

**BIBLIOTHÈQUE
DE L'AUTOMATICIEN**

15

410
C99

sous la direction de Pierre Naslin

A. DEWEZE

**traitement
de
l'information
linguistique**

F. Castellan



DUNOD

BIBLIOTHÈQUE DU CERIST

BIBLIOTHÈQUE DE L'AUTOMATICIEN

Sous la direction de P. NASLIN

1. — G. M. RUSSEL. — *Modulation et codage dans les systèmes automatiques.*
2. — J. A. ASELTINE. — *Les méthodes de transformation dans l'analyse des systèmes linéaires.*
3. — A. RALSTON et H. S. WILF. — *Méthodes mathématiques pour calculateurs arithmétiques.*
4. — J. BENEŠ. — *Dynamique statistique des circuits de régulation.*
5. — A. Ja. LERNER. — *Commandes optimales par commutation.*
6. — P. CASTELLO. — *Clé des schémas électriques (2^e édition).*
7. — P. LEFÈVRE. — *Optimalisation statistique des systèmes dynamiques.*
8. — R. E. BELLMAN et S. E. DREYFUS. — *La programmation dynamique et ses applications.*
9. — P. NASLIN. — *Circuits logiques et automatismes à séquences (2^e édition).*
10. — J. KUNTZMANN. — *Algèbre de Boole.*
11. — J. BRUNIN. — *Logique binaire et commutation.*
12. — Sheldon S. L. CHANG. — *Synthèse des systèmes de commande optimale.*
13. — P. NASLIN. — *Introduction à la commande optimale.*
14. — J. M. DOYLE. — *Impulsions et circuits impulsionnels.*
15. — A. DEWEZE. — *Traitement de l'information linguistique.*

Sous presse :

- J. LECORGUILLIER. — *Les vannes électroniques à électrodes de déclenchement et leurs applications.*
- A. E. ROGERS et T. W. CONNOLLY. — *Applications industrielles du calcul analogique.*
- H. SOUBIÈS-CAMY. — *Les techniques binaires et le traitement de l'information (2^e édition).*
- P. I. TCHINAIEV. — *Systèmes auto-adaptatifs.*
- L. N. VOLGUINE. — *Programmation des calculatrices numériques de commande.*
- D. B. WELBOURN. — *Précis de commande automatique pour ingénieurs mécaniciens.*

En préparation :

- J. HENRY-BAUDOT. — *Les moulins électriques en automatique appliquée.*
- J. KUNTZMANN et P. NASLIN. — *Algèbre de Boole et machines logiques.*
- J. LÛB. — *L'identification pratique des processus industriels.*
- R. MOLLE. — *Les composants hydrauliques et pneumatiques de l'automatique.*
- J. THILLIEZ. — *Commande numérique des machines.*

BIBLIOTHÈQUE DE L'AUTOMATICIEN
Sous la direction de P. NASLIN

TRAITEMENT DE L'INFORMATION LINGUISTIQUE

PAR L'HOMME
PAR LA MACHINE

PAR

A. DEWEZE

Chef du Service de Documentation
Établissements Merlin & Gerin, Grenoble

PRÉFACE DE

E. DELAVENAY

Agrégé de l'Université
Président-fondateur de l'Association de Traduction automatique
et de Linguistique appliquée



DUNOD
PARIS
1966

PRÉFACE

L'expression même de « machines à penser » donne souvent naissance à une double mystique — une pour et une contre — chez ceux qui, également mal informés, y voient soit un sacrilège, soit la paresseuse justification d'espérances démesurées. Le livre de M. DEWEZE, intellectuellement probe et ample-ment informé, inspiré d'une solide expérience des problèmes de la documen-tation, de la traduction humaine et automatique, de l'électronique et des calculatrices, contient, avec les éléments essentiels de la démystification, d'ex-cellents antidotes à la mystique des machines à penser. Il dresse un inventaire du possible, ouvrant des perspectives modestes mais sûres quant au dévelop-pement des procédés d'automatisation.

Dans ce que l'on désigne couramment du nom de pensée (mis à part le phénomène de la conscience de soi, fondamental, mais non absolument néces-saire dans tous les actes de l'esprit), quelle est au juste la part des coïncidences au niveau neuro-psychologique, et quelle est celle d'un processus *sui generis*, plus noble, caractéristique de l'esprit humain? Plus n'est besoin aujourd'hui de disputer pour savoir si les bêtes ont de l'esprit. On comprend mieux que naguère les mécanismes des associations d'idées, sinon de la volition affirma-trice du moi. La dispute s'est déplacée : les machines peuvent-elles avoir de l'intelligence, c'est-à-dire par exemple, restituer davantage d'information qu'on ne leur en a fourni à un moment donné? M. DEWEZE parvient à démon-trer que cette possibilité existe dans un cas particulier, celui d'un automate de recherche documentaire; cela sans jamais entrer dans le débat philosophique sur la « supériorité » de l'homme sur la machine, qu'il laisse aux philosophes. « Chaque fois », écrit-il, « que, au cours de cette étude, nous comparons les performances de l'homme et de la machine, nous le faisons à l'aide de quantités mesurables : vitesse de perception, capacité de mémoire, temps d'accès à un élément mémorisé, etc. »

Aussi trace-t-il entre les organismes biologiques et les machines électro-niques actuellement réalisées une ligne de démarcation bien nette, qui est celle de la faculté d'apprentissage et d'adaptation — dépendant elle-même de méca-nismes de régulation en retour à partir de la reconnaissance des formes visuelles ou sonores. Partant d'un modèle analogique de recherche bibliographique et aboutissant aux perceptions, il initie donc son lecteur, selon un plan d'ensemble judicieux, aux étapes successives de la simulation par les machines de certaines

formes d'intelligence, faisant chaque fois les comparaisons qui s'imposent avec les organes humains correspondants.

Cette exploration des divers aspects du traitement de l'information linguistique par l'homme et par la machine, résolument limitée à l'examen de problèmes concrets et précis, permet néanmoins de circonscrire nos connaissances actuelles et de définir les directions vers lesquelles s'oriente utilement la recherche de machines plus habiles, aux réactions plus souples que celles des premières calculatrices, douées de facultés plus rapprochées de celles de l'homme, donc capables de simuler de plus près un ensemble plus complet et plus varié de ces mécanismes de la pensée dont, grâce aux machines, on apprécie mieux aujourd'hui le fonctionnement.

Jusqu'à quel point des machines douées de mémoires associatives se révéleront-elles capables d'imiter certains processus intellectuels s'élevant au-dessus des simples coïncidences de signes, tels que l'abstraction et la généralisation? Que peut-on espérer de l'auto-apprentissage des machines? M. DEWEZE ouvre sur ces problèmes des perspectives volontairement sobres et réalistes. Il faut lui savoir gré d'avoir procédé à cet inventaire riche et complet des ressources que nous offre le traitement de l'information par des procédés automatiques de plus en plus complexes et dérivés du simple calcul mécanique.

E. DELAVENAY,

Agrégé de l'Université.

Président-fondateur de l'Association de Traduction automatique
et de Linguistique appliquée.

AVANT-PROPOS

La démarche intellectuelle suivie au cours de cette étude est empirique, en ce sens qu'il nous fallait d'abord définir un système documentaire capable de fonctionner dans un avenir proche, avec des équipements électroniques disponibles sur le marché.

Nous avons gardé constamment présente à l'esprit l'idée de définir un système capable de rendre des services immédiats pour la diffusion rapide et la recherche rétrospective de documents.

Mais, ce faisant, nous nous sommes efforcé de déterminer une méthode de codage et d'adressage, telle que l'information enregistrée puisse être utilisée à des tâches plus évolutives, pouvant aller jusqu'à l'exploitation des capacités heuristiques du modèle décrit, dont les résultats peuvent être interprétés par l'homme ou, dans une étape ultérieure, grâce à un auto-apprentissage de la machine.

Nous établissons nettement la distinction entre l'automatique documentaire réalisable avec les machines existantes et des mécanismes cognitifs plus évolués, auxquels la structure des calculatrices actuelles ne convient guère, dans la mesure où, pour établir les associations nécessaires, on a besoin d'un espace à n dimensions.

Si, au cours de cette étude, nous avons examiné les aspects linguistiques du traitement de l'information, c'est que le langage est le principal outil de communication entre les hommes.

Certains problèmes de linguistique, tels que la traduction automatique des langues, ne sont que des aspects particuliers du problème sémantique plus général qui est proposé à l'automatique documentaire.

Dans la deuxième partie de cette étude, nous montrons qu'un automate documentaire ou linguistique parfaitement adaptatif devrait être éduqué avec une interaction étroite entre les premier et deuxième systèmes de signalisation, au sens de Pavlov.

Après avoir étudié le problème de la reconnaissance des formes visuelles et sonores, par l'homme et par la machine, nous définissons la structure des dispositifs électroniques qui permettraient de réaliser un automate linguistique adaptatif.

Le lecteur, que certains détails technologiques, nécessaires en tant qu'éléments de démonstration, n'intéresseraient pas, peut, sans inconvénient, ne lire que l'introduction et les chapitres moins spécialisés. Il aura quand même une vue d'ensemble des problèmes et des voies indiquées pour leur solution, ainsi qu'une information la plus à jour possible sur l'état d'avancement des techniques et connaissances.

En ce sens, cet ouvrage s'adresse non seulement à l'ingénieur automaticien et au documentaliste, mais encore au linguiste, au phonéticien et au neurophysiologiste intéressés par les problèmes de simulation des mécanismes de la perception et des fonctions cérébrales supérieures.

Enfin, sans vouloir confondre un modèle avec la réalité, les lois d'associativité et d'auto-structuration énoncées ici sont applicables de façon plus générale à tous les messages utilisés par les organismes vivants, et à la notion même d'émergence et d'adaptation de la Vie.

TABLE DES MATIÈRES

INTRODUCTION.....	1
-------------------	---

PREMIÈRE PARTIE

AUTOMATIQUE DOCUMENTAIRE ET MÉCANISMES COGNITIFS

CHAPITRE 1. — Un modèle documentaire analogique.....	11
CHAPITRE 2. — Transformation par autoadressage du modèle analogique en un système discret.....	22
2. 1. Zone d'adresse. Matricule du document.....	24
2. 2. Fichier des classifications et facettes.....	25
2. 3. Fichier des citations. Adressage automatique.....	26
2. 4. Fichier des titres et lexiques organisés.....	27
2. 5. L'expression des relations entre classifications et citations.....	28
2. 6. Modulations du fichier des classifications au cours d'une recherche rétrospective.....	29
2. 7. Modulations du fichier des citations au cours d'une recherche rétrospective.....	30
2. 8. Modulations des fichiers de classifications et de citations sur bande magnétique.....	31
2. 9. Remarques sur l'immatriculation des documents.....	33
2.10. Exemples d'exploitation de l'information fournie par les citations.....	36
2.11. Classifications et structuration des messages.....	39
CHAPITRE 3. — Nécessité, conception et perspectives des mémoires associatives.....	42
3. 1. Le problème de l'adressage multiple et de l'accès direct à une zone de carte.....	42
3. 2. Une mémoire associative capacitive.....	44
3. 3. Une mémoire associative optoélectronique.....	47
3.3.1. Accès par la zone d'adresse et lecture d'une seule carte..	52
3.3.2. Accès par une zone quelconque et lecture de plusieurs cartes.....	52
3.3.3. Accès libre en longueur d'adresse variable et lecture de plusieurs cartes.....	53

3. 4. Perspectives de réalisation des dispositifs optoélectroniques..... 57
 3. 5. Perspectives d'utilisation des mémoires optoélectroniques associatives 59

CHAPITRE 4. — Comparaison des mémoires associatives avec la mémoire humaine 61

4. 1. Bilan informatif d'une mémoire associative statique..... 61
 4. 2. Matérialisation des souvenirs dans le cerveau 62
 4. 3. La mémoire humaine et l'adressage automatique 63
 4. 4. La vitesse d'acquisition et de traitement par le cerveau 64
 4. 5. L'apprentissage dans les organismes vivants..... 65

CHAPITRE 5. — Apprentissage et autoapprentissage des machines 68

CHAPITRE 6. — Enseignement mutuel de l'homme et de la machine 78

DEUXIÈME PARTIE

RECONNAISSANCE DES FORMES VISUELLES ET SONORES

CHAPITRE 7. — Notion de bruit et de reconnaissance d'une structure 85

CHAPITRE 8. — Quelques procédés de lecture automatique..... 94

8. 1. Distorsions et parasites des caractères 95
 8. 2. Les deux aspects de la lecture automatique : analyse et identification des caractères 96
 8. 3. Les systèmes à tête de lecture magnétique 98
 8. 4. Principe des procédés d'analyse optique 100
 8.4.1. Procédés à masque et à disque de Nipkow 100
 8.4.2. Procédés à matrice (ou rangée) de cellules photoélectriques..... 101
 8.4.3. Procédés d'analyse séquentielle par tube à rayons cathodiques 104
 8.4.4. Procédés d'analyse asservie 105
 8. 5. Les procédés d'identification 108
 8.5.1. Identification de critères externes de caractères normalisés ou identification d'un jeu restreint de caractères courants..... 108
 8.5.2. Identification logique d'après la position de critères dans une trame 112
 8.5.3. Repérage de critères topologiques (asservissement aux contours du caractère ou quantification angulaire) 121
 8.5.4. Identification de l'écriture manuscrite cursive..... 127
 8.5.5. Identification à l'aide d'un système analogique et quantifié..... 129

CHAPITRE 9. — Problèmes et procédés de traitement du langage parlé 135

9. 1. Analyse, reconnaissance et synthèse du langage parlé 135
 9. 2. La compression de bande et le Vocoder 137

BIBLIOTHEQUE DU CERIST

9. 3.	Caractéristiques du spectre de la parole	138
9. 4.	La classification binaire des phonèmes	141
9. 5.	La représentation bidimensionnelle des traces de formants	142
9. 6.	Les variations de distribution du spectre	144
9.6.1.	Les variations d'amplitude	144
9.6.2.	Les variations de la position des formants	145
9.6.3.	Les variations de la fondamentale	146
9. 7.	Les analyseurs et leurs possibilités actuelles de reconnaissance ...	147
9.7.1.	L'identification de phonèmes et le « Phonétographe » ...	148
9.7.2.	Les syllabes phonétiques et la « Phonetic Typewriter » ...	149
9.7.3.	L'identification de mots entiers prononcés isolément	153
9.7.4.	L'identification d'après d'autres critères	154
9.7.5.	Utilisation de l'information linguistique	156
9. 8.	La synthèse de la parole	157
9. 9.	Comparaison des procédés de reconnaissance des formes sonores ..	159
9.10.	Comparaison des perspectives d'application industrielle des techniques de reconnaissance automatique des formes visuelles et sonores	160
CHAPITRE 10. — Facultés et organes de reconnaissance dans les organismes biologiques		164
10.1.	L'aptitude des organismes biologiques aux opérations de reconnaissance	164
10.2.	Les aires gnosiques de l'écorce cérébrale	166
10.3.	Le langage et le conditionnement des réflexes	168
10.4.	Une étude expérimentale de la rétine corticale	170
CHAPITRE 11. — Les perceptrons et les matrices d'apprentissage		176
11.1.	Les perceptrons et leurs dérivés	176
11.2.	Un convertisseur de signaux sonores en signaux optiques : le Sceptron	184
11.3.	Les matrices d'apprentissage	187
11.4.	Associations de concepts et mécanismes d'invention	197
CHAPITRE 12. — Conclusions		202
GLOSSAIRE		213

INTRODUCTION

Notre conception de l'Univers dépend des connaissances que nous avons des processus de sa perception.

A. MOLES.

Depuis quelques années, on parle volontiers de traitement de l'information, sans préciser de quelle information il s'agit. Nous distinguerons :

— le *traitement de l'information* fournie par la *mesure de grandeurs physiques*. La machine reçoit l'information sur son environnement par des capteurs (microphones, thermistances, cellules photoélectriques) qui convertissent les signaux d'entrée en signaux électriques analogiques. Sur ces signaux, ou sur leur transformation en trains d'impulsions codées, on peut effectuer des calculs automatiques destinés à la régulation ou à la commande d'actions physiques.

— le *traitement de l'information mathématique* : calculs de systèmes physiques, statistiques économiques et démographiques, calcul opérationnel, etc. L'homme introduit les données dans la machine à l'aide de cartes ou de bandes perforées ou — plus récemment — à l'aide de lecteurs automatiques de caractères normalisés.

— l'homme reçoit l'information sur son environnement par l'intermédiaire de ses organes sensoriels. Ce sont également des capteurs qui convertissent les signaux extérieurs (optiques, acoustiques, thermiques) en trains d'impulsions codées envoyés aux aires réceptrices corticales. Les deux principaux canaux sensoriels (l'œil et l'oreille) transmettent au cerveau des images visuelles et sonores. L'homme a appris à schématiser ces images et à les manipuler de façon commode à l'aide des mots du langage (système de verbalisation ou second degré de signalisation de Pavlov). Le langage est le message verbalisé permettant à l'homme de communiquer avec ses semblables. La mathématique est un *langage symbolique* particulier dont la formalisation facilite le traitement par des machines. La linguistique a pour objet l'étude des *langages naturels*. *L'information linguistique* est susceptible d'être traitée à l'aide de machines construites par l'homme.

— la *transmission et la diffusion de l'information linguistique* par la télégraphie, la téléphonie, la radiodiffusion, la télévision et la presse sont des aspects particuliers du traitement de l'information linguistique. Le message est émis sous une forme par la source et perçu sous la même forme par le récepteur. Le message subit une double conversion d'ordre physique (par exemple : transformation d'oscillations sonores en oscillations électriques ou électromagnétiques par l'émetteur et nouvelle transformation en oscillations sonores par le récepteur. Mais le message ne subit pas de *traitement sémantique*.

Ce qui nous intéresse ici, ce sont les opérations effectuées par l'homme ou par la machine sur la *signification de l'information linguistique*.

L'homme a d'abord utilisé des signes idéographiques pour représenter les images verbales. Par la suite, l'écriture alphabétique lui a permis d'exprimer des concepts abstraits.

A l'heure actuelle, une grande partie des connaissances acquises par l'homme est consignée dans des documents divers : livres, films, disques, tableaux, etc. Les connaissances décrites à l'aide du langage verbalisé sont principalement enregistrées dans des documents écrits ou imprimés.

* * *

Dans la première partie de cette étude, nous ne considérerons que les documents imprimés : livres, revues, rapports techniques.

La signification d'un document imprimé est détectée par une série d'intégrations successives portant sur la signification d'unités linguistiques de dimensions croissantes et sur la signification des associations de ces unités à des niveaux de plus en plus complexes. Nous aurons, dans l'ordre :

- les lettres,
- les mots,
- les phrases,
- les ensembles de phrases.

La signification globale du document dépend de la signification des éléments considérés isolément à un certain niveau et de la signification qu'ils prennent lorsqu'ils sont associés à des éléments de même niveau. La signification du mot polysémique se colore selon l'éclairage de son contexte. La syntaxe introduit une signification au niveau de la phrase : une suite de mots, même si aucun d'eux n'est polysémique, n'est qu'une énumération susceptible de plusieurs interprétations.

L'ensemble des phrases est rendu cohérent par une suite logique dans les relations entre phrases. La logique de l'exposé constitue donc une sorte de syntaxe entre phrases, que le documentaliste doit détecter pour appréhender la signification globale du document afin d'en faire le résumé.

Aucune véritable analyse documentaire automatique ne pourra se faire tant que la machine ne saura pas reconnaître la signification des unités linguistiques à leurs divers niveaux.

Nous noterons, en passant, que la traduction automatique des langues n'est qu'un aspect particulier du traitement automatique de l'information linguistique. Lorsque la machine documentaire sera capable de détecter la signification d'un texte en langage naturel et de la transposer en son langage propre, le problème de la traduction automatique sera résolu. En effet, la même idée, les mêmes signifiés, exprimés par un auteur japonais, russe ou français se réduiraient, dans le langage interne de la machine, à un même ensemble de signifiants. La traduction automatique des langues deviendrait alors un problème secondaire, car il s'agirait seulement d'effectuer, à partir d'un ensemble de signifiants en langage machine, la synthèse d'un ensemble de signifiants dans le langage naturel-cible. Cette opération devient purement mécanique dès que la signification du message dans le langage naturel-source a été détectée par les procédés de l'automatique documentaire et traduite dans le langage intérieur de la machine.

* * *

Le modèle analogique décrit dans le premier chapitre permet d'enregistrer et d'exploiter les relations entre documents selon deux aspects :

— les relations de convergence et de divergence des documents, établies par les filières de citations bibliographiques et la modulation de ces relations dans le temps;

— les relations des documents d'après leur position dans plusieurs systèmes de classification et la possibilité de moduler les relations de classification afin d'établir entre les concepts des relations variables selon les points de vue.

Le choix du traitement de l'information à un niveau aussi élevé, compatible avec le traitement de l'information linguistique à d'autres niveaux, répond à deux objectifs :

— la définition d'un système documentaire basé sur ce principe et décrit dans le chapitre 2.

— la commodité d'illustrer le principe et l'exploitation du modèle décrit. Ce second objectif a été atteint par l'exemple des classifications documentaires, des coefficients de relation entre classes et concepts (sans lesquelles il n'y a pas de connaissance possible) et des matricules de documents illustrant le principe de l'auto-adressage décrit dans les chapitres 2 et 3.

Mais les coefficients de relation et les diverses modulations qu'ils permettent entre documents et classes de documents pourraient être utilisés pour exprimer la syntaxe des relations sémantiques entre phrases ou entre mots de lexiques organisés. Nous verrons dans les chapitres suivants que le modèle associatif décrit est utilisable à tous les niveaux du traitement de l'information linguistique.

La réalisation et la mise à jour d'un modèle analogique serait extrêmement coûteuse. Dans le deuxième chapitre nous examinerons comment remplacer les branchements et les liaisons du modèle analogique par des notations discrètes : adressage automatique à l'aide des matricules alphanumériques des documents et expression des relations par des coefficients numériques pouvant être modulés par des dispositifs à seuil. Le modèle discret peut être réalisé, tenu à jour et exploité automatiquement à l'aide de calculatrices électroniques classiques.

* * *

Les mémoires utilisées dans les calculatrices électroniques ne permettent pas des associations suffisamment rapides. Pour le traitement de l'information linguistique, il faudrait réaliser des mémoires associatives décrites dans le chapitre 3.

La souplesse d'association de ces mémoires les rend particulièrement utiles pour l'analyse prédictive qui permet de tirer parti des propriétés statistiques du langage.

A l'échelle microscopique du traitement linguistique (reconnaissance automatique de phonèmes ou de lettres) on peut améliorer les résultats en utilisant les probabilités des digrammes et des trigrammes.

A l'échelle macroscopique du traitement linguistique (analyse morphologique des mots et syntaxique de la phrase et analyse sémantique à ces deux niveaux) les probabilités liées permettraient de déterminer les enchaînements les plus vraisemblables entre les catégories sémantiques et grammaticales des mots.

La mémoire associative permanente décrite dans le chapitre 3 est intéressante par sa rapidité et sa sûreté d'accès à une très grande quantité d'information. Ce type de mémoire peut-il restituer une quantité d'information supérieure à celle que l'homme y a introduite? En un certain sens, oui, car l'homme est incapable de prévoir tous les cheminements possibles par association. Comme, par définition, la mesure de la quantité d'information est la mesure de l'imprévisible, on peut extraire du modèle une *information nouvelle*. Encore faut-il savoir poser les questions pertinentes permettant de les retrouver.

* * *

Dans le chapitre 4 nous comparons les performances des mémoires associatives avec la vitesse d'acquisition, d'accès et de traitement de l'information linguistique par le cerveau humain. Précisons immédiatement notre propos : le débat sur la supériorité de l'homme ou de la machine ne nous intéresse pas, il est l'affaire des philosophes qui utilisent pour cela des méthodes spéculatives qui leur sont propres.

Chaque fois que, au cours de cette étude, nous comparons les performances de l'homme et de la machine nous le faisons à l'aide de quantités mesurables :

vitesse de perception, capacité de mémoire, temps d'accès à un élément mémorisé, etc.

On ne peut *démontrer* les mécanismes du langage (et, par conséquent, ceux de la pensée, dans la mesure où le langage est l'un des principaux moyens de communication de la pensée) qu'à l'aide d'un *modèle physique* apte à les *simuler*.

Les êtres vivants (organismes biologiques) sont doués de facultés d'apprentissage et d'adaptation que ne possèdent pas les grandes calculatrices électroniques réalisées à ce jour.

* * *

Dans le chapitre 5, nous montrons que ces machines sont incapables d'établir elles-mêmes des règles, sinon dans les conditions dictées par l'homme : elles n'ont pas *besoin d'apprendre*.

Le véritable apprentissage repose sur la faculté innée d'établir des règles en fonction d'expériences antérieures, suivies d'une information en retour sanctionnant la réussite ou l'échec. Nous décrivons un modèle adaptatif faisant son propre apprentissage. Ce modèle est réalisable moyennant le remplacement de certaines structures permanentes de la mémoire associative décrite dans le chapitre 3 par des dispositifs électroniques ou optroniques qui ont déjà été réalisés à d'autres fins.

Les calculatrices électroniques utilisées actuellement ont été mises au point dans le but de gagner du temps dans les opérations de calcul. Leurs mérites sont incontestables et elles ont permis des recherches théoriques que, sans elles, l'homme n'aurait jamais pu entreprendre, faute de temps pour effectuer certains calculs.

Ces machines sont relativement faciles à réaliser et présentent des garanties de rentabilité immédiate pour les utilisateurs, auxquels elles apparaissent comme des « adultes » dès leur sortie d'usine.

Au contraire, les mécanismes cognitifs décrits ici et réalisables par l'homme ne seront peut-être jamais rentables et leur mise au point sera le fruit d'une longue patience.

* * *

Pour passer du langage naturel au langage interne des machines mathématiques utilisées à ce jour, l'homme utilise des langages de programmation. Pour les calculs scientifiques (Algol) ou les calculs de gestion (Cobol), ces langages tendent de plus en plus à permettre à l'homme de rédiger les instructions dans un langage naturel simple, l'organe de programmation de la machine faisant les transpositions nécessaires. De tels langages ont également été étudiés en vue de leur adaptation spécialisée à l'automatique documentaire (W.R.U., Syntol).

La réalisation de mécanismes cognitifs échappant aux déterminismes de l'information introduite par l'homme dans leur structure verra s'établir une

nouvelle forme de dialogue entre l'homme et la machine. L'homme pourra alors, non seulement utiliser les machines pour enseigner à des humains, comme le veut la pédagogie cybernétique, mais tirer la leçon de l'enseignement mutuel que se communiqueront l'homme et le mécanisme cognitif. Cette perspective est examinée dans le chapitre 6, à la fin de la première partie.

* *
* *

L'essentiel du véritable apprentissage repose sur la faculté d'établir, dans une situation donnée, des règles optimales de décision en fonction de décisions antérieures face à une situation comparable. Pour cela, il faut que la première décision soit suivie d'une information en retour sanctionnant la réussite (récompense simulée par une facilitation d'accès ultérieur) ou l'échec (punition simulée par une inhibition). On peut montrer que, en ce sens, il n'y a pas différence fondamentale entre la possibilité de conditionnements à partir d'excitants absolus du premier système, et le conditionnement à partir de mots qui peuvent n'avoir aucun équivalent dans le premier système, tels que les mots abstraits : Amour, Justice, Démocratie (ou des équivalents avilis : Palais de Justice, etc.).

Mais il est plus facile et exact de sanctionner les résultats d'un apprentissage sur des formes objectives (visuelles ou sonores) que sur des images verbalisées (mots ou phrases) ou sur des associations entre ces images (système documentaire).

La deuxième partie de cette étude est précisément consacrée à l'examen des problèmes de reconnaissance automatique des formes visuelles et sonores par la machine.

Dans le chapitre 7, nous définirons la *forme* ou *structure*, à partir de la notion de *bruit* et de la convention préalable nécessaire à toute *reconnaissance* et *désignation d'une forme*. Au cours de ce chapitre, nous insistons particulièrement sur les diverses opérations de la reconnaissance : analyse, schématisation, écart par rapport au schéma minimal, restructuration, généralisation et établissement de catégories nouvelles.

Dans les chapitres 8 et 9 nous examinerons les équipements de reconnaissance automatique réalisés industriellement ou en cours d'expérimentation. Nous limitons d'ailleurs notre propos aux formes du langage parlé (phonèmes ou leurs assemblages) et écrit (lettres), mais ceci n'est qu'une question de dimensions du catalogue des associations de formes élémentaires.

Mais les mécanismes biologiques de la perception, dont nous donnons un aperçu dans le chapitre 10, laissent loin derrière eux les dispositifs électroniques décrits dans les chapitres 8 et 9. La supériorité actuelle des aptitudes des organismes biologiques à la reconnaissance peut être mise sur le compte de la très grande richesse des neurones et de la complexité de leurs associations.

Dans le chapitre 11, nous examinons les dispositifs (perceptrons et matrices d'apprentissage) qui, par leur richesse d'interconnexions, seraient plus aptes à l'auto-apprentissage nécessaire à la reconnaissance des formes et à l'établisse-

ment de catégories nouvelles, que les systèmes électroniques, décrits dans les chapitres précédents, mais réalisés à des fins industrielles strictement utilitaires.

* *

Nous ne pouvons et ne voulons prédire l'utilisation qui pourrait être faite de mécanismes cognitifs et heuristiques.

Certaines applications, telles que la lecture automatique de textes imprimés, le codage des phonèmes, la traduction automatique et l'analyse automatique de documents scientifiques présentent un intérêt économique certain.

D'autres recherches pourraient n'avoir pour seul résultat que d'apprendre à l'homme à mieux connaître ses propres mécanismes perceptifs et cognitifs.

Il est possible que les méthodes actuelles de traduction automatique permettent d'obtenir économiquement, dans un avenir raisonnable, des traductions bilingues de bonne qualité. Mais nous ne pensons pas que cette qualité soit perfectible sans l'utilisation de mécanismes cognitifs capables d'appréhender la signification du message traité et ce, dans les deux systèmes de signalisation.

Nous ne savons pas combien de temps il faudrait pour réaliser un modèle cognitif capable de s'adapter à des situations dont un enfant de deux ans se tire à merveille. Rien ne permet d'affirmer ou de nier que nous soyons capables de réaliser un tel modèle. Seule une expérience fort prolongée pourrait nous l'apprendre. Mais combien de centaines de millions d'années la nature a-t-elle mis pour extraire, des profondeurs des premiers océans, le merveilleux enfant que nous venons d'évoquer?

* *

Chacun des chapitres est suivi d'une bibliographie sommaire référant à des ouvrages de base à la portée des lecteurs non spécialisés et nous avons regroupé les termes spécialisés dans un glossaire à la fin de l'ouvrage. Cette étude est donc accessible à tous ceux qui s'intéressent aux problèmes du traitement de l'information linguistique. La bibliographie mentionne également les ouvrages les plus récents, grâce auxquels nous avons tenté, au cours de notre exposé, de faire le point des connaissances dans les divers domaines abordés.

* *

Nous avons examiné le traitement du langage à deux niveaux extrêmes :

- au niveau de la lettre et du phonème;
- au niveau de la signification des documents.

Pour être complète, cette étude aurait dû aborder les problèmes de traitement au niveau du mot et de la phrase. La description formelle des langues

naturelles appartient aux linguistes et aux spécialistes de la traduction automatique. Nous sommes persuadé que la description linguistique, rendue obligatoire pour l'élaboration de bons programmes de traduction automatique, permettra de mieux connaître les mécanismes de la pensée.

La transition pourrait alors s'établir entre les dispositifs de reconnaissance des formes et le niveau supérieur de l'analyse automatique de la signification du message linguistique à l'aide de mécanismes cognitifs. A la fin de cet ouvrage, nous examinerons quel type d'automate pourrait assurer cette transition.

BIBLIOGRAPHIE SOMMAIRE*

- V. BELEVITCH : « Langage des machines et langage humain », Hermann (Paris, 1959).
 L. COUFFIGNAL : « La Cybernétique », Que Sais-je?, n° 638, P.U.F (Paris, 1963).
 E. DELAVENAY : « La machine à traduire », « Que sais-je? », n° 834, P.U.F. (Paris, 1959).
 JAGLOM et JAGLOM : « Probabilités et Information », Dunod, (Paris, 1959).
 A. MOLES : « Théorie de l'information et Perception esthétique », Flammarion (Paris, 1958).
 A. MOLES, B. VALLANCIEN : « Communications et Langages », Gauthier-Villars (Paris, 1963).
 J. POYEN, J. POYEN : « Le langage électronique », « Que sais-je? », n° 900, P.U.F., (Paris, 1960).
 C. E. SHANNON : « A mathematical Theory of Communication », *Bell System Tech. Journal*, 1948, p. 379-423, 623-656.
 N. WIENER, « The human Use of human Beings », University Press (Cambridge, 1950).
 J. C. GARDIN et alt. : « Le Syntol », Gauthier-Villars (Paris, 1964).
 B. C. VICKERY : « Techniques modernes de documentation. Analyse des systèmes de recherches de documents », Dunod (Paris, 1962).

(*) Une bibliographie plus spécialisée est indiquée à la suite de chaque chapitre.