

GEORGES

VIGNAUX

LES SCIENCES COGNITIVES

une introduction



Le
LIVRI
de
POCHI

bibli
essai

821 DL-01001974-14842

LES SCIENCES COGNITIVES
UNE INTRODUCTION

DU MÊME AUTEUR
Les Sciences cognitives
L'argumentation - Essai de logique descriptive - Paris, 1978
Le Discours et son statut cognitif - Paris, 1982
L'information et la perception - Paris, la direction de
G. VIGNAUX et K. FALL, Presses de l'université du Québec
Montréal, 1989



16° Z
24820
(193)

LA DÉCOUVERTE

DU MÊME AUTEUR

L'Argumentation. Essai d'une logique discursive, Droz, Genève, 1976.

Le Discours acteur du monde, Orphrys, Paris, 1988.

L'Informatique en perspective (sous la direction de G. VIGNAUX et K. FALL), Presses de l'université du Québec, Montréal, 1989.

7755778

7

GEORGES VIGNAUX

Introduction

LES SCIENCES COGNITIVES
EN NOUVEL ESPACE D'ESQUISSE

Les Sciences cognitives

Une introduction

Les sciences cognitives (1950), sont des sciences qui se sont pour objet de décrire, d'expliquer et de comprendre, de saisir les principales dispositions et capacités de l'esprit humain — langage, raisonnement, perception, connaissances motrices, planification... —. Définition à laquelle nous adhérons, mais qui s'est par là même présentée un double inconvénient : celui, d'abord, d'inciter à penser en termes « totalitaires » des sciences dites cognitives — « une nouvelle science de l'esprit » —, celui ensuite d'encourager à croire qu'il n'y avait et que de « scientifiques », voire de « dévotés », et non de « philosophiques ». En vérité, la construction se trouve plus que propre à certaines parties plus qu'à la « science » de l'homme et de son monde, qu'on peut reconnaître en ses spécificités, et de ses conditions.

Un premier exemple de ces parties plus — impliquant en l'occurrence — est de faire comprendre le raisonnement des sciences cognitives, une situation où il y a une « chose », dans un contexte particulier, la situation de l'intention, que ce soit la première des sciences de l'homme, celle de l'intention, ou la situation d'une certaine « intention » qui consiste de marquer les sciences cognitives et dont l'avant dire la possibilité, mais d'où nous nous dirigeons difficile à les définir, comme le souligne Auerbach, mais de manière « extensive ».

1391

LA DÉCOUVERTE

Les sciences cognitives (1950) sont des sciences qui se sont pour objet de décrire, d'expliquer et de comprendre, de saisir les principales dispositions et capacités de l'esprit humain — langage, raisonnement, perception, connaissances motrices, planification... —. Définition à laquelle nous adhérons, mais qui s'est par là même présentée un double inconvénient : celui, d'abord, d'inciter à penser en termes « totalitaires » des sciences dites cognitives — « une nouvelle science de l'esprit » —, celui ensuite d'encourager à croire qu'il n'y avait et que de « scientifiques », voire de « dévotés », et non de « philosophiques ». En vérité, la construction se trouve plus que propre à certaines parties plus qu'à la « science » de l'homme et de son monde, qu'on peut reconnaître en ses spécificités, et de ses conditions.

DL-0061994-14048 757/6/CT3

GEORGES VIGNAUX

Les sciences cognitives

Genève, 1976.

The introduction

G. VIGNAUX et K. PAUL, Professeurs de l'Université de Québec, Montréal, 1989.

1989

VERTE



© Éditions La Découverte, Paris, 1991, pour l'édition française.

Introduction

LES SCIENCES COGNITIVES, UN NOUVEL ESPACE D'ENJEUX...

La genèse historique des sciences cognitives

Les sciences cognitives, selon D. Andler¹ (1989), sont ces sciences qui « ont pour objet de décrire, d'expliquer et, le cas échéant, de simuler les principales dispositions et capacités de l'esprit humain — langage, raisonnement, perception, coordination motrice, planification... ». Définition à laquelle nous adhérons, mais qui n'est pas sans présenter un double inconvénient : celui, d'abord, d'inciter à penser au caractère « totalisant » des sciences dites cognitives — « une nouvelle science de l'esprit » ? —, celui ensuite d'incliner à croire qu'il n'y aurait là que du « scientifique », voire du descriptif, et non du « philosophique ». En vérité, la conjoncture se révèle plus que propice à certains partis pris quant à la « nature » de l'humain et des modèles qu'on peut construire de ses spécificités ou de ses conduites.

Un premier exemple de ces partis pris — moderniste en l'occurrence — est de faire remonter la naissance des sciences cognitives aux alentours d'il y a trente-cinq ans, dans un contexte particulier, déterminé par la naissance de l'informatique et les premiers développements des techniques de traitement de l'information. D'où effectivement une certaine « mémoire » qui continue de marquer les sciences cognitives et dont l'avenir dira la durabilité, mais d'où aussi cette double difficulté à les définir, comme le souligne Andler : tantôt de manière « extensionnelle » (selon leurs objets d'études) tantôt,

1. Les références des travaux et ouvrages cités dans l'ensemble du livre sont réunies dans la bibliographie (organisée par chapitre) figurant en fin de volume.

en revanche, de façon « intensionnelle », en regard des options théoriques qu'elles avancent, et à l'occasion, sous forme réductrice quant à nos capacités humaines. C'est dire qu'en la circonstance toute définition se voit d'emblée problématisée, mais sans doute est-ce là l'enjeu passionnant de cet espace stimulant de débats et réflexions que constituent aujourd'hui les sciences cognitives.

En résumé, bien qu'en vérité il s'agisse souvent d'interrogations sur l'homme qui remontent aux tout débuts de la philosophie, on s'entend à n'accorder à ces « nouvelles » sciences qu'une histoire relativement récente (Gardner, 1985). La datation d'origine serait vers la fin des années trente puis des années quarante, lorsque, précisément, le logicien A. Turing va concevoir, en 1936, le premier modèle mathématique de l'ordinateur programmable, puis vers 1950, reformuler, cette fois d'un point de vue philosophique, cette conception d'une « machine intelligente » (Hodges, 1988). Toute cette période voit l'apogée de la logique mathématique, laquelle va construire les concepts et les outils nécessaires aux nouveaux calculs. C'est alors la naissance, vers 1943, de la *première cybernétique*, expression de ce projet de réduction matérialiste sinon mécaniste du mental au physique (Andler, 1989).

Cette première cybernétique, impulsée notamment par N. Wiener et W. McCulloch, s'est développée de 1946 à 1953, à l'occasion d'une dizaine de sessions interdisciplinaires (les conférences Macy) auxquelles participèrent aussi bien des mathématiciens, des psychologues, des physiologistes que des sociologues, des linguistes et des anthropologues. Cela a sans doute orienté considérablement le développement des sciences cognitives, au point que perdurent aujourd'hui certains effets conceptuels des modélisations alors conçues. Plusieurs colloques fameux eurent un impact déterminant, dont celui de New York (« Teleological Mechanisms », 1946) et le Hixon Symposium de 1948. Tout ce mouvement d'idées et travaux fut surtout soutenu durant près de vingt ans, depuis le laboratoire que McCulloch fonda et dirigea au MIT jusqu'à sa mort en 1969. C'est à cette époque, enfin, que von Neumann conquiert sa réputation en développant la *théorie des automates*, contribuant de façon capitale autant aux futures conceptions d'ordinateurs

qu'aux diverses théorisations sur la nature des processus mentaux.

La naissance des sciences cognitives

Les sciences cognitives, de là, vont naître rapidement. Pour nombre d'historiens-épistémologues du mouvement, la date symbolique est celle de 1956, année où, d'une part, se tient le « Symposium on Information Theory » du MIT, rassemblant psychologues et linguistes intéressés à intégrer leurs travaux en vue d'une simulation des processus cognitifs sur ordinateur, et, d'autre part, est organisée cette « rencontre de Darmouth » où sera officiellement déclarée la naissance de « l'intelligence artificielle ». A partir de là, l'objet « cognition » est né, et différentes disciplines vont s'efforcer de lui attribuer des contenus et des orientations spécifiques.

En 1960, les psychologues J. Bruner et G. Miller prennent ainsi l'initiative de fonder à Harvard, un Center for Cognitive Studies qui va s'illustrer pendant au moins une dizaine d'années. Cette même année paraît l'ouvrage de Miller, Pribam et Galanter, *Plans and the Structure of Behavior*, mettant en cause le béhaviorisme classique des travaux psychologiques et proposant de lui substituer l'approche cybernétique. Un peu plus tard, en 1967, U. Neisser publie son ouvrage, *Cognitive Psychology*, plaidant en faveur d'une nouvelle approche psychologique, différente cependant de cette approche cybernétique, défendue notamment par Newell et Simon pour qui l'ordinateur peut fournir le modèle des fonctionnements de l'esprit humain.

Les années soixante-dix sont celles de la promotion, un peu partout, de centres interdisciplinaires, de revues, de congrès et de travaux donnant corps à l'appellation « sciences cognitives ». Un réseau de chercheurs et d'échanges internationaux se met alors en place. De là, diverses évolutions vont s'opérer, qui prendront forme dans les années quatre-vingt. L'intelligence artificielle, qui semblait devoir être la référence dominante, va perdre de son importance en regard des neurosciences bien sûr, mais aussi au profit de nouvelles représentations des phénomènes humains, issues des divers

concours de la philosophie, de la sémantique cognitive ou de l'anthropologie.

Ces mouvements peuvent encore laisser croire aux sceptiques, attachés aux fonctionnements disciplinaires classiques, que les études cognitives ne seront guère plus qu'une mode, une étape dans des restructurations des savoirs. Il n'en est rien : les affrontements visibles sont aussi indices de nombreux échanges invisibles : nous sommes là en présence d'une « révolution » au sens copernicien dans les formulations de nos savoirs et de nos méthodes, mais dont il est bien sûr difficile aujourd'hui de prévoir toutes les conséquences. La meilleure façon de le comprendre est de s'en assurer au travers des progrès accomplis par les différentes disciplines concernées et des convergences ou divergences dont ces évolutions peuvent témoigner.

Les disciplines des sciences cognitives

Quelles disciplines peut-on considérer comme directement impliquées dans les sciences cognitives ? Assurément, il s'agit des neurosciences, de l'intelligence artificielle, de la philosophie, de la psychologie et de la linguistique. Qu'est-ce qui peut ainsi les réunir ? C'est à la fois la préoccupation des rapports esprit/cerveau (*mind/brain*) et celle des modélisations possibles de ce rapport, mais aussi l'analyse des fonctionnements ainsi impliqués ou des conduites par là dérivées. Plus globalement, sans doute s'agit-il de construire une nouvelle « science » des phénomènes constitutifs de nos appareils psycho-biologiques et des interactions entre ces appareils et nos comportements y compris sous leurs formes hautement symboliques telles que les langages et les cultures.

Un paradoxe est à souligner à ce propos — comme le fait D. Andler —, c'est celui de l'anthropologie : impliquée dès les premières conceptualisations historiques des sciences cognitives, celle-ci semble aujourd'hui reléguée à un rôle secondaire ; ce qui peut surprendre s'agissant ici de construire une science des spécificités de l'humain dans ce qu'elles ont d'universel au travers des variations. Un tel « oubli » tôt ou tard sera réparé en même temps sans doute que seront reconnues les contribu-

tions importantes de la psychologie sociale, assurée de faire passerelle entre l'« individuel » et le « collectif » ou de la linguistique, à condition que celle-ci intègre pleinement les préoccupations du rapport entre langage, sens et expressions culturelles.

Dans l'immédiat, il importe bien sûr de poursuivre la recherche en direction du développement de modèles aussi bien logico-mathématiques — au moins pour se constituer des outils —, que neuro- et physico-sensoriels — afin de mieux comprendre à la fois les fonctions internes et les manifestations externes de nos « équipements de base ».

Les recherches sur la cognition

C'est pourquoi le classement des recherches en cognition peut s'opérer de façon tripartite (Imbert, 1987). Il y a tout d'abord une première grande famille d'études des *aptitudes ou compétences cognitives* au sens classique du terme : le raisonnement, le langage, la perception et l'action. Mais là où, autrefois, on n'observait que les réalisations extérieures de ces grandes fonctions tout en les considérant comme corrélées au sein de l'organisme et dans les différentes classes de situations, on tend de plus en plus aujourd'hui à les étudier chacune spécifiquement et à différents niveaux relativement indépendants les uns des autres.

De là vient que, d'un côté — c'est la deuxième famille de recherches cognitives —, on s'attachera à spécifier les constitutions et réalisations matérielles de différents types de *mécanismes*, neurophysiologiques (en biologie) ou électroniques voire mécaniques (en intelligence artificielle) ; et, d'un autre côté — la troisième famille —, on s'efforcera de caractériser fonctionnellement les mises en œuvre de ces mécanismes jusqu'à pouvoir les décrire sous forme de processus décomposables en *opérations élémentaires*, puis les modéliser en termes de propriétés formalisables.

Dans ce troisième et dernier cas apparaissent des clivages significatifs entre recherches, selon que les unes vont s'attacher à observer la structure d'*état initial* de nos équipements — maturation nerveuse et capacités —, en reliant cet « état »

structurel à nos différentes aptitudes à l'apprentissage, tandis que d'autres études porteront sur l'analyse des différentes formes d'*état final* que vont prendre nos systèmes naturels, après apprentissage et acquisition de capacités fonctionnelles spécifiques. Les difficultés demeurent quant aux moyens de relier correctement ces états origines à ces états finaux. Ces difficultés ne sont pas moins grandes lorsqu'il s'agit d'appliquer la même approche évolutive en intelligence artificielle et de savoir dissocier les « états initiaux » des « états finaux » de la machine (*machine learning*).

Ce rangement tripartite des recherches en sciences cognitives semble satisfaisant tant qu'on ne se préoccupe pas de savoir où placer les recherches pourtant bien établies et fructueuses que la psychologie générale et la psychologie sociale ont développées depuis longtemps en ce qui concerne les processus de mémoire, d'attention ou les phénomènes de représentation sociale. Il en va de même de nombre de travaux atypiques qui s'inscrivent dans la perspective « connexionniste » (Andler, 1989).

Cette difficulté à classer les problématiques en cours s'accroît encore lorsqu'il s'agit de les départager selon les hypothèses que les uns et les autres vont appliquer au « système » cognitif humain et selon les choix stratégiques et méthodologiques qui vont en découler. Les débats sont nombreux et vifs, les affrontements, d'écoles.

D. Andler propose un schéma regroupant les trois *hypothèses ontologiques* les plus communément répandues ; schéma qui offre le mérite de la claire simplicité :

1. La description et l'explication à un niveau purement physicaliste (bio-chimico-physique) des phénomènes cognitifs se révèlent insuffisantes et doivent donc être complétées par un *niveau représentationnel* : « Les états des systèmes physiques considérés représentent des *informations* et c'est dans cette mesure qu'ils sont, outre des états physiques, des états cognitifs » ;

2. Les transformations par lesquelles vont passer ces états ne sont pas seulement alors physiques, mais peuvent être considérées comme des *calculs sur les représentations* dont ces états sont porteurs ;

3. Si tout phénomène cognitif peut se définir comme proces-

sus établissant relation entre un stimulus et la réaction ou réponse à ce stimulus, on ne peut limiter l'analyse et la définition à chaque fois aux effets observés de ce processus ; en vérité, l'essentiel de tout processus se fonde sur ce qui se passe dans l'intervalle séparant le stimulus de la réponse et il importe de le définir le plus relativement possible vis-à-vis de ses effets ; ce qui veut dire qu'il y a toujours *autonomie des processus internes* en regard de ce que laisserait supposer la simple considération des phénomènes qui les suscitent ou qu'ils engendrent.

Les grands choix méthodologiques et épistémologiques

Ces hypothèses vont susciter des choix méthodologiques et épistémologiques non négligeables :

1. Les phénomènes cognitifs peuvent être repérés, décrits et analysés voire *simulés un par un* ; sans doute ressuscitent-ils nombre de questions philosophiques dont il importe de retrouver la tradition (Gardner, 1985) ;

2. Les sciences cognitives sont nécessairement interdisciplinaires et, par là, un de leurs objectifs est celui de contribuer à la formulation de *concepts nouveaux, de type transversal*, qui ne seront qu'étapes vers une future « science de la cognition » ;

3. Si on ne peut nier l'importance dans la cognition de l'intervention des facteurs sociaux, affectifs, historiques et culturels, on peut faire choix dans une première étape de les ignorer ; ce qui ne va pas sans conséquences épistémologiques graves.

Voilà donc les idées. Elles ne suffisent pas cependant à expliquer les raisons de ce développement considérable des disciplines cognitives, ces dernières décennies. Il aura fallu, on le verra plus loin, le prodigieux essor des neurosciences et les progrès spectaculaires accomplis par l'informatique notamment au travers du développement technique d'ordinateurs de plus en plus performants et remettant même en cause les architectures classiques de construction et de fonctionnement des machines.

Pour ne citer que quelques exemples :

1. en neurochimie, on a pu développer l'exploration systé-

matique des substances (neurotransmetteurs et hormones) intervenant dans la biochimie des processus cérébraux et, par là, révolutionner les conceptions traditionnelles et simplistes du cerveau et du neurone ;

2. en neurophysiologie, on sait aujourd'hui enregistrer l'activité de neurones isolés et étudier les multiples situations neuro-électriques de l'organisme ;

3. en neuroanatomie, les techniques d'imagerie cérébrale permettent une précision jamais atteinte dans les cartographies neuronales ;

4. en informatique, en mathématiques et en physique enfin, la mise au point de théorisations, de logiciels et de langages de « haut niveau » favorise une souplesse inattendue dans le traitement des données et la modélisation.

Ce foisonnement de données et d'approches est peut-être à l'origine, d'une part, de certaines confusions quant à l'objet des sciences cognitives et, d'autre part, des positions tranchées que ces mêmes sciences peuvent susciter sous forme d'affrontements philosophiques où la « croyance » entre parfois pour une plus large part que la simple sagesse épistémologique. Ainsi en va-t-il de l'opposition centrale entre *cognitivism*e et *connexionnisme*. Nous le verrons au terme de ce panorama dont l'ambition est d'être introductif, à défaut d'une exhaustivité rendue impossible par la multiplicité des travaux.

L'intelligence artificielle Rêves, réalités et mythes

L'intelligence artificielle : un rêve très ancien

La naissance de l'intelligence artificielle peut apparaître comme contemporaine, du moins dans l'acception moderne du terme, à savoir : *l'utilisation de l'ordinateur comme moyen de simulation des processus naturels ou comme support de capacités ordinairement attribuées à l'intelligence humaine*. En vérité, le désir de réaliser des « machines intelligentes » date de l'Antiquité. Et sans doute, l'une des premières mentions du genre pourrait se retrouver au chant XVIII de l'*Illiade*, dans lequel Héphaïstos, dieu du Feu, avait construit des tables à trois pieds munies de roulettes qui se déplaçaient toutes seules dans les palais des dieux, sortes de robots avant la lettre (Pitrat, 1985) !

Il y eut aussi, dans la tradition juive, le Golem, automate à forme humaine, fait de bois et d'argile, et qui devenait serviteur du rabbin capable d'inscrire un mot magique sur son front. Plus tard, au XVI^e siècle, on se mit à rêver d'automates simulant le corps humain. Ce rêve perdura. Au XVII^e siècle, Descartes, lui aussi fasciné par les automates, introduisit l'idée d'« animal-machine », avec déjà l'espoir que certaines activités humaines pouvaient être simulées mécaniquement. Cet espoir prit forme au XVIII^e siècle avec Vaucanson, lequel construisit d'abord un automate joueur de flûte traversière, puis surtout, en 1738, son célèbre canard capable de nager, de battre des ailes — chaque aile était composée d'environ deux mille pièces ! —, d'avalier du grain et de rejeter des fientes, en l'occurrence des boulettes de pain colorées...

La figure frappa les esprits de l'époque et, en 1747, La Met-

trie osa franchir le pas en publiant *L'Homme machine* où, pour la première fois, se voyait remise en cause la distinction entre l'homme et la machine. En 1769, un automate encore plus complet vit le jour : celui de Kempelen, prenant la figure d'un Turc assis sur un coffre et placé devant un échiquier. Cet automate eut un grand succès ; c'était en fait une supercherie : un joueur nain s'y trouvait dissimulé, assurant le jeu des pièces. Mais le rêve était lancé et le xx^e siècle allait le concrétiser : en 1912, Torrès y Quevedo réalisait un automate à même de jouer les finales roi et tour contre roi, et, en 1945, K. Zuse fut le premier à programmer les règles du jeu d'échecs.

En 1949, C. Shannon, le célèbre concepteur de la théorie de l'information, mit au point les principes de base d'une méthode pour jouer aux échecs. L'année suivante, le logicien anglais A. Turing construisit à son tour un programme d'échecs qu'il avait simulé manuellement (Turing, 1983). Conjointement, on ne cessait d'inventer des automates en vue de reproduire, dans un premier temps, le comportement animal : ce furent, en 1929, Philidog qui pouvait suivre un rayon lumineux et aboyer en fonction de l'intensité de ce rayon, puis vers 1950, les « tortues » de Grey Walter et le « renard » d'Albert Ducrocq, déjà autonomes et capables d'explorer partiellement un environnement. Vers 1947 surtout, A. Turing fut le premier à envisager l'hypothèse de la réalisation d'une créature intelligente et non humaine ; le rapport qu'il écrivit sur ce sujet ne fut guère apprécié à l'époque bien qu'il s'agît déjà de la définition de ces applications qui composent aujourd'hui l'intelligence artificielle : jeux, résolutions de problèmes et démonstrations de théorèmes et traduction automatique.

Alors — mais est-ce un paradoxe ? —, les plus fervents propagandistes d'une future « intelligence artificielle » furent les écrivains. Déjà, c'était à l'un d'entre eux, le Tchèque Karel Capek, que l'on devait la trouvaille du mot « robot », lequel signifie simplement « travail forcé » en langue tchèque. Il y eut surtout une production considérable de science-fiction concourant notamment à la célébrité d'Isaac Asimov et mêlant rêves futuristes et angoisses d'un monde futur asservissant l'homme à la machine.

Mais l'imagination anticipatrice ne suffit pas à fonder une discipline. Si l'intelligence artificielle a connu l'essor que l'on

sait lors de ces trente dernières années, c'est grâce véritablement aux progrès conjoints de l'informatique, de la logique, de la psychologie et des sciences du vivant.

Les contributions historiques des sciences du vivant et de la psychologie

Il n'y aurait pas, en effet, d'intelligence artificielle sans cette longue histoire de la médecine qui marque la pensée occidentale jusqu'à ce qu'on parvienne à expliquer nombre des fonctionnements du corps humain, et notamment du système nerveux. Bien des aspects de notre cerveau demeurent encore mystérieux, mais, dès les années cinquante, des modèles de neurones ont pu être développés à partir d'hypothèses précises, inspirées du « câblage » neuronal observé de diverses familles animales, depuis les insectes jusqu'à la pieuvre, la grenouille et le chat.

L'une des plus célèbres machines en réseau, le Perceptron, conçu par F. Rosenblatt en 1958, se fondait ainsi sur un dispositif électrique modélisant grossièrement la cellule nerveuse et constitué d'éléments tous câblés entre eux à l'instar des neurones dans le système nerveux. Une longue lignée de machines en sont issues, utilisées en particulier pour le traitement des images. Plusieurs formules sont possibles en vue de constituer de telles machines.

La méthode la plus simple consiste à définir trois couches : une première couche dite de *réception*, qui reçoit l'information de l'extérieur, telle une cellule photoélectrique excitée par la lumière ; une deuxième couche dite d'*association*, recevant les signaux émis par la première couche et agissant sur la troisième, qui donne la *réponse* du système. Naturellement, ces machines en réseau ont des limites, dans la mesure où elles ne peuvent assurer qu'un traitement global des images, mais depuis quelques années, les travaux sur les modèles connexionnistes ont relancé l'intérêt pour de telles architectures multicouches.

Parallèlement à ces progrès des neurosciences, d'autres travaux allaient inspirer la genèse de l'intelligence artificielle, ceux-là directement issus des recherches des psychologues.

Dès la fin du XIX^e siècle ainsi, A. Binet avait attiré l'attention sur les calculateurs prodiges, dont le plus célèbre fut Inaudi. En l'observant, Binet remarqua que la force de sa mémoire provenait autant d'un certain nombre de trucs que de résultats partiels qu'Inaudi connaissait déjà. Les observations sur différents sujets montraient encore que certains calculateurs prodiges disposaient d'une mémoire auditive, et d'autres d'une mémoire purement visuelle. A la veille de la Seconde Guerre mondiale, il y eut surtout les travaux du psychologue néerlandais de Groot qui étudia de façon systématique les stratégies des joueurs d'échecs en corrélation notamment avec leurs motricités oculaires. On s'aperçut ainsi, grâce à de Groot, que la force d'un grand maître aux échecs ne provient pas de la séquence des coups réalisés, mais bien des quelques seuls coups essentiels qu'il sait choisir. Ces observations eurent une influence considérable sur la conception de nombreux programmes d'échecs bien sûr, mais aussi en vue de la résolution de problèmes. Et c'est ici que la logique allait intervenir (Pitrat, 1985).

Les développements de la logique à la source des progrès de l'intelligence artificielle : la machine de Turing

Historiquement, ce sont d'abord les travaux du logicien français J. Herbrandt, mort en 1931, lequel avait déjà mis au point une méthode mécanisable afin de démontrer des théorèmes mathématiques. Herbrandt ne songeait guère aux ordinateurs, mais sa méthode inspira les premières recherches sur les programmes « mathématiciens » et continue de le faire. Il y eut ensuite Alan Turing qui établit un modèle théorique de machine, appelé depuis machine de Turing. Ce modèle allie la simplicité à l'extraordinaire puissance, en même temps qu'il traduit, avec une lenteur apparemment épuisante, toutes les opérations que peut réaliser un ordinateur classique. Il n'est donc pas inutile de décrire au lecteur ce modèle.

1) Dans une machine de Turing, la décomposition des processus en opérations est d'abord ramenée à la limite du possible, et la partie mémoire est représentée sous forme d'un ruban illimité des deux côtés et divisé en cases. Il est clair qu'une

machine qui existe réellement ne peut avoir une mémoire infinie (ruban infini), mais la machine de Turing n'a l'ambition de se représenter que comme un schéma idéal, reflétant la possibilité théorique d'agrandissement de l'étendue de la mémoire. En ce sens, elle assure le lien entre la notion d'algorithme et la notion de machine avec mémoire potentiellement illimitée.

La machine de Turing dispose d'un nombre fini de signes (ou symboles) : S_1, S_2, \dots, S_k , qui forment ce qu'on appelle l'« alphabet extérieur », grâce auquel on code à la fois l'information fournie à la machine et celle qu'elle élabore. Il faut ajouter — et c'est capital —, l'existence du *signe vide* (par définition : S_1) dont l'insertion dans une case quelconque du ruban (de la mémoire) efface le signe qui y était contenu et la laisse vide.

A un stade arbitraire du travail de la machine, chaque case ne contient qu'un signe. Chaque information conservée sur le ruban est représentée par un assemblage fini de signes de l'alphabet extérieur, disposés un par un, dans certaines cases du ruban. Au début donc, l'information initiale est marquée sur le ruban, et le travail de la machine s'effectue selon des « temps » successifs, au cours desquels va s'opérer la transformation de l'information initiale en informations intermédiaires (à la fin de chaque étape, la réunion des signes marqués sur le ruban exprime l'information intermédiaire correspondante), c'est-à-dire qu'en regard de l'information initiale $I(I)$, deux cas sont alors possibles :

— Soit, après un nombre de pas déterminé, la machine s'arrête, donne un signal d'arrêt, et sur le ruban se trouve marquée une certaine information B . Dans ce cas, on dira que la machine est applicable à l'information initiale $I(I)$ et la transforme en information résultante B ;

— Soit encore l'arrêt et le signal d'arrêt ne se produisent jamais, et dans ce cas, on dira que la machine n'est pas applicable à l'information initiale $I(I)$. En résumé, la machine ne résout une certaine classe de problèmes que si elle est applicable aux conditions arbitraires de cette classe de problèmes et que si elle est à même de transformer une information en solution de ce type de problèmes et dans le code utilisé.

2) Dans la machine de Turing, le système des opérations

élémentaires, et avec lui le système des commandements, est simplifié à une « adresse », c'est-à-dire telle qu'à chaque temps une seule case de la mémoire entre en jeu seulement : à chaque temps particulier, le commandement prescrit le remplacement d'un seul signe *si* gardé dans la « case vue » (la case de mémoire en jeu à l'étape considérée), par un autre signe quelconque *sj*. Pour $j = i$, cela signifie que le contenu de la case vue n'est pas changé ; pour $j \neq i$, cela signifie que, si un signe quelconque est gardé dans la case vue, on l'efface. L'idée qui préside à cette simplification est que le contenu des diverses cases, dont on a besoin pour le processus, se cherche en contrôlant progressivement toutes les cases, l'une après l'autre, jusqu'à ce que ne se manifeste plus de case dont on ait besoin. Apparemment, cela semble allonger la durée du processus, mais en vérité, on gagne beaucoup de temps : dans les commandements du programme, à la place d'adresses arbitraires dans les cases vues, on peut se limiter à utiliser en tout trois adresses standard, représentées ainsi :

- mettre sous la case vue la case voisine de droite,
- mettre sous la vue la case voisine de gauche,
- garder sous la vue la même case qu'auparavant.

La particularité spécifique de la machine de Turing est donc que le traitement de l'information se fait dans le « bloc logique », qui peut se trouver dans un nombre fini de positions indiquées par des séries de signes déterminées (q_1, q_2, \dots, q_m). Le bloc a deux canaux d'entrée : par l'un d'eux, à chaque stade du travail de la machine (à chaque temps), passe le signe de la case vue, par l'autre, le signe q_l de la position qui est prescrite au bloc au temps considéré ; par le canal de sortie, le bloc envoie dans la case vue le signal correspondant « transformé » s_j , qui est une fonction à une valeur des signaux s_l, q_l donnés à l'entrée. Les commandements qui fixent le travail de la machine à chaque temps distinct ont donc la forme : Dq_l, Gq_l, Nq_l ($l = 1, 2, \dots, m$), où le premier signe change l'adresse de la case vue, le second prescrit au bloc logique la position convenable. Les signes G, D, N, q_1, \dots, q_m constituent l'« alphabet intérieur » de la machine.

La spécificité de la machine de Turing est donc qu'on impose aussi au bloc logique l'élaboration, à chaque temps considéré, du commandement qui sera présenté dans le bloc

de conduite au début du temps suivant. De la sorte, le bloc logique a, outre un canal pour la distribution des signes s_j , deux autres canaux pour la distribution des signes de commandement à exécuter au temps suivant. Cela signifie que le bloc logique réalise une fonction qui fait correspondre à chaque paire de signes s_i, q_n (prise parmi l'ensemble de toutes les semblables paires, qui sont au nombre de k, m) un triple de signes s_j, P, q_l . De cela résulte que le travail de la machine de Turing est complètement défini par la fonction logique que réalise le bloc logique. En d'autres termes, deux machines de Turing, avec un schème fonctionnel commun, sont identiques au plan de ce qu'elles font.

On peut donc parler d'un « schème de structure » commun à toutes les machines de Turing, lequel schéma reflète la distinction entre mémoire extérieure et mémoire intérieure. La mémoire extérieure est représentée par les cases d'un ruban sans fin, destiné à garder l'information codée en symboles de l'alphabet extérieur ; la mémoire intérieure est représentée par deux cases pour la garde des commandements à exécuter immédiatement : la *Q-case* garde le signe de position et la *P-case*, le signe de déplacement. Dans ces deux cases, sont retenus les signes P, q_l , obtenus à la sortie du bloc logique, pendant le temps de travail considéré, jusqu'au commencement du temps suivant, lorsqu'ils entrent dans le bloc de conduite. La fonction de ce bloc de conduite est, dès lors, d'assurer le déplacement du ruban d'une case au plus, selon le *P-sign*e qui y entre.

Le travail de la machine de Turing, en résumé, se fait de la façon suivante : avant de la mettre en marche, on insère sur le ruban l'information initiale et, dans le « champ de vision » de la machine, on place une case initiale déterminée ; dans les cases *P* et *Q* on met les signes de position et de déplacement initiaux. Le processus qui suit s'effectue dès lors automatiquement et en application du schème fonctionnel de la machine.

Ce processus peut apparaître désespérément lent, et impuissant à résoudre de multiples problèmes. Cela fut prétexte d'ironie pour les adversaires de l'intelligence artificielle. En vérité, une machine de Turing universelle — laquelle correspond à la théorie mathématique des fonctions récursives — constitue un modèle idéal des machines abstraites destinées à effectuer

potentiellement tous les types de calculs exprimables par des algorithmes.

Une machine de Turing, pour simplifier, c'est la composition : a) d'un organe central, capable, comme on l'a vu, de prendre un nombre fini d'états physiques internes ; b) d'une mémoire externe — la bande illimitée de papier ; c) de périphériques (tête de lecture et tête d'écriture) à même d'extraire ou d'inscrire sur la mémoire externe des « informations », représentées par des suites de symboles d'un alphabet. Le fonctionnement de la machine est donc dirigé par un programme dont les instructions de base sont, d'une part, des changements d'état internes en fonction de symboles reconnus dans la mémoire externe, et, d'autre part, des instructions d'écriture en fonction des mouvements de la bande qui défile devant les périphériques de lecture et d'écriture. Toute opération de calcul consiste alors : 1) à mettre la machine de Turing en état initial ; 2) à lui faire reconnaître progressivement une suite initiale sur la bande ; 3) à transformer cette donnée initiale du calcul, en exécutant les instructions du programme, en vue de construire le résultat du calcul à inscrire sur la bande. Le fonctionnement d'une machine de Turing est donc entièrement tributaire d'un programme. Le modèle des machines de Turing est en vérité une théorie mathématique de l'effectuabilité potentielle, de la calculabilité et de la démontrabilité. Une machine de Turing modélise donc la compétence d'un calculateur idéalisé opérant sur des suites de symboles au moyen de règles de réécriture, et c'est en ce sens qu'elle demeure un puissant référent des logiques encore à l'œuvre en intelligence artificielle.

L'émergence de la cybernétique

Les progrès suivants dans cette genèse de la théorie des machines furent réalisés au cours de la Seconde Guerre mondiale, lorsqu'il s'est agi de développer, pour les besoins militaires, différents types de dispositifs du domaine aéronautique, dont, en particulier, les radars. Pour mettre au point des mécanismes de commande pouvant prendre en compte des variations déterminées de paramètres tels que la température, l'orientation, la pression, les chercheurs — mathématiciens et

électroniciens — s'intéressèrent de très près aux mécanismes neurophysiologiques analogues, intervenant dans la régulation du corps humain. Ces concours interdisciplinaires amenèrent à fonder la *cybernétique*, nouvelle discipline ayant pour vocation de modéliser mathématiquement les processus de traitement de l'information. De là viendront la plupart des fondateurs de l'intelligence artificielle. Entre-temps, les avancées de la linguistique avaient permis de défricher nombre de problèmes liés au traitement du langage naturel et, surtout, les progrès industriels aidant, l'apparition régulière de nouveaux ordinateurs de plus en plus rapides et puissants allait contribuer au développement effectif de cette communauté de travaux et de scientifiques qui, dès 1956, adopterait le nom d'« intelligence artificielle ».

La fondation de l'intelligence artificielle

A partir des années cinquante, apparaissent ainsi les premiers programmes de calcul formel et peu à peu les informaticiens découvrent que leurs machines peuvent servir à bien autre chose qu'à effectuer des calculs, et notamment qu'elles peuvent « manipuler des symboles ». 1956 devient effectivement l'année clé de l'intelligence artificielle. Deux événements y contribuent : tout d'abord la mise au point du premier « vrai » programme d'intelligence artificielle — le Logic Theorist, de Newell, Shaw et Simon, qui peut démontrer des théorèmes de la logique des propositions du type « si p et q , alors p » et « si *non* (*non p*) alors p » ; ensuite la réunion au collège de Darmouth d'une douzaine de chercheurs, réunion au cours de laquelle Newell et Simon présentent leur Logic Theorist, et proposent le terme d'« intelligence artificielle » pour mieux dénommer cette nouvelle perspective de travaux visant à reproduire et à modéliser des conduites jusque-là réservées aux êtres humains.

Rapidement, la recherche va prendre essor dans de multiples domaines : programmes d'échecs, de géométrie, de résolution de problèmes, etc., favorisés par de nouvelles générations d'ordinateurs et, surtout, la mise au point de méthodes de programmation adaptées au traitement des symboles dont la plus

célèbre, Lisp, fut développée à partir de 1960 par J. McCarthy. L'enthousiasme est contagieux : en 1958, Newell et Simon lancent leur fameux défi, en prédisant qu'en 1968, dans dix ans au plus tard, un programme verrait le jour, qui serait champion d'échecs et démontrerait un important théorème mathématique. Cette prédiction ne s'est pas réalisée.

En vérité, les programmes de ces débuts de l'intelligence artificielle demeuraient très combinatoires, c'est-à-dire qu'ils étaient capables de tester un très grand nombre de possibilités, sans pour autant pouvoir examiner chacune d'elles en profondeur. Par exemple, s'agissant des programmes de jeux d'échecs, de tels programmes envisagent d'abord tous les coups possibles et, pour chacun de ces coups, passent en revue toutes les ripostes adverses et cela fait, considèrent encore tous les coups possibles. Ces programmes vont donc, comme on dit, jusqu'à la profondeur 3 (ami-ennemi-ami), et si l'on souhaite aller plus loin, le nombre de combinaisons à examiner augmente à une vitesse considérable : en milieu de partie, il y a ainsi une quarantaine de coups possibles et, dès lors, augmenter la profondeur d'une unité multiplie le temps de calcul par quarante environ, de deux unités par 40×40 , etc. Autrement dit, analyser davantage de combinaisons implique à chaque fois une croissance coûteuse des temps de calcul qui n'est guère suivie d'une amélioration significative des résultats. La seule solution est de reconsidérer l'approche des problèmes et le mode de leur traitement. Tel est l'objectif des recherches les plus actuelles en intelligence artificielle.

Un autre cas, encore plus éclairant, est celui de la traduction automatique. Dès 1956, se fabriquent des programmes qui traduisent, plutôt fort mal que bien, du russe vers l'anglais. Il s'agit en général de dictionnaires dont les compléments grammaticaux demeurent rudimentaires.

Ces tentatives maladroites auront eu au moins la pertinence d'aider à mieux profiler un certain nombre de principes généraux à l'intelligence artificielle, à savoir : 1) la nécessité de créer des programmes généraux, c'est-à-dire que les connaissances spécifiques à un domaine doivent être indépendantes du programme lui-même, lequel n'est alors qu'un simple *moteur* adaptable à différents types de domaines ; 2) l'importance de l'*apprentissage* : le système doit être capable de découvrir et

d'explorer lui-même les connaissances qui lui sont nécessaires. Ces deux problèmes demeurent centraux dans les développements contemporains.

Vers un modèle des conduites et capacités humaines

De même qu'il en avait été des automates au XVIII^e siècle, les premiers succès de l'intelligence artificielle ont ainsi fasciné les chercheurs, et certains envisageaient allégrement de parvenir à construire des machines à l'égal de l'intelligence humaine et, qui sait, supérieures. D'où l'emprunt de nombreux programmes aux résultats d'expériences de psychologie. Conjointement, les projets issus de la cybernétique, en écho des modèles de la cellule nerveuse, continuaient de faire leur chemin, suite au Perceptron de F. Rosenblatt (1958), dont l'élément de base, on l'a vu, était un schéma plutôt grossier du neurone. La situation se développa de la sorte, jusqu'au milieu des années soixante. Puis un tournant s'opéra aux alentours de 1965, lorsque apparurent de nouveaux ordinateurs plus puissants. Depuis, deux approches continuent de coexister en intelligence artificielle : la première, de type combinatoire, utilise la machine pour expérimenter le maximum de possibilités ; la seconde s'efforce d'approcher de modèles les plus proches de nos capacités humaines, en ayant recours pour ce faire à diverses contributions interdisciplinaires.

Ces deux types d'approche ont connu des fortunes diverses. Il est manifeste que la première, l'approche combinatoire, connut un grand succès dans les années soixante-dix, succès favorisé par la puissance des nouveaux calculateurs. Ainsi, en reprenant l'exemple du jeu d'échecs, les possibilités de calcul offertes, la diminution des coûts temporels ont fait qu'en pratique on a pu arriver à des profondeurs de l'ordre de huit, avec des programmes ne commettant jamais d'erreurs dans le calcul des ripostes à l'adversaire, et en 1967, R. Greenblatt a pu enfin proposer un programme d'échecs à même de battre sans difficulté un joueur « normal ». A la même époque, de nouveaux et nombreux automatismes — les robots — ont vu le jour, souvent bien plus complexes et passionnants que les robots industriels qui depuis se sont répandus. Nous y reviendrons.

Mais dans ce même temps où la combinatoire était poussée à ses limites, une équipe de l'université Stanford mettait au point un programme nommé Dendral, destiné à analyser les résultats obtenus en spectrographie de masse. La quantité de connaissances nécessaires dans ce cas est considérable, défiant toute combinatoire, et il fallut bien évidemment bâtir un système capable de gérer de tels stocks de connaissances. C'était l'ancêtre de nos modernes systèmes experts. Une révolution nouvelle s'installait en intelligence artificielle. La discipline avait acquis droit de cité : en 1969, se tint le premier congrès international des spécialistes du domaine et, en 1970, paraissait le premier numéro de la revue *Artificial Intelligence*.

L'expansion de l'intelligence artificielle

A partir de ce congrès de 1969, les différents travaux, qui jusque-là se glissaient au hasard des multiples colloques d'informatique, se regroupèrent dans la revue *Artificial Intelligence* et, de 1970 à 1980, les progrès se multiplièrent. Ce fut le cas, d'abord, dans le traitement des langues naturelles : après d'innombrables tentatives pour passer directement du russe à l'anglais, on comprit qu'était nécessaire une étape intermédiaire qui devait consister à représenter le sens indépendamment de la structure en surface de la langue naturelle, et qu'il fallait donc concevoir deux programmes : un *analyseur* chargé de cette représentation sémantique du texte, et un *générateur* pour reconstruire cette représentation dans l'autre langue. A partir de là, on sut mieux comment donner des instructions à un robot, dans la mesure où ces ordres pouvaient être traduits en « représentation du sens » et, de là, en « programme autonome ».

Conjointement, les progrès des machines aidant, des stocks considérables de connaissances purent être installés en ordinateur. Un des premiers systèmes experts fameux, Mycin, mis au point à Stanford en 1974, se révélait ainsi capable de diagnostiquer, parmi de multiples données, une infection bactérienne et de prescrire les traitements nécessaires.

Là encore, les progrès conceptuels furent acquis lorsqu'on comprit la nécessité de séparer la base de connaissances et le

programme — ou « moteur d'inférences » — chargé de les traiter. Dans un programme ordinaire, en effet, les connaissances sont intriquées aux instructions destinées à les utiliser : on dit qu'elles sont sous forme « procédurale ». Dans un système expert, les connaissances sont stockées, sans mode d'emploi immédiat : on dit qu'elles sont sous forme « déclarative ». Il est beaucoup plus commode de fournir à un système des connaissances « déclaratives » parce qu'on peut les modifier facilement, en ajouter ou en retrancher. Cela nécessite en retour une grande qualité du moteur d'inférences destiné à les traiter, mais une fois réalisé, il n'est plus remis en cause et, du fait de sa séparation vis-à-vis de la base de connaissances, on peut encore lui demander, à des fins de contrôle, le raisonnement qu'il a suivi. C'est encore vers 1975 qu'A. Colmerauer conçut le langage Prolog, dont on sait les multiples applications depuis.

Depuis lors, l'intelligence artificielle n'a plus guère à prouver la pertinence de ses méthodes et son rôle au sein des sciences cognitives. Cela ne signifie pas pour autant que tous les problèmes soient résolus. En particulier, deux questions demeurent cruciales : si l'on veut utiliser efficacement une quantité importante de connaissances, il faut un moteur d'inférences capable de choisir à chaque fois lesquelles considérer et de savoir comment les employer : c'est le problème de la formulation des *métaconnaissances* nécessaires pour une bonne exploitation des systèmes ; l'autre question est celle de la mise en relation interne des connaissances multiples et afférentes à des domaines de savoir complexes : aucune stratégie préconçue n'est possible ; il faudra bien parvenir à des systèmes capables d'*apprentissage*, c'est-à-dire de construire, à partir de connaissances antérieures, de nouvelles connaissances. Deux problèmes redoutables qui rendent encore discutable l'unité des approches en intelligence artificielle (Pitrat, 1990).

D'où la délimitation des champs spécifiques qui la sous-tendent : systèmes experts, simulation du raisonnement, traitement du langage, reconnaissance de la parole, représentation des connaissances et robotique.

Les systèmes experts

S'il est coutumier d'entendre que les systèmes experts raisonnent « comme des experts humains », cela ne veut rien dire d'autre que leur démarche ressemble plus ou moins à celle qu'adopterait un expert sur un problème donné. La plupart des systèmes experts ne sont que des « simulacres » de raisonnements, mais, en ce sens, leur vertu essentielle — en dehors des services considérables qu'ils peuvent rendre — est bien celle de poser question sans cesse quant à la nature du raisonnement humain.

Pour concevoir un système expert, il faut d'abord analyser le raisonnement d'un expert humain sur le domaine considéré en vue de coder ce raisonnement en termes logiques et informatiques, puis il faut construire en machine une synthèse artificielle de telle sorte que l'ordinateur procède de la façon la plus proche possible de ce raisonnement. Tout repose, au départ, sur la méthodologie d'entretien avec cet expert humain, et ensuite sur la capacité à analyser les différents types de raisonnement (déduction, implication, intuition, etc.) impliqués dans le domaine en question, afin de modéliser ces raisonnements. Dans la pratique, donc, l'expérience, la compréhension et le jugement du concepteur de système vont jouer un grand rôle.

Une théorie admise fait distinguer trois niveaux de connaissances, lesquels vont différencier les trois phases d'étude et de mise au point d'un système expert : un niveau structurant, un niveau conceptuel et un niveau cognitif.

Le *niveau structurant* organise le mécanisme qui construit la « certitude », laquelle va dépendre de l'état des connaissances dans le domaine. Cette certitude va varier selon qu'il s'agira de physique ou de médecine ou de géologie. De cette structuration du domaine vont dépendre les types de représentations formelles des connaissances élaborées et les processus déductifs établis à l'intérieur de ces connaissances.

Le *niveau conceptuel* consiste à définir et spécifier les différents types de concepts opératoires utilisés par l'expert. Ces concepts correspondent aux « nœuds » du raisonnement d'expertise. A chacun sera associé un « descripteur élémentaire » exprimé dans le formalisme précédent.

Le *niveau cognitif* enfin rassemble l'ensemble des connais-

sances mobilisées par l'expert. Ces connaissances doivent toutes pouvoir être exprimées dans le formalisme du premier niveau et à travers les descripteurs du niveau conceptuel.

En résumé, un système expert est conçu pour fonctionner comme intermédiaire entre l'expert humain d'un domaine de connaissances et l'utilisateur potentiel de ces connaissances, utilisateur qui doit pouvoir acquérir une pratique analogue à celle de l'expert grâce à l'observation des fonctionnements du système. Surtout, ce que les systèmes experts introduisent et que les programmes informatiques classiques ne prennent pas en compte, ce sont des modes de raisonnement du type *savoir-faire*, autrement dit des *heuristiques* (du grec *heuriskein* : trouver), qui sont des processus d'exploration, de découverte et de mise en œuvre de savoirs locaux ou composites. De plus en plus, on s'approche ainsi de la façon dont chacun de nous procède pour se construire des connaissances au cours de la vie et lors de la rencontre avec de nouvelles expériences.

Une architecture en trois composants

C'est pourquoi l'essentiel d'un système expert repose sur son architecture en trois composants :

— La *base de connaissances*, qui contient l'ensemble des informations relatives au domaine traité et qui est écrite dans un langage de représentation des connaissances ;

— La *base de faits (mémoire de travail)*, qui enferme les données des problèmes à traiter et qui, en conservant la trace des raisonnements produits, joue le rôle de mémoire auxiliaire ;

— Le *moteur d'inférences* enfin, c'est-à-dire le programme destiné à utiliser les données et les heuristiques de la base de connaissances pour résoudre les différents problèmes afférents à ces informations et à ces données.

Le moteur d'inférences a donc un rôle important, qui est celui de sélectionner, de valider et de déclencher les règles nécessaires à la résolution de tout problème posé. Ce moteur d'inférences peut être de conception générale et applicable à une gamme variée de bases de connaissances. Mais ce n'est pas un modèle de déduction généralisée ; il existe un grand

nombre de moteurs d'inférences et la recherche à ce niveau mobilise toujours les efforts des scientifiques.

Les outils logiques

Les questions logiques et techniques ouvrent ici un vaste champ d'interrogations à caractère cognitif quant aux statuts et aux modèles du raisonnement « naturel » : une déduction classique et contraignante peut fort bien, en effet, ne pas aboutir tout en donnant une conclusion « vraie », là où un raisonnement « approximatif » se révélera efficace bien que produisant des résultats à première vue erronés. D'où l'importance des travaux logiques les plus récents sur le « flou », le « vague » et « l'incertain » permettant de prendre en compte et de modéliser des continuums de possibilités allant du vrai au faux et où chaque événement sera, par exemple, affecté d'un coefficient de vraisemblance compris entre -1 (le définitivement faux) et $+1$ (le vrai). Mais jusqu'ici, les chercheurs ont été impuissants à construire des systèmes capables de modéliser simultanément des raisonnements « certains » complexes et des raisonnements « incertains ».

L'objectif général est cependant de parvenir à élaborer un système général constitué d'un nombre imprécisé de couches de *métaconnaissances*, qui constitueraient en quelque sorte la base de base de connaissances d'un futur système potentiel en stratégies de métaraisonnement et en expertise d'autres systèmes. Tel est le projet auquel s'attachent des chercheurs comme J. Pitrat, R. Kowalski ou A. Bundy (Ganascia, 1985, 1990 ; Pitrat, 1990).

Dans l'immédiat, la construction des systèmes experts continue donc d'emprunter à une gamme d'outils, plus ou moins cohérente, et selon cette méthodologie en trois étapes : le « dialogue » avec un expert humain d'un domaine, la formalisation des connaissances recueillies, la mise au point finale de la base de connaissances. Le dialogue avec l'expert humain est crucial : il s'agit de détecter les points clés de la connaissance, les nœuds du raisonnement, d'identifier avec précision les objets, les concepts et les fils conducteurs qui orientent le parcours des savoirs et des savoir-faire de l'expert. Il ne faut, à ce propos, négliger aucun détail linguistique : les mots techniques ne

suffisent pas ; encore faut-il repérer les ambiguïtés, les mots du vocabulaire courant et leurs emplois, les qualifications, etc. La présence de linguistes se révèle ici précieuse.

Dans une seconde étape, il importe d'organiser les raisonnements sous-jacents à ces discours de l'expert, et cela est d'autant plus difficile qu'on est en présence de processus cognitifs et expressifs « naturels », lesquels ne suivent pas forcément les étapes analytiques (décompositionnelles), familières au discours scientifique. Chacun de nous, lorsqu'il s'agit d'analyser un problème, fait d'abord un raisonnement global, puis des contradictions apparaissent, et ce raisonnement s'affine jusqu'à parvenir à une conclusion exempte de controverse. C'est le raisonnement du *sens commun* qui intrique connaissances générales, expertises locales et heuristiques de construction du savoir à partir des savoir-faire. Modéliser ce type de raisonnement, compte tenu de la diversité de ses formes argumentatives et stratégiques, n'est pas une mince affaire. Cet objectif concerne autant les linguistes, les cognitivistes que les logiciens et les informaticiens.

C'est pourquoi se sont développés tant de travaux modernes sur les *coefficients de vraisemblance*, et surtout sur l'introduction de la *non-monotonie* dans les systèmes de production et les logiques de représentation. Les coefficients de vraisemblance permettent de construire des raisonnements incertains en affectant des poids de vérité ou de vraisemblance aux inférences successives, pour ensuite combiner ces poids entre eux afin de définir des stratégies de raisonnement. Du côté des procédures non monotones, on dira que tel est le cas d'un système si le retrait d'un événement, dans une base de faits cohérente, est susceptible de conduire à une contradiction. Exemple classique : si « Zoé est une autruche » et que le système expert en infère qu'elle est un oiseau qui ne vole pas, et qu'ensuite on retire de la base de données l'événement « Zoé est une autruche », le système va pouvoir inférer que Zoé est un oiseau qui vole, puisqu'elle n'est plus une autruche. On a donc une contradiction à résoudre du fait de la coexistence de « Zoé vole » et de « Zoé ne vole pas » (Ganascia, 1985).

Cette gestion de la « non-monotonie » se révèle coûteuse, et peut même conduire à ce qu'on appelle une « explosion combinatoire », c'est-à-dire l'impossibilité de contrôler les multiples

paramètres du raisonnement et de sa modélisation en termes de sous-procédures. Ce problème théorique et pratique motive actuellement les chercheurs et des réalisations en ce sens ont vu le jour : le Truth Maintenance System de Doyle, les réseaux sémantiques à propagation de Fahlman ou encore la circonscription de McCarthy. Mais on ne peut procéder que cas par cas, et à chaque fois, en élucidant le plus exhaustivement possible les raisonnements des experts en vue d'élaborer des niveaux structurants. Cela semble parfaitement atteignable dans des systèmes du type aide au diagnostic médical, où il s'agit souvent de caractériser une situation à l'aide d'un descripteur (par exemple : maladie = rougeole, symptômes = fièvre, boutons, etc.) ; cela l'est moins dans les situations de résolution de problèmes et de planification des tâches où, comme en robotique, le problème à résoudre n'est défini que par un ensemble de contraintes à optimiser alors que, parallèlement, l'environnement évolue dans le temps même de l'exécution de ces tâches et qu'il faut mémoriser les actions exécutées pour éviter des séquences d'opérations dont l'une effacerait la précédente. Pour y parer, il faut donc savoir introduire des règles générales contrôlant les actions accomplies, d'où la nécessité de faire recours à des logiques d'ordre supérieur.

Mais les problèmes de constitution de systèmes experts ne s'arrêtent pas à cette simple première phase du niveau structurant. Une fois un formalisme établi à partir de ce niveau, il reste à savoir comment utiliser ce formalisme pour orienter l'écriture de la base de connaissances. Pour ce faire, il faut d'abord définir un langage d'expression — c'est le niveau conceptuel —, et ensuite adapter ce langage à la transcription des connaissances de l'expert. Dans le premier cas, l'important est de déterminer l'ensemble des descripteurs et des prédicats employés par l'expert ; dans le second cas, il s'agit de traduire correctement le type de savoir en question et les connaissances circonscrites. Problèmes délicats qui font que les systèmes experts sont toujours encore en phase de développement, mais qui, en même temps, font considérablement progresser la recherche en sciences cognitives.

L'enjeu est en effet considérable, à la fois technique, industriel et théorique. Du côté industriel, c'est la multiplication actuelle des systèmes experts pour une variété considérable de

domaines. De nombreux savoir-faire du monde de l'entreprise peuvent être partiellement rationalisés, assurant ainsi des gains de productivité non négligeables ; d'autres surtout doivent être mémorisés, sous forme de « capitalisation des savoirs », ne serait-ce que pour éviter leur perte en cas de disparition d'un expert. Il s'agit autant de permettre à des profanes d'accéder à une connaissance que de garantir une entreprise, une compétence, un investissement technologique contre la « fuite des cerveaux » (Ganascia, 1985, 1990).

Du côté technique, il s'agit de cette révolution qu'apportent les nouveaux ordinateurs mis sur le marché : de puissants systèmes experts peuvent être maintenant installés sur des machines de taille réduite, aisément manipulables et à grande diffusion.

Du côté théorique, si les rêves futuristes de machines « supérieures » en intelligence à l'homme se révèlent peu fondés, il est manifeste que c'est du côté des systèmes experts que les progrès les plus importants sont à attendre vers une « intelligence artificielle ». Cela non seulement au plan des connaissances, mais surtout dans le développement de modèles aptes à gérer ces connaissances sous forme de métaconnaissances en machines capables, dès lors, d'acquérir et d'utiliser de nouvelles connaissances, et surtout d'explicitier mieux que nous ne savons le faire les processus par lesquels nous-mêmes apprenons et nous représentons nos propres connaissances (Prince, 1991 ; Hoc, 1991).

La représentation des connaissances

Le thème de la représentation des connaissances envahit aujourd'hui le champ des travaux en intelligence artificielle. Les ambiguïtés n'en sont pas pour autant levées : de quelles « connaissances » s'agit-il ? Et qu'appelle-t-on « représentation » ? Savoir répondre à une liste de questions ne signifie pas forcément « connaître » un sujet. Toute vraie connaissance suppose l'usage à bon escient des informations qui sont les nôtres. Quant aux représentations qu'on pourra se faire de ces informations, elles dépendront de multiples facteurs dont chacun, isolément, pourra les orienter différemment : la situation,

le domaine concerné, la personne visée, les « notions » organisant la connaissance, le « langage » dont on dispose, etc. Les classifications du savoir sur un domaine n'ont ainsi d'autre vertu que celle de nous aider à « ranger nos idées ». Comment ensuite les organiser, les orienter, tel est effectivement le problème des métaconnaissances nécessaires pour *apprendre à savoir utiliser les connaissances* et, par suite, à « délier » les moteurs d'inférences inhérents à de multiples systèmes experts (Krivine, David, 1991).

Si le propre d'un système expert est de savoir « raisonner », cela, dans bien des cas, passe encore par la représentation des connaissances en termes de règles de production du type : « *Si* prémisses *alors* conclusions. » Ces règles sont souvent des propositions logiques reliées par des opérateurs « et », « ou », « non ». Exemple : la proposition « *Si* (A et B) *alors* C » qui organisera des séquences du type : « *Si* pression normale et débit normal, *alors* pompe en bon état. » De telles règles, empruntées à la logique des propositions, permettent la formalisation d'algorithmes simples pour les moteurs d'inférences, lesquels vont avoir pour rôle de déterminer quelles règles seront applicables à tel moment donné et sur tel ou tel problème de connaissances. Cependant, l'objectif des chercheurs étant de parvenir à des formalismes de représentation « déclaratifs », c'est-à-dire exprimant les connaissances indépendamment de la manière de les exploiter, la logique des propositions se révèle vite impuissante à y contribuer. D'où le recours massif à la logique mathématique ou encore à la logique des prédicats, ou logique du premier ordre, inspirant ainsi les langages Prolog (mis au point par A. Colmerauer) ou Snark (conçu par J.-L. Laurière de l'université Paris-VI) ou encore Tango (développé par M.-O. Cordier, au laboratoire de recherche en informatique d'Orsay).

En règle générale, le reproche qu'on peut faire aux différents systèmes jusqu'ici développés est qu'ils sont contraints de suivre des raisonnements nécessairement rigoureux sous la forme exclusivement déductive. La logique des propositions comme celle des prédicats sont des logiques catégoriques, ne distinguant que deux valeurs de vérité : le vrai et le faux. Il n'en va pas de même dans les fonctionnements naturels de nos connaissances, lesquelles procèdent plus souvent de l'à-peu-

près que du « vrai » *stricto sensu*. Souvent, les experts consultés avant toute élaboration d'un système expert témoignent de ces jeux d'approximation qui vont fonder une compétence humaine.

Mais un autre problème provient de ce que les données auxquelles on appliquera des règles ne sont pas toujours exemptes d'erreurs ou d'imprécisions ; auquel cas il faut nécessairement leur appliquer un « coefficient de vraisemblance ». Enfin, il peut arriver que plusieurs règles aboutissent à une même conclusion, mais avec des coefficients de vraisemblance différents ; ce qui implique de faire recours à la théorie des probabilités. Par exemple, dans la succession de règles déductives suivantes :

- « Si la voiture cale lorsque le moteur est au ralenti, c'est probablement (coefficient 0,7) à cause du système d'allumage ;
- Si la voiture démarre difficilement lorsque le moteur est froid, c'est probablement (coefficient 0,6) à cause du système d'allumage » (Gallaire, 1985).

En admettant que les faits soient vérifiés, il y a donc deux règles approximatives conduisant à la même conclusion : le système d'allumage. Mais en quoi et comment convergent-elles ? On ne peut le trancher que par un système de pondération des règles, sous forme de valeurs approximatives, telles que les formula un des premiers systèmes experts : le système Mycin, réalisé par E.H. Shortliffe et son équipe, à Stanford, pour simuler le raisonnement médical. Mycin utilise une technique « floue » pour conseiller les médecins sur le choix d'un antibiotique afin de traiter les infections bactériennes avant qu'on ait pu identifier avec certitude le germe responsable de telles infections ; les mécanismes de son moteur d'inférence sont fondés sur quelques centaines de règles du type : « Si : le germe est positif à la coloration de Gram, *et* que sa morphologie est celle d'un coque, *alors* : cela induit (0,7) qu'il s'agit d'un streptocoque. » En conséquence, on peut soumettre au système chaque dossier d'un patient, par exemple : « Homme de quarante-cinq ans ; céphalée depuis dix jours ; ponction lombaire ; protéines : 253 mg/dL ; globules blancs : 790/mm³ dont 89 % de lymphocytes. » La tâche du système est alors de sélectionner les règles susceptibles de s'appliquer à ce cas et d'en dégager les hypothèses diagnostiques adéquates, puis de

remonter de ces hypothèses à leurs conditions origines (*chaînage arrière*) pour solliciter éventuellement d'autres informations, d'évaluer ensuite le degré de certitude des diagnostics probables et, enfin, de prescrire la thérapeutique la mieux appropriée (Gallaire, 1985 ; Pitrat, 1986 ; Ganascia, 1990).

Pensée naturelle, catégories et prototypes

Tout cela témoigne, encore une fois, de l'approximation inhérente à nos modes de pensée naturels, de même que de l'imprécision des connaissances « en situation », par suite de la difficulté de modéliser de telles stratégies de raisonnement, faute d'outils logiques adéquats. Nous savons bien, par exemple, que tous les oiseaux ne volent pas, mais si on nous dit qu'un oiseau est venu percuter un avion — chose plus fréquente qu'on n'imagine —, nous le croirons volontiers, parce que communément nous admettons que « les oiseaux volent ». Autrement dit, nous raisonnons, faute de mieux, sur des « globalités », des « catégorisations » nécessairement incomplètes (Dubois, 1991).

C'est ce qu'on appelle aujourd'hui le *raisonnement par défaut* (Kayser, 1987). Cela implique de réfléchir à l'opportunité de nouvelles logiques, dites *non monotones*, c'est-à-dire capables de prendre en compte des vérités « relatives » comme il en va dans nos raisonnements quotidiens. Des progrès ont été faits en ce sens, mais il demeure la difficulté de devoir traiter en réalité d'*objets* et de *relations* entre ces objets, surtout lorsqu'il s'agit de construire des bases de données.

D'où, face à ces problèmes, la multiplication de réseaux sémantiques visant à représenter des domaines de connaissances, tous empruntant à la psychologie ou à la linguistique. L'idée est toujours d'établir un graphe exprimant des relations entre éléments constitutifs d'un univers, une sorte de modèle du monde dans lequel se situe le problème à traiter. De cette façon, il apparaît acceptable et commode d'organiser des concepts, de les classer, de les hiérarchiser ; par exemple : « mammifère = vertébré », « mammifère = porteur de mamelles », etc., peuvent s'intégrer dans un graphe sous forme de nœuds conceptuels reliant entre eux « mammifère », « vertébré » et « mamelles » *modulo* des relations du type *est-un* ou

possède. On dira alors que le « sens » de chaque nœud du graphe est défini par l'ensemble des concepts qui seront rattachés à ce nœud par l'intermédiaire de relations prédéfinies.

Les réseaux sémantiques ont connu une grande vogue il y a quelques années. Ils ne sont plus guère utilisés aujourd'hui, mais on les retrouve sous-jacents à bien des formalismes, où, quelle que soit la complexification des règles avancées, celles-ci reposent toujours sur une description et une classification d'*entités* initiales, sans trop qu'on ait bien le temps de vérifier la pertinence de telles classifications.

Une approche plus moderne se fonde sur la notion de prototype (*frame* en anglais). Il s'agit de modèles visant à organiser des univers en « classes d'objets ». L'usage courant, lorsqu'on représente une classe d'objets, est de considérer les propriétés que ces objets partagent en commun — par exemple : mammifère = animal à sang chaud, porteur de mamelles, allaitant ses petits, *type* : la « vache », etc. La description qu'on fait alors s'inspire toujours d'un « représentant » de la classe en question, lequel reçoit le statut de prototype. Et comme les classes naturelles s'organisent en hiérarchies, ce processus de décomposition en propriétés et caractéristiques, se répétera de même pour les sous-classes : « lapin », « chat », etc.

Si l'on considère un prototype de la classe « chat », il comportera nécessairement une partie fixe commune à tous les éléments de la classe (« tous les chats »), et une partie variable caractérisant chaque instance et différente d'une « sous-espèce » de chat à une autre. On dira que les attributs représentent l'ensemble des caractéristiques attachées aux éléments d'une classe. Certains *attributs* vont avoir le statut non pas de données naturelles ou numériques, mais de « procédures » — de sous-programmes — organisant des dérivations ou des extensions à partir d'un prototype, exemplifiant une classe. En résumé : d'un côté, les attributs contribuent à définir les « états » de l'objet ; de l'autre, les procédures aident à construire ce que cet objet « sait faire » ou « ce qu'on peut lui faire faire », soit en agissant sur son propre état interne, soit en le faisant agir sur d'autres objets. On peut ainsi construire des relations entre attributs de prototypes et, lorsqu'un objet sera créé, les valeurs qu'on attribuera à ses attributs permettront alors de le distinguer à l'intérieur d'une classe. Ces

notions d'objet et d'attribut ont été reprises à l'intérieur d'une nouvelle famille de langages de programmation, apparemment mieux adaptés au traitement des connaissances ; le plus connu de ces langages est *Smalltalk*, conçu par A. Kay chez Xerox. Ces langages visent à intriquer structures de données et procédures.

Les problèmes sémantiques et cognitifs

La question qui demeure est celle des moyens de trier entre les types de formalismes existants selon encore les applications qu'on en attend. La plupart de ces formalismes ne sont que des variantes les uns des autres. Cette situation témoigne d'un réel déficit de modèles sémantiques (Visetti, 1991). Alors, il ne reste au mieux qu'à tenter de privilégier la neutralité des représentations que le chercheur construit, et cela impose d'employer le plus souvent des formalismes « déclaratifs », impliquant des sélections précises dans les informations qu'on intègre. Par exemple : on peut choisir d'insérer la connaissance : « Si un vertébré allaite, alors c'est un mammifère », mais on peut aussi déclarer symétriquement : « Si un vertébré est un mammifère, alors il allaite ses petits. » Intégrer de telles règles symétriques pose de nombreuses difficultés : le danger est celui d'aboutir à des raisonnements de type strictement circulaire, pour la simple raison qu'il demeure difficile encore de gérer techniquement de telles équivalences, pourtant naturelles à nos raisonnements ordinaires (Rastier, 1991).

Autre problème déductible du précédent : celui de savoir harmoniser certaines complémentarités des formalismes. Exemple : un système expert devant diagnostiquer des pannes mécaniques devra non seulement être compétent en technique des dispositifs à gérer, mais encore capable d'intégrer à la fois les lois de la physique nécessaire et les enseignements des dysfonctionnements ou échecs rencontrés. Aucun formalisme n'est capable d'organiser des compétences aussi différentes. Il faut en convoquer plusieurs, chacun devant se charger d'un type de problème, sans qu'on sache trop les conséquences de ce choix sur le formalisme voisin. Nous ne disposons à l'heure actuelle d'aucune certitude sémantique et cognitive quant à ce que pourrait être un système « multifonctionnel » sinon univer-