, *b*), des caractéristiques *à la fois* sémantiques (organisation paradigmatique) et logiques (organisation syntagmatique) du langage documentaire.

2. Cette première observation conduit naturellement à la question suivante, souvent posée, concernant la place du SYNTOL par rapport à l'ensemble des systèmes documentaires existants. Remarquons tout d'abord que le SYNTOL n'est pas à proprement parler un langage, immédiatement défini par un lexique et/ou par une grammaire propres, mais plutôt un cadre logico-linguistique où peuvent venir se couler la plupart des langages documentaires ainsi définis, à quelque niveau d'élaboration et pour quelque champ d'application que ce soit. Ce point paraît avoir été souvent mal compris, voire méconnu, sans doute parce que nous ne l'avions pas suffisamment mis en relief dans la présentation du SYNTOL. Pour les besoins de l'expérimentation initiale (*infra*, chap. 5), il fallut en effet choisir une incarnation particulière du modèle : élection d'un domaine scientifique de référence, constitution de listes de descripteurs pour l'indexation des documents propres à ce domaine, adoption d'un type de structure pour l'organisation sémantique de ces listes ou « lexiques », sélection enfin d'une certaine grammaire pour l'expression de rapports syntaxiques entre descripteurs dans la représentation de chaque document. Dans notre esprit — et aussi, croyions-nous dans nos écrits — il était clair que l'ensemble de ces choix définissait non

(3) *Op. cit.* p. 22-5 : voir aussi F. ALOUCHE, N. BELY, R.-C. CROS, J.-C. GARDIN, F. LEVY, J. PERRIAULT, *Economie générale d'une chaîne documentaire mécanisée*, Paris, Gauthier-Villars 1967 (abrégé ci-dessous *Economie générale*) pp. 109-111, où l'on propose une codification des « lexiques documentaires » (ex. : classifications hiérarchiques) au moyen d'ensembles de syntagmes : [(R, x, y) (R , y, z)].

pas *le* SYNTOL, mais *une* modalité parmi d'autres du SYNTOL, ou encore, dans notre terminologie même, un des « états » possibles du modèle (*infra*, pp. 17-8, 44, 90, 96-7, 109-112, 167-8, etc.). Pourtant, les cas de confusion ne manquèrent pas, où critiquant tel ou tel de ces choix, l'on pensait mettre en cause le modèle tout entier, contrairement à son principe même. Le phénomène à cet égard le plus déroutant est l'obstination que l'on met à débattre des problèmes de syntaxe, dans les langages documentaires, comme s'il existait une réponse universellement valide à la double question suivante : a) est-il nécessaire d'exprimer tout ou partie des rapports logiques observés entre les descripteurs, dans les représentations indexées ? b) si oui, quel est le nombre optimal des relations qu'il convient de différencier, et de quelle nature sont-elles ? Répétons une fois encore que le SYNTOL ne prend parti sur aucun de ces points : les « états » les plus simples du système ne comportent aucune syntaxe (*infra*, p. 110), et si nous avons choisi de présenter ici une version plus élaborée, à quatre relations syntaxiques (pp. 48-51), ce fut pour illustrer la manière dont celles-ci devaient être définies et ordonnées, en vertu des exigences de généralité et de flexibilité propres au SYNTOL, et non bien sûr pour suggérer l'adoption universelle de la même grammaire, dans toutes les applications documentaires (4).

Malheureusement, la reconnaissance de cette fonction abstraite du SYNTOL risque de conduire à d'autres malentendus : s'il ne s'agit pas d'un langage d'indexation, mais seulement d'un moyen d'exprimer dans un même format les propriétés structurelles des divers outils sémantiques et syntaxiques conçus pour l'analyse de « corpus » particuliers, n'est-il pas abusif de prétendre à la création d'un système documentaire nouveau, comme ce livre même semble le postuler ? La **réponse tient en deux points** : (a) En premier lieu, il faut avoir à l'esprit que le langage d'indexation, si raffiné soit-il, n'est jamais qu'un des éléments nécessaires à la constitution d'un « système documentaire » ; ce dernier est aussi défini par l'ensemble des procédures mises en jeu pour l'« enregistrement » et pour la « recherche » des informations,

(4) Cette liberté de l'interprétation syntaxique apparaît sans doute plus nettement dans l'ouvrage en anglais déjà cité : J. C. GARDIN, SYNTOL, etc. (1965) pp. 25-9; elle est en outre tout à fait manifeste dans le compte rendu d'une application SYNTOL récente, où deux « états » différents furent considérés pour l'exploitation du même corpus : l'un plus sommaire, *sans aucune syntaxe*, l'autre au contraire plus raffiné, comportant non plus quatre mais *dix relations* : voir *Economie générale*, pp. 87-8 et 92.

indépendamment du langage dans lequel celles-ci sont indexées. On retrouve cette distinction dans le présent ouvrage, dont la première moitié concerne les aspects linguistiques du modèle (pp. 26-106), tandis que la seconde traite de ses caractéristiques opératoires, questions de langage mises à part (5). (b) D'autre part, s'il est vrai que le SYNTOL a d'abord été conçu comme un modèle auquel on pouvait ramener nombre de systèmes documentaires superficiellement différents, il reste que rien n'empêche d'utiliser le modèle de façon inverse, pour composer un ou plusieurs systèmes nouveaux, correspondant à des interprétations inédites. C'est la situation à laquelle nous sommes d'ailleurs arrivés, en présentant le modèle à travers une de ses manifestations construite pour les besoins de la démonstration. On comprend dans ces conditions qu'une certaine confusion ait pu s'instaurer, quant à l'objet même de l'entreprise.

3. Celle-ci est évidemment fondée, du point de vue linguistique, sur le postulat de la nécessité d'une forme quelconque d'« indexation », dans les systèmes documentaires. Comme cette thèse était déjà combattue à l'époque où paraissait la première édition de l'ouvrage, le lecteur est en droit de s'interroger à nouveau sur l'état présent de la question. Il est heureusement facile de le résumer en quelques mots. (a) Après une période d'engouement pour le « traitement des informations en langage naturel », les contradictions et les inconséquences actuelles de cette démarche (*infra*, pp. 26-31) n'ont pas manqué d'apparaître (6); et celle-ci n'est plus avancée sous sa forme brute qu'à l'occasion de rencontres éphémères entre les problèmes d'automatique documentaire et d'amusants visionnaires qui en ont une connaissance de seconde main. (b) En revanche, toutes les applications notables de l'informatique en documentation continuent à faire appel à des langages d'indexation, dont la mise au point et l'entretien retiennent chaque

(5) Un ouvrage entier vient d'ailleurs d'être consacré aux seules propriétés opératoires du SYNTOL, sans référence particulière à ses aspects linguistiques : *Economie générale*, etc., déjà cité (*supra*, note 3).

(6) Le mot « actuelles » est important : toutes les objections que nous voyons à la thèse du traitement de l'information scientifique en langage naturel disparaissent si l'on se place à une étape future du développement humain, où le langage naturel se réduirait à une seule langue (mais laquelle ?), et où cette langue se confondrait avec le langage scientifique lui-même. Les deux conditions sont nécessaires, et chacun est libre d'imaginer une échéance historique à son goût, mais non d'éluder le paradoxe.

année plus de soin (7). (c) Cependant, le projet même d'une automatisation éventuelle de l'analyse documentaire pourrait encore laisser penser que ces langages d'indexation n'auront d'utilité qu'un temps; il est facile de montrer que la vérité est exactement inverse, en ce sens que les algorithmes d'analyse automatique exigent tous le recours à une culture sémantique étendue — pour ne parler que de celle-là — c'est-à-dire en dernier ressort un outillage linguistique beaucoup plus riche même que les seuls « lexiques » ou « thesaurus » des langages documentaires actuels (*infra*, pp. 32-35) (8). (d) Sans doute est-on libre d'imaginer que

(7) Voir par exemple les travaux qui se multiplient aux Etats-Unis depuis quelques années, avec le concours financier d'organismes gouvernementaux, pour la constitution de lexiques ou thesaurus spécialisés, destinés à des applications documentaires sur machines : *Thesaurus of ASTIA Descriptors*, Armed Services Technical Information Agency, Arlington, Virginia, 1re éd., mai 1960; 2e éd., décembre 1962 — *Chemical Engineering Thesaurus*, American Institute of Chemical Engineers, New York, 1961 — *Thesaurus of Descriptors, Key-words and Cross References for Indexing and Retrieving the Literature on Water Resources Development*. U. S. Department of the Interior, Bureau of Reclamation, Denver, Colorado 1963 (édition provisoire) — *Thesaurus of Engineering Terms*, Engineers Joint Council, New York, 1re éd., mai 1964; 2e éd. annoncée en 1967 — *Thesaurus of Metallurgical Terms*, American Society of Metals, 1966 — *Thesaurus of FAA Descriptors*, 2e éd., Federal Aviation Agency, Washington 1965 — *Thesaurus of ERIC* [Educational Research Information Center] *descriptors : Phase I*, U. S. Department of Health, Education and Welfare, juin 1966 — *DoD-wide Technical Thesaurus*, Department of Defense, Washington, à paraître en 1968. En Europe, les études systématiques sont en revanche peu nombreuses; citons cependant l'*Euratom Thesaurus, Key-words Used within Euratom's Nuclear Energy Documentation Project*, Presses Académiques Européennes, Bruxelles 1964; 2e éd. parue en 1966 sous le titre *Euratom Thesaurus Indexing terms used within Euratom's Nuclear Documentation System*, 1re partie, Centre for Information and Documentation, Euratom, Bruxelles — le *TDCK-Circular Thesaurus System*, Centre d'Information Scientifique et Technique de l'Armée, La Haye, 4e éd., mai 1966.

(8) La démonstration de ce point de vue dépasserait malheureusement les limites de cette préface. Bornons-nous à citer les recherches à nos yeux les plus avancées sur le sujet, où la part des outils sémantiques dans l'analyse mécanique est évidente : aux U.S.A., G. SALTON, *The SMART Automatic Document Retrieval System — An Illustration*, Communications of the Association for Computing Machinery, vol. 3, n° 6, juin 1965, pp. 391-8, avec liste des rapports plus détaillés diffusés par le Computation Laboratory, Harvard University, sur le système SMART; en U.R.S.S., D. G. LAKHUTI, V. S. ČERNJAVSKIJ, *Ob Algoritmičeskom razpoznavanii značenij omonimov*, Naučno-tekhniceskaja Informatsija 1965, 1, pp. 46-7 (expériences d'analyse automatique du russe vers le langage documentaire PUSTO-NEPUSTO); en France, les travaux de la Section d'Automatique

cette « culture » sera un jour élaborée par la machine elle-même, grâce aux procédures d'« apprentissage » que nous évoquions il y a cinq ans (pp. 36-7); les réserves que nous annoncions alors, non sur le principe mais sur l'échéance pratique des recherches entreprises sur le sujet, conservent toute leur actualité (9).

4. Les trois points que l'on vient de passer en revue concernent les traits les plus généraux du SYNTOL : son nom (§ 1), sa fonction (§ 2), ses fondements (§ 3). D'autres observations portent sur certains détails seulement du système qu'il est bon par conséquent de reprendre, à la lumière de critiques ou de faits nouveaux apparus depuis la présentation initiale.

Citons tout d'abord, et plutôt pour mémoire, les reproches adressés à la terminologie. La nouveauté de certaines expressions a gêné plus d'un lecteur; ainsi l'opposition entre les axes « paradigmatique » et « syntagmatique » du langage documentaire, la dénomination des relations (« consécutive », « associative », etc.), les termes désignant les « modulations », et cette expression elle-même, etc. Souvent, il est vrai, les notions visées ne sont pas nouvelles, et peut-être eût-il fallu s'efforcer de conserver des appellations plus familières aux documentalistes (par exemple, « classification » au lieu d'« organisation paradigmatique », etc.). Si nous n'en avons rien fait, c'est que ces appellations courantes sont aussi ambiguës qu'elles sont répandues; il suffit pour s'en convaincre de parcourir la littérature spécialisée, ou d'entendre les débats habituels de la profession, plus souvent fondés sur des divergences non reconnues de terminologie que sur des oppositions de fond. Tout reste à faire, selon nous, pour que la « science de l'information », comme on la nomme dans les pays anglo-saxons, commence à mériter son titre, et d'abord un effort de rigueur dans la définition et dans l'usage des termes spécialisés, s'il en est. On peut déplorer le choix que nous avons fait de tel ou tel terme, jugé particulièrement abscons (ex. : « paradigmatique ») ou natu-

Documentaire (C.N.R.S.) menés avec le concours de la Délégation Générale à la Recherche Scientifique et Technique : M. COYAUD et N. SIOT-DECAUVILLE, *L'Analyse automatique des documents,* Mouton, Paris 1967; A. BORILLO, J. VIRBEL, etc. *Etudes sur l'indexation automatique,* 4 rapports ronéotypés (1966-7), à paraître dans un ouvrage en préparation (1969).

(9) Voir par exemple le compte rendu des travaux mentionnés plus loin (p. 36, note 1), dont une première phase est aujourd'hui achevée : *Rapport final sur les méthodes d'apprentissage automatique appliquées à la documentation* [ronéotypé], établi sous la direction de B. JAULIN (Centre de Calcul de la Maison des Sciences de l'Homme), en particulier les pp. I-1 à 25.

rellement ambigu (ex. : « variations »); mais on ne saurait échapper pour autant à l'obligation de définir de façon stricte les notions auxquelles les documentalistes se réfèrent (ou devraient se référer) constamment, sous quelque étiquette que ce soit. Le glossaire publié à la fin du volume marque un pas dans ce sens (10); rien ne s'oppose à ce que l'on en change les *termes*, à condition de respecter (ou d'affiner) la pertinence des *notions* qu'ils dénotent.

Cela dit, la forme de notre exposé n'est assurément pas sans tache: nous n'en prendrons pour exemple que l'usage certainement abusif du mot « formel », dans des sens métaphoriques et changeants qui n'ont rien de recommandable : en opposition tantôt avec « sémantique » (ex. : les catégories formelles de la grammaire, contrastées avec les classes sémantiques du lexique, pp. 54 sq.; ou encore les données formelles ou sémantiques mises en jeu dans l'interprétation des syntagmes ambigus, pp. 52, 61, etc.), tantôt avec « réel » (ex. : les rapports formels que dénote la relation « coordinative », contrastés avec les rapports réels que recouvrent les autres, p. 49), etc. La qualification « formelle » devrait pourtant être maniée avec d'autant plus de précaution que l'on entend décrire une construction dont ce devrait être la vertu principale; et l'on ne peut que faire appel à l'indulgence des théoriciens — linguistes ou logiciens — pour laisser subsister des écarts de ce genre (malheureusement nombreux) dans une présentation qui pêche finalement plutôt par défaut que par excès, du point de vue de la terminologie.

5. Ces questions de forme mises à part, les commentaires qu'appelle un nouvel exposé du SYNTOL se répartissent commodément en trois groupes, selon qu'ils concernent : (a) les détails de l'outillage linguistique, sur le plan syntagmatique tout d'abord (pp. 44-87) , (b) puis paradigmatique (pp. 88-106) ; (c) ou bien plutôt l'appareillage logique du système (métalangage, programme), pour l'exploitation automatique des « données » formulées au moyen des outils précédents (pp. 107-167).

Sur le premier point — « organisation syntagmatique » — nous avons déjà eu l'occasion de répondre à deux objections :

(10) Voir aussi les contributions plus étendues du Groupe d'Etude sur l'Information Scientifique, dans la même direction : *Informations sur les techniques documentaires,* Bulletin des Bibliothèques de France, 12e année, n° 6, juin 1967, pp. 214-222 et pp. 232-4.

l'une, justifiée, visant la confusion possible entre la partie et le tout dans le nom même du SYNTOL (§ 1); l'autre, au contraire mal fondée, où l'on conteste l'utilité de telle ou telle relation syntaxique particulière, voire de toute relation, pour les besoins supposés invariants de l'analyse et de la recherche documentaires (§ 2). Si cette dernière position ne fait que souligner la résistance des faux problèmes dans la « science de l'information », il n'en subsiste pas moins deux questions sérieuses, relatives toutes deux à la seule caractéristique fondamentale de la grammaire du SYNTOL, à savoir la réduction à des « graphes » composés de syntagmes binaires { Ri, *a*, *b* }, où Ri peut recevoir différentes interprétations.

(a) La première de ces questions porte sur le choix de ce format « syntagmatique », où l'on redoute parfois toutes sortes d'incommodités : une difficulté d'apprentissage et de maniement plus grande, pour les analystes, que dans le cas réputé inverse des indicateurs de « rôle » attachés à chaque descripteur; un encombrement accru des représentations indexées dans quelque mémoire que ce soit; l'impossibilité de bénéficier alors des avantages d'une organisation « inverse » (voir Glossaire, *s. v.*), etc. Nous ne contestons pas la réalité occasionnelle de ces inconvénients, dont certains sont d'ailleurs évoqués ici même (*infra*, pp. 145-6); mais le problème pour nous est ailleurs, et réside encore une fois dans le souci de généralité qui commande la structure du modèle. Les procédés d'expression syntaxique réputés plus simples, par indicateurs de « rôle » ou affixes fonctionnels (ex. : Agent, Effet, Moyen, etc., opérateurs binaires), ne restent simples en fait que si les représentations documentaires le sont aussi, c'est-à-dire lorsqu'elles se réduisent à un petit nombre de descripteurs, correspondant idéalement à une proposition logique et une seule. Il est facile de montrer que le procédé *doit* se compliquer, et se compléter d'un codage de type relationnel (opérateurs *n*-aires) dès lors qu'une même représentation comporte plusieurs propositions distinctes, tout en conservant la forme non-linéaire propre au langage documentaire. Cette démonstration est d'ailleurs esquissée ici même (p. 47), et nous ne la développerons pas davantage, nous bornant à répéter que le choix d'un format relationnal dans le SYNTOL est précisément destiné à écarter ce genre de limitation. Ses inconvénients, s'il en est, comptent donc moins à nos yeux que l'avantage d'une liberté totale d'extension, dans l'analyse syntaxique du discours; rien n'empêche au demeu-

rant de repasser d'un graphe SYNTOL à une expression équivalente en termes de « rôles » si la commodité de l'exploitation paraît l'exiger.

(b) Cette option admise, il reste à justifier celle d'un format binaire (Ri, *a*, *b*) plutôt que ternaire (Ri, *a*, *b*, *c*), ... ou *n*-aire; ou mieux la question est celle-ci : comment exprime-t-on par les seuls syntagmes binaires du SYNTOL des relations ternaires (ex. : « *a* entre *b* et *c* », « si *a* présent, *b* entraîne *c* », etc.), voire plus complexes (ex. : « effets comparés de *a* sur *b* et de *c* sur *d* », etc.) ? A vrai dire, le problème était posé dès la première présentation du SYNTOL (*infra*, pp. 74-5); mais l'insistance des commentaires ultérieurs sur cette difficulté laisse penser qu'il convenait de la traiter plus sérieusement, alors même qu'elle ne s'était guère manifestée, en pratique, dans les applications expérimentales du langage. La question fut donc reprise lors d'un séminaire consacré au SYNTOL, il y a quelques années; et l'on trouvera dans le compte rendu en anglais de cette réunion les compléments d'information voulus (11).

(c) Une dernière observation se recommande enfin, concernant un trait de l'organisation syntaxique du SYNTOL qui nous paraît aujourd'hui mal venu : il s'agit des « opérateurs » que l'on envisageait d'associer à certains types de syntagmes, généralement pour préciser l'interprétation de la relation sous-jacente, sans toutefois renoncer à utiliser celle-ci seule, dans une acception délibérément plus vague, chaque fois que le domaine d'application (ou plus immédiatement le contexte de la représentation) autorisait l'ellipse : *infra*, pp. 61-6. A l'usage comme à la réflexion, il apparut que ces spécifications complémentaires pouvaient être prises en charge par le « système central » des relations (pp. 48-9), dont il suffisait de poursuivre les subdivisions hiérarchiques pour retrouver les différenciations voulues, tout en conservant la faculté de regroupements (ou d'ellipses) éventuels (voir p. 49, fig. 6). On trouvera un exemple de ce parti dans un « état » syntaxique plus raffiné du modèle (10 relations), utilisé pour une application-pilote aujourd'hui publiée (12).

6. Passons maintenant à l'autre axe d'organisation du SYNTOL, l'axe « paradigmatique ». Un regret s'est manifesté de plu-

(11) J. C. GARDIN, *SYNTOL*, etc. (voir plus haut, note 2), pp. 29-33, et pp. 89-90.

(12) *Economie générale* (voir note 3), p. 87, fig. 19.

sieurs parts, tout d'abord, à propos du silence volontairement observé dans cet ouvrage sur le contenu des classifications évoquées (pp. 90, 98, etc.). La raison de cette discrétion était double : (a) en premier lieu, comme pour la syntaxe, il était à craindre que l'on confondît alors l'exemple particulier et le modèle abstrait; (b) d'autre part, ces classifications, si étendues fussent-elles, n'en restaient pas moins assez pauvres du point de vue structural qui nous occupe : elles ne mettent en jeu qu'une seule « dimension », selon la terminologie proposée plus loin (p. 90), c'est-à-dire que la relation analytique est unique, et de plus non spécifiée. Quoi qu'il en soit, ces sous-produits de l'étude sont aujourd'hui publiés (13), et l'on peut vérifier qu'ils intéressent probablement davantage les spécialistes des domaines considérés — physiologues, psychologues, et anthropologues — que les sémiologues attachés à l'étude des lexiques documentaires, d'un point de vue formel.

Une observation de plus de conséquence touche à la manière dont sont présentés les principes de l'organisation paradigmatique du SYNTOL (pp. 88-97). L'accent est mis en fait sur *une* forme d'organisation particulière, d'allure hiérarchique (pp. 91-97), celle-là même qui commande la structure des lexiques spécialisés que l'on vient d'évoquer; et les indications préalables sur la double ouverture du modèle, « multidimensionnel » et « multivoque » (p. 90), ne suffisent sans doute pas à marquer ici encore le départ entre le formalisme général et ses différentes interprétations. Sur le premier, nous avons déjà été conduits à préciser que les « syntagmes » { Ri, *a*, *b* } d'où le SYNTOL tire son nom constituent en fait le mode d'expression commun aux deux axes de l'organisation du langage, syntagmatique et paradigmatique (*supra*, § 1). Il eût donc été logique de présenter celle-ci de la même manière que celle-là, c'est-à-dire en partant des caractérisations de R, où l'on retrouvait : (a) la binarité, à nouveau (les « relations analytiques » ou « sémantiques » ne peuvent être prises en compte qu'entre des termes du

(13) Voir *Le Syntol, Etude d'un système général de documentation automatique*, rapport EUR 423 f à la Communauté Européenne de l'Energie atomique, tome III (*Exemples de lexiques*), Presses académiques européennes, Bruxelles 1964. Dans ce volume sont présentés les lexiques mentionnés plus loin (pp. 98-9) sous les sigles P (Psycho-physiologie), Q (Psychologie) et T (Champ commun); les lexiques R/S (Sociologie/Ethnologie) et T (Nature et Techniques) ont fait l'objet d'une diffusion limitée, sous forme ronéotypée, par le Centre d'Analyse Documentaire pour l'Afrique Noire (Ecole des Hautes Etudes) et par la Section d'Automatique Documentaire (C.N.R.S.).

lexique pris deux à deux); (b) l'ouverture « multivoque » du modèle, en ce sens que tout terme *a* du lexique peut être mis en relation avec plusieurs autres, quelle que soit l'orientation de la relation : $\{ \vec{R}\ a, (b, c, d \ldots) \}$ et/ou $\{ \overleftarrow{R}\ a, (b, c, d \ldots) \}$; (c) l'ouverture « multidimensionnelle » enfin, l'interprétation de R pouvant varier d'un syntagme à l'autre : (Ri, *a, b*), (Rj, *a, c*), (Rk, *b, d*), etc. La démonstration pratique de l'universalité du modèle devient alors aisée : considérant une organisation sémantique quelconque, dans le domaine de la documentation (liste de descripteurs, classification, thesaurus, etc.), on observe qu'elle peut être complètement décrite ou « traduite » par un ensemble de syntagmes du type précédent, et caractérisée formellement par les propriétés — le plus souvent implicites — de R : nombre, nature, règles d'orientation, conditions d'emploi, etc.

Faute de place, nous devrons nous limiter à cette esquisse de reformulation, pour les pages relatives à la structure paradigmatique du SYNTOL (pp. 88-90), et renvoyer à des études plus récentes où les mêmes questions ont été reprises d'une manière plus détaillée (14).

7. Reste l'aspect non plus linguistique, mais logique du système, c'est-à-dire l'ensemble des moyens proposés pour manipuler les données précédentes — syntagmatiques et paradigmatiques — aux fins de la « recherche documentaire » sur calculateurs.

L'outil sans doute le plus neuf, sur ce plan, est celui des « commutations » automatiques exposées aux pp. 108-125, sous leur deux espèces : les « variations », par lesquelles on peut traiter par le même programme des ensembles de documents indexés selon des « états » plus ou moins élaborés du langage, du point de vue syntaxique en particulier; et les « modulations », qui permettent de contrôler l'expression des « questions » posées au corpus, par des relâchements successifs de contraintes linguistiques opérés automatiquement lorsque le nombre de « réponses » est inférieur à un certain seuil fixé par le demandeur.

(14) J. C. GARDIN, *Free classifications and faceted classifications*, dans *Classification Research*, Proceedings of the Second International Conference (FID/CR Committee on Classification Research), Elsinore, 14-18 sept. 1964, éd. P. Atherton, Copenhague, Munksgaard 1965, pp. 161-188; du même auteur, *Eléments d'un modèle pour la description des lexiques documentaires*, Bulletin des Bibliothèques de France, 11e année, n° 5, mai 1966, pp. 171-182.

Touchant les « modulations », certains malentendus semblent s'être fait jour, que cette nouvelle édition fournit l'occasion de dissiper. On s'est parfois étonné d'une contradiction apparente entre le principe conditionnel du mécanisme, et le fait qu'il ait été appliqué systématiquement, *a priori,* dans le traitement de toutes les « questions » considérées pour l'expérimentation du système (p. 180). En d'autres termes, l'objection était la suivante : plutôt que de formuler d'abord une question stricte (ou « question brute », dans notre terminologie, *cf.* p. 112), et d'en étendre ensuite le champ par des modulations, n'est-il pas à la fois plus simple et plus rationnel d'exprimer immédiatement la question élargie, et de faire ainsi l'économie du calcul des modulations en machine ? Tournée de cette façon, la critique est bien sûr irréprochable; mais elle tombe si l'on veut bien garder à l'esprit la fonction réelle des modulations, dans une exploitation concrète; trois remarques s'imposent à ce sujet. (a) La première est que la manière systématique d'appliquer les modulations, dans l'expérience relatée plus bas (chap. 5), tenait précisément au fait qu'il s'agissait d'une *expérience,* où l'on voulait éprouver le fonctionnement de ces opérations, du point de vue du langage et du programme tout à la fois. C'est la raison pour laquelle aucun mécanisme conditionnel n'avait alors été inclus dans le programme, comme il est dit à la page 129. (b) En second lieu, l'absence de critères conditionnels ne suffit pas à disqualifier le principe des modulations; certaines d'entre elles, en effet, doivent être comprises non seulement comme un moyen d'étendre le champ d'une recherche documentaire, *a posteriori* (c'est-à-dire lorsque l'automate constate que les critères conditionnels ne sont pas satisfaits), mais aussi comme la seule manière concise de désigner certains sous-graphes, dans l'organisation syntagmatique (représentations documentaires) ou paradigmatique (classifications lexicales). C'est le cas notamment de la « médiation » dans le premier cas, et de la « sommation » dans le second; tout lecteur peut s'en convaincre en essayant d'imaginer ce qu'il en coûterait d'avoir à énumérer explicitement dans une question non modulée les syntagmes ou les termes implicitement visés par ces deux opérations. (c) Des sous-graphes de ce genre interviennent non seulement dans la formulation des « questions » posées à l'automate, mais aussi dans la présentation des « réponses » fournies : on peut souhaiter par exemple que celles-ci soient ventilées par termes, dans un ordre correspondant à l'organisation du lexique, ou encore par types

de configuration syntaxique, etc. Le formalisme des modulations est dans ce cas encore un moyen commode d'exprimer sous une forme abrégée des conditions souvent complexes de tri, pour la phase finale d'*édition;* on trouvera dans l'exposé d'une version élargie du métalangage SYNTOL une illustration de cette démarche (15), qui n'avait pas été évoquée dans le présent ouvrage.

8. Cette allusion au métalangage SYNTOL nous conduit à une autre catégorie de commentaires, touchant à la nature des programmes présentés dans ce volume pour l'expérimentation du système (chap. 4). On a souligné avec raison que les symboles proposés pour le contrôle des opérations effectuées par l'ordinateur sur les expressions SYNTOL (pp. 133-7) ne faisaient que refléter certaines macro-opérations inhérentes à la recherche documentaire en général, et aux mécanismes modulatoires du SYNTOL en particulier, sans souci d'une adaptation aux contraintes du mode opératoire de l'automate; en d'autres termes, cet embryon de métalangage était orienté vers une catégorie de problèmes déterminés, plus qu'il ne l'était vers les modalités du traitement automatique, de telle sorte qu'il risquait de peser lourdement sur l'efficacité du programme général. L'objection est tout à fait valide, et sans doute confirmée aux yeux d'autrui par les chiffres de performance publiés ici même (pp. 147-150); mais elle est la conséquence d'un choix délibéré de notre part, qui n'a sans doute pas été suffisamment marqué. La première application du SYNTOL sur machine n'avait d'autre but qu'expérimental; nombre de fonctions du programme n'étaient elles-mêmes définies que par rapport à cette expérimentation (ex. : p. 132), et nous admettions alors que l'impératif premier était de favoriser une expression aussi « naturelle » que possible des questions du point de vue du demandeur, plutôt que de l'obliger à faire l'apprentissage d'un langage de programmation plus proche des exigences de la machine, et partant plus efficace. Ce point de vue a évidemment changé dès lors qu'il s'est agi par la suite de concevoir un programme d'exploitation réaliste, celui-là même que nous annoncions sous le nom de « SYNTOL B » (pp. 150-165). Les critiques du métalangage ne sauraient par conséquent ignorer aujourd'hui les éléments nouveaux apportés par la description du « programme élargi » (*ibid.*), dans un ouvrage plus récent (16). Ce n'est pas dire que les fonctions et la

(15) *Economie générale*, pp. 157-8 et pp. 169-175.
(16) *Ouv. cit.*, pp. 141-168 et pp. 196-208.

forme du métalangage plus évolué auquel nous nous sommes arrêtés (1967) suppriment l'objection que l'on évoquait : le compromis nécessaire entre les exigences pour le moment opposées de l'homme et de la machine, quant aux formes d'expression les plus agréables à chacun, n'est pas matière à des choix faciles, et il faut souhaiter au contraire que des divergences de vues continuent à s'exprimer sur ce sujet, plus nombreuses même que par le passé, pour favoriser le progrès de ces hypothétiques « DOLs » (Documentation Oriented Languages) dont les préfigurations sont encore si rares (17).

Des remarques du même genre pourraient s'appliquer à d'autres aspects de l'exploitation sur machine, qui affectent également l'efficacité du système : ainsi, la codification externe des données, et leur mode d'organisation interne en machine. Si ces deux points n'ont guère été débattus dans le présent ouvrage, c'est à nouveau parce que nous ne nous attachions guère aux considérations de performance; et il faut ici encore renvoyer le lecteur à l'ouvrage précité, où sont exposés d'une manière plus circonstanciée : a) des règles d'écriture assouplies, pour l'enregistrement des lexiques (18) et des représentations documentaires (19); b) les arguments locaux qui nous ont conduits à conserver une organisation « directe » de la mémoire, comme dans le programme décrit plus loin (pp. 145-6), malgré les avantages apparemment décisifs de l'organisation « inversée » (20).

9. L'ouvrage en question, plusieurs fois évoqué dans cette préface, n'est rien d'autre que le compte rendu de l'application-pilote annoncée en 1964 (*infra*, pp. 219-224). Son objet devait être, rappelons-le, de simuler le fonctionnement d'une « chaîne de fabrication documentaire » réelle, exploitée au moyen des outils logiques et linguistiques du SYNTOL, ceci à deux fins : (a) compléter sur le plan de l'organisation pratique des tâches (collecte, analyse, traitement et diffusion des informations) la définition d'un modèle jusqu'alors envisagé surtout d'un point de vue théorique (langages documentaires, système de programmation); (b) fournir les éléments d'un premier bilan économique, dans l'hypothèse de chaînes documentaires diversement intégrées, c'est-à-dire caractérisées chacune par une gamme de fabrications plus

(17) *Ouv. cit.*, pp. 251-254.
(18) *Ouv. cit.*, pp. 117-8.
(19) *Ouv. cit.*, pp. 150-1.
(20) *Ouv. cit.*, pp. 74-5 et 246-8.

ou moins étendue (*infra*, pp. 220-4). Tels sont en effet les deux aspects de l'« économie générale » dont traite l'ouvrage précité (21); et c'est à ce dernier que devra se reporter le lecteur curieux de connaître l'aboutissement de certains projets évoqués au chapitre 6 du présent volume, ou de vérifier la validité des anticipations économiques avancées au chapitre 7, en guise de conclusions (pp. 228-9).

La mécanisation dont il est fait état dans cette dernière étude est encore limitée aux opérations de la recherche documentaire *stricto sensu*, selon le parti exposé il y a quatre ans (*infra*, pp. 21-2); l'analyse elle-même, sous ses deux modalités principales (« résumé » et « indexation »), est encore laissée à la compétence des hommes, et non pas réduite à son tour aux calculs de l'automate. Une question risque donc de se poser au lecteur averti : les arguments que nous avancions en 1964 pour justifier cette restriction provisoire (*infra*, pp. 32-5) sont-ils encore valides aujourd'hui ? Et les recherches que nous évoquions, sur la mécanisation d'une indexation définie précisément comme le passage d'une langue naturelle à un langage documentaire conforme aux prescriptions du SYNTOL (p. 34) n'ont-elles pas au contraire déjà rendu anachronique la mécanisation partielle dont nous venons d'achever l'examen ?

Curieusement, nous serions tenté de répondre, d'une manière apparemment contradictoire, par l'affirmative à chacune des deux questions. En ce qui concerne tout d'abord la seconde, il est exact que l'on peut à présent faire état de résultats expérimentaux d'une qualité suffisante pour autoriser le projet d'applications documentaires où l'indexation serait elle-même prise en charge par la machine, sans que l'on ait à renoncer à la précision ou à l'intelligence habituellement attribuées à l'activité humaine équivalente (22). Cependant, là comme ailleurs, ces résultats ne sont

(21) *Economie générale*, pp. 3-9.

(22) Voir M. COYAUD et N. SIOT-DECAUVILLE, *L'analyse automatique des documents*, Mouton 1967, 2e Partie; à compléter par quatre rapports de la Section d'Automatique Documentaire (aujourd'hui Laboratoire d'Automatique Documentaire et Linguistique, C.N.R.S.), diffusés sous forme ronéotypée en 1966-7 — *Etudes sur l'indexation automatique*, par A. BORILLO, J. VIRBEL, etc. — et prochainement repris dans un ouvrage à paraître en 1969 sur le même sujet; voir aussi J. C. GARDIN, *Recherches sur l'indexation automatique des documents scientifiques*, Revue Française d'Informatique et de Recherche Opérationnelle, 1re année, n° 6 (1967, pp. 27-46).

obtenus qu'au prix d'« investissements intellectuels » assez notables, dont on ne voit pas que les techniciens de l'information scientifique aient encore accepté, en France du moins, le caractère inévitable : construction de lexiques ou de thesaurus spécialisés pour les domaines d'application retenus (voir *supra*, note 7), établissement des correspondances voulues entre les mots ou expressions du ou des langage(s) naturel(s) considéré(s) et les termes de tel ou tel lexique, recherche des cas — plus nombreux qu'on ne le croit généralement — de correspondances multivoques (ex. : mots apparaissant tantôt isolément, tantôt dans des expressions composées qui en altèrent le sens; homographes, etc.), choix des critères de résolution les plus efficaces pour chacun de ces cas particuliers, etc. (23). Limité à *un* champ d'application à la fois, ce genre d'étude ne dépasse pourtant pas les capacités d'une petite équipe de spécialistes informés à la fois de la terminologie propre à ce domaine et de la méthodologie générale de l'indexation automatique; il est à craindre néanmoins que plusieurs années ne passent avant qu'on ne se résigne à cette « lapalissade [selon laquelle]... la qualité des produits documentaires dépend de la richesse et de la précision des données qui servent à les fabriquer », fût-ce sur le calculateur le plus puissant... (24). Dans l'intervalle, il y aura place pour des applications partiellement mécaniques, du genre de celles que nous avons considérées plus haut, où le rôle de la machine est seulement de compiler les informations bibliographiques requises, sous quelque forme que ce soit, à partir de données documentaires fournies par des analystes humains (25).

(23) Des travaux étrangers récents mettent en évidence la nécessité d'outils linguistiques du même ordre, pour l'indexation automatique de documents scientifiques en langue anglaise (ex. : J. O'Connor, *Automatic subject recognition in scientific papers, an empirical study*, Journal of the Association for Computing Machinery, vol. 12, n° 4 (1964), pp. 490-515; G. Salton, voir les travaux cités à la note 8, etc.) aussi bien qu'en russe (ex. : D. G. Lakhuti, V. S. Tchernjavskij, *Ob algoritmečeskom razpoznavanij omonimov*, Naučnotekhničeskaja Informatsija, 1965, n° 1, pp. 46-7, etc.).

(24) *Economie générale*, p. 237.

(25) Toutes les entreprises de documentation automatique actuellement en fonction ou en projet — à l'exception de celles qui visent moins encore, à savoir la fabrication d'index à partir de titres ou de résumés bruts (procédés KWIC, KWIT, etc.) — appartiennent à cette catégorie : voir la liste-échantillon donnée plus haut, note 7, où les outils lexicographiques cités sont tous destinés au personnel chargé d'assurer le fonctionnement des systèmes documentaires, *partiellement* mécaniques par conséquent.

10. S'il en est ainsi, dira-t-on, comment se peut-il que les mécanismes finalement assez sommaires de la recherche documentaire n'aient pas déjà fait l'objet de modèles analogues à celui que nous commencions à ébaucher au début de l'actuelle décennie ? L'absence de références à des travaux de cet ordre dans la première présentation du SYNTOL traduisait notre ignorance d'un système documentaire à vocation tout aussi générale qui apparaissait en U.R.S.S. à peu près vers le même moment (26). Il est d'autant plus nécessaire de réparer aujourd'hui cette omission que le modèle linguistique proposé présente de fortes analogies avec les caractéristiques structurelles du SYNTOL : même distinction entre le plan paradigmatique et le plan syntagmatique de mise en relation des termes, même format binaire pour l'expression des relations différenciées sur chaque plan, etc. Ce n'est pas ici le lieu d'entreprendre une comparaison systématique des deux modèles; mais il faut au moins compléter cette référence, malencontreusement omise en 1964, par d'autres qui concernent des travaux plus récents, où les mêmes analogies se font jour, à l'insu, semble-t-il, des auteurs; citons par exemple le « Relational Data File » (27) et le « DEACON » (28), dont le dernier état fait apparaître le rôle central d'une unité d'expression syntagmatique diversement nommée « binary relational sentence » chez l'un, « triangle » chez l'autre, à propos de laquelle reparaissent la plupart des problèmes traités dans la première présentation du SYNTOL : conséquence de la réduction au format binaire, règles d'inférence ou de « développement », choix des relations disponibles sur le plan paradigmatique et syntagmatique, nature des macro-opérations utiles dans les graphes constitués sur chacun des deux plans, etc. Il n'est pas jusqu'à des études d'un ordre moins terre à terre,

(26) T. A. GRJAZNUKHINA, L. F. PSHENITČNAJA, E. F. SKOROKHOD'KO, *Sistema Informatsionnogo poiska*, Institut de Cybernétique, Académie Ukrainienne des Sciences, Kiev 1964.

(27) R. LEVIEN et M. E. MARON, *Relational Data File : a tool for mechanized inference execution and data retrieval*, Memorandum RM-4793-PR, The Rand Corporation, Santa Monica, déc. 1965; aussi *Relational Data File I : Design Philosophy; II : Implementation* dans *Proceedings of the Third National Colloquium on Information Retrieval*, mai 12-13, 1966, University of Pennsylvania.

(28) C. LONGYEAR, *Memory Structures in DEACON Natural Language questioning-answering systems*, presented at the Fourth Annual Meeting of the Association for Machine Translation and Computational Linguistics, 26-27 July, 1966, University of California, Los Angeles (document P-129, General Electric Co., Santa Barbara, California).

et détournées de tout souci « appliqué », où l'on retrouve les mêmes préoccupations de ce que l'on pourrait appeler le calcul sémantique appliqué à des réseaux formés de syntagmes binaires librement enchaînés (29); la convergence de ces travaux, nés dans des circonstances pourtant tout à fait séparées, est sans doute une justification raisonnable de la réédition présente.

J.-C. GARDIN
Marseille, le 3 janvier 1968

(29) G. H. LEECH, *Systems and Structures of Meaning, an Outline of a Semantic Theory* (sous presse), où l'unité fondamentale de la représentation est encore un triplet, appelé « prédication », dont les constituants sont deux éléments terminaux T1 et T2 liés par un terme médian M.

AVANT-PROPOS

La documentation automatique est depuis quelques années le sujet d'innombrables exposés, sous des titres divers : enregistrement et recherche des informations (de l'anglais, « Information Storage and Retrieval »), traitement automatique de l'information scientifique, automatique documentaire, etc. La manière de de l'aborder est tout aussi variée. Dans le cas le plus simple, l'on se borne à décrire l'emploi d'un procédé ou d'un *appareil* particulier, destiné à faciliter en pratique les opérations de tri caractéristiques de maints travaux documentaires. Ailleurs, c'est au contraire une « théorie générale » que l'on vise, où la mécanisation s'entend plutôt dans un sens figuré, comme l'observation des *mécanismes mentaux* qui sous-tendent l'activité documentaire, que celle-ci porte sur l'analyse ou sur la recherche rétrospective des informations contenues dans les textes scientifiques. Les exposés du premier type donnent de la documentation automatique une image indûment simplifiée, voire simpliste, qui conduit certains au vain espoir de résoudre par le seul emploi de machines et de la technique dite du « presse-bouton » toutes les imperfections de nos moyens actuels en matière d'information scientifique. Les seconds, inversement, risquent d'entretenir la thèse tout aussi erronée de l'impuissance des machines à exécuter les tâches éminemment « intelligentes » jusqu'ici confiées à des cerveaux humains, dans le même ordre d'activité. Entre ces deux points de vue extrêmes, il y a place pour des attitudes plus nuancées, où l'on envisage l'automatisation comme un partage, d'ailleurs mouvant, entre l'homme et la machine, dans l'accomplissement des opérations distinctives de la documentation.

Le « système » que nous présentons est un de ces compromis. En cette qualité il n'offre à nos yeux que l'intérêt d'un *exemple*,

et c'est à ce titre seul qu'il convient de le considérer. Nous soulignerons d'ailleurs dans les deux premiers chapitres, et à nouveau dans le dernier, les limites probables de sa validité. Il nous a néanmoins semblé que la description d'un système de documentation automatique réellement en usage pouvait illustrer, mieux qu'un manuel théorique, la nature des problèmes liés à une mécanisation même partielle des travaux documentaires. En ce sens, les solutions particulières que nous avons choisies importent moins peut-être que le dessin d'une problématique générale, construite à partir de ces observations concrètes.

Cette volonté pragmatique de l'exposé est la seule excuse que nous puissions donner de la manière ingrate de sa présentation. On n'y trouvera en effet que fort peu de matière à une lecture suivie, mais plutôt des indications à consulter, sur tel ou tel aspect particulier des méthodes proposées, pour l'analyse et la recherche automatique d'informations : règles linguistiques (chap. 3), mode d'exploitation sur calculateur (chap. 4), résultats expérimentaux (chap. 5), champ d'application (chap. 6). L'ouvrage n'a pas été conçu comme une « vue d'ensemble » des problèmes théoriques de la documentation automatique, logiquement ordonnés, mais au contraire comme une mosaïque de recettes dont l'intérêt est surtout qu'elles répondent à un souci pratique : permettre la recherche automatique d'informations, sur calculateurs. C'est d'ailleurs pour faciliter cette *consultation* de l'ouvrage, à défaut de sa lecture, que nous l'avons fait suivre d'un index relativement développé. Un glossaire, placé en fin de volume, rendra le même service, la terminologie employée ne correspondant à aucun usage bien établi, même hélas, pour les mots les plus courants (« analyse documentaire », « représentation », « mot-clé », etc.).

Les recherches sur le SYNTOL sont le fruit d'une coopération entre plusieurs organismes. Outre le Centre National de la Recherche Scientifique (Section d'Automatique Documentaire) et la Maison des Sciences de l'Homme (Centre de Calcul) — elle-même bénéficiaire d'une subvention de la Fondation Ford aux Etats-Unis — deux institutions ont bien voulu apporter une contribution financière à nos travaux, à savoir la Communauté Européenne de l'Energie Atomique (Bruxelles) et la Délégation Générale à la Recherche Scientifique et Technique (Paris). Enfin, plusieurs

collaboratrices ont été associées de façon régulière à l'expérimentation du SYNTOL : citons notamment Britta EISENREICH, Natacha GARDIN, Françoise IZARD et Radmila ZYGOURIS; qu'elles veuillent bien trouver ici l'expression de nos remerciements.

R. C. CROS (Section d'Automatique Documentaire, C.N.R.S.).
J. C. GARDIN (Section d'Automatique Documentaire, C.N.R.S.).
F. LEVY (Centre de Calcul, Maison des Sciences de l'Homme).

Les chapitres 4 et 5 sont dûs respectivement à R. C. CROS et F. LEVY, les autres à J. C. GARDIN.