

Jean-Paul Baquiast

# Pour un principe matérialiste fort



DécohérenceS

**Compléments du livre**

**"Pour un principe matérialiste fort"**

## **Chapitre 4**

### **Construction de systèmes de conscience artificielle**

**Chapitre 4, Section 1 :**

**Un cahier des charges fonctionnel pour une conscience artificielle**

**Chapitre 4, Section 2 :**

**Les recherches de Sony-CSL Paris**

**Chapitre 4, Section 2 :**

**Le programme ECAgents. Acquisition du langage par les robots**

**Chapitre 4, Section 3 :**

**Réalité virtuelle, vie artificielle, automates cellulaires**

## Chapitre 4, Section 1 :

### Un cahier des charges fonctionnel pour une conscience artificielle

Il ne s'agit pas de revenir sur la question de la conscience humaine, mais de se placer dans la position de l'informaticien concepteur de systèmes calculables, qui propose une vision de la conscience artificielle à partir des techniques les plus sophistiquées de l'intelligence artificielle. L'informaticien est un scientifique qui fait des modèles et qui réalise des systèmes. Autrement dit, il fait appel à la science des modèles calculables.

Précisons par ailleurs que ce que nous présentons ici ne prétend pas définir la façon dont tous les informaticiens devraient ou ne devraient pas spécifier un projet de cette nature. Il s'agit d'une possibilité parmi d'autres. Mais comme les responsables de projets sont généralement très discrets sur la façon dont ils voient les choses, nous ne pouvons rien dire de ce que pourraient être d'autres approches. Nous devons par contre remercier Alain Cardon qui, dans la grande tradition d'ouverture de l'université française, n'a pas hésité à publier ses idées depuis leurs origines.

Le problème est le suivant : la meilleure manière d'imiter les phénomènes complexes caractérisant la conscience humaine, dont l'organisation fait intervenir des actions à de multiples niveaux, n'est pas de découvrir le système d'équations miraculeux qui permettrait de prévoir à coup sûr leur comportement à tout instant. Il est de concevoir un système médiateur artificiel, basé sur de très nombreuses entités informatiques autonomes, évolutives et coopératives. Ces entités produiraient, par le fait de leur capacité à s'auto-organiser, un état semblable à celui du phénomène réel étudié mais qui serait lisible et de plus totalement paramétrable. Il s'agit de faire correspondre à un phénomène réel complexe, que l'évolution a produit sur de longues périodes de temps par adaptations - sélections, un système de même nature organisationnelle mais artificiel, fondé sur des processus calculables, paramétrable et finalement compréhensible. Alors, que seront l'architecture de tels systèmes, leur structure et l'élément minimal leur permettant de s'organiser en se calculant d'une manière semblable au phénomène réel complexe qu'ils viseront à représenter ?

#### Poser le problème

Quel est l'élément minimal, la brique de base, à partir de laquelle construire un système qui doit ressentir et penser artificiellement, un système générateur

d'émotions et éprouvant des sensations. Concevoir un système, c'est d'abord le modéliser, c'est-à-dire décrire sa structure et ses fonctions, afin de préciser comment on peut et on va le fabriquer. Ensuite, il faut le construire effectivement, ce qui est un travail éventuellement très minutieux. Finalement, on le mettra en expérimentation dans son environnement d'usage puis on le livrera aux utilisateurs. Il ne s'agit pas pourtant d'un processus de réalisation linéaire. L'expérimentation nécessitera généralement de reprendre toute la démarche, depuis l'analyse jusqu'à la réalisation en cours. C'est un processus propre à l'Intelligence Artificielle: construire des systèmes d'une façon dite "incrémentielle".

Mais pour construire, il faut savoir quel est précisément le problème posé et le modèle. Or l'énoncé de ce problème particulier, réaliser une machine pensante, est évidemment difficile. On procédera par étapes et la première étape, en s'inspirant du paradigme de l'évolution du vivant, est de concevoir un système capable de produire en lui des émotions. Le vivant dit inférieur, même s'il ne génère pas de pensées conceptuelles, éprouve en effet des émotions. Ensuite, partant de ce premier système, il sera possible en l'augmentant, de concevoir un système générant des pensées conscientes. Ainsi semble avoir procédé le vivant dans son évolution.

Il s'agit donc de concevoir d'abord, avec des entités spécifiquement informatiques, c'est-à-dire en utilisant des concepts venant de l'informatique et qui se réalisent essentiellement sous forme de programmes, un système qui génère des émotions. Il faudra pour ce faire que le système ait un corps, qu'il ait des capteurs lui permettant de sentir les changements physiques de son environnement et qu'il ait des effecteurs lui permettant de mouvoir des organes. Ce corps lui permettra d'avoir une sensibilité, appréciée comme telle, aux phénomènes externes (exogènes) et aux mouvements internes (endogènes). Mais remarquons à ce stade qu'une émotion n'est pas ressentie consciemment, elle consiste seulement en la réaction automatique d'un corps sensible à des stimuli.

Il faudra ensuite qu'à partir de ce corps sensible, le système puisse produire des sensations, c'est-à-dire qu'il puisse éprouver du plaisir, de l'ennui, de la peine, et enfin qu'il génère des faits de conscience, c'est-à-dire qu'il se représente, pour son compte et non pour le compte de son constructeur ou de son utilisateur, certaines choses de l'environnement, en conférant à ces choses de la signification dans l'espace et dans le temps. Bien plus, il faudra que le système ait envie de penser et ceci de façon aussi constante que possible. Ce dernier point, le souci de penser, sera le problème le plus délicat à résoudre. Comment un système artificiel, construit avec des composants physiques, des capteurs et leurs systèmes de commande, puis des processeurs qui ne réalisent en fait que des calculs même s'ils sont distribués, peut-il avoir envie de penser à quelque chose, ici et maintenant, et finalement comment fait-il pour penser ? Comment peut-il vouloir, pour penser, se servir de son expérience présente et de sa mémoire ? Cette mémoire, dans ce cas, n'aura pas grand-chose à voir avec la mémoire à base d'adresses des ordinateurs classiques,

mais elle sera sémantique et sélective pour retrouver en mémoire des sensations, des événements, des concepts.

Pour progresser, nous devons approfondir toutes ces questions. Il s'agit, ne nous le dissimulons pas, d'un enjeu considérable.

Qu'est-ce, d'abord, que penser artificiellement ? Je ne me pose pas la question, quand je pense, de savoir ce qu'est pour moi la pensée. Je ne saurais d'ailleurs pas y répondre. Pour un système artificiel qui aura à penser artificiellement, on doit par contre se poser la question et y répondre. Un tel système, en effet, doit être conçu dans sa totalité, construit avec une précision aussi grande que possible et finalement se comporter d'une façon que l'on pourra, en principe, décrire et suivre.

Mais comment ? Est-ce que penser signifie raisonner ? On sait ce qu'est le raisonnement et même comment raisonner en mathématiques et en informatique. C'est ce que l'on appelle produire des inférences, c'est-à-dire travailler sur l'équivalent du logos, déduire, induire à partir d'axiomes et en utilisant des règles. L'Intelligence Artificielle s'y applique avec des résultats très satisfaisants depuis de nombreuses années. Mais dans ce cas, les bases du raisonnement sont fournies à la machine, elles sont programmées à l'avance et la machine ne fait que les utiliser, dans le bon ordre et rapidement. Il faut bien voir dès ce stade ce qui distinguera un système artificiel conscient d'un automate préprogrammé. Le chien Aibo, que Sony avait commercialisé jusqu'à l'année 2005, disposait de dizaines de milliers de sous-programmes, écrits par des centaines de programmeurs. Ces programmes lui permettaient de faire face à des situations prévues à l'avance: par exemple remuer la queue pour donner l'apparence du contentement quand son propriétaire lui caressait la tête. Mais il s'agissait d'un comportement stéréotype sur le mode stimulus-réponse n'autorisant aucune adaptation et n'ayant aucune signification symbolique – pour le chien Aibo tout au moins, car le propriétaire du chien au contraire pouvait se laisser tromper et prêtait au robot des sentiments qu'il n'avait pas. Par contre, notre système artificiel se trouvera confronté à des objets du monde qu'il devra distinguer, par construction, mais dont il ne saura rien en dehors du fait qu'il se heurte à eux, qu'il les appréhende et qu'il doit les concevoir. Quel sera alors le rapport entre ces choses et les pensées qu'il développera à leur propos?

La question fondamentale est bien la suivante : est-ce qu'on peut ramener le fait de penser tel que nous le comprenons quand il s'agit de nous à la production d'un système informatique, fut-il complexe ? La définition d'un système, en informatique, est nécessairement précise : c'est un ensemble fini d'entités de calculs en relation, qui forment un tout, qui sont agrégées et communicantes. Le système est entouré d'un bord ou interface qui lui permet de communiquer avec son utilisateur et avec le système d'exploitation (operating system) gérant toutes les activités sur le hardware de la machine. Ce bord le délimite et assure sa

permanence. Un système informatique existe sous forme de programmes, est fini et est entièrement descriptible avant, pendant et après exécution des programmes.

Notre réponse à cette question fondamentale sera affirmative : il sera possible de ramener le fait de penser à l'exécution d'un certain système, mais il nous faudra étendre la notion de système. Pour ce faire, nous allons choisir la voie systémique et constructiviste : nous posons que nous pouvons construire d'une certaine façon un système qui va générer des états très particuliers que nous appellerons des pensées artificielles et qui aura une certaine conscience de lui-même. Il s'agira d'un problème d'informatique de calculabilité, n'ayant à ce stade que peu à voir, sinon rien, avec la conscience humaine.

Nous évacuerons donc de cette façon le débat sur la relation entre corps, pensée et conscience humains d'une part et pensée artificielle, conscience artificielle d'autre part. Nous construirons un ensemble d'automates très évolutifs qui produiront certaines choses que d'aucuns appelleront de la pensée, d'autres non. Peu nous importera, pourvu que notre système fonctionne et se comporte comme nous, sinon aussi bien que nous.

Remarquons que le même raisonnement devrait s'appliquer à la pensée animale. Plutôt que de décider a priori que les animaux ne peuvent héberger une conscience humaine – ce qui paraît d'ailleurs évident – nous devrions simplement nous demander comment ils font pour réaliser des choses qui pour nous relèveraient de la conscience, et même d'une conscience assez subtile et élaborée. Nous découvririons alors peut-être des phénomènes qui nous échappent encore.

## **La conception du système**

Nous avons évoqué jusqu'à présent la nécessité de faire appel à l'informatique pour concevoir notre système. Mais cela ne suffira pas. Le système devra disposer d'un corps, et cela concerne la robotique. Depuis une dizaine d'années, on estime que le travail sur les émotions et la pensée artificielle ne peut être fait qu'avec des machines dotées d'un corps muni de très nombreux capteurs et effecteurs (des milliers sinon plus). On a besoin ici de corporéité : il ne peut exister de conscience, même artificielle, en dehors d'un substrat matériel qui lui donne un ancrage et sa raison d'être. Les robots nécessaires seront malheureusement encore des robots rares et chers, des robots de recherche très éloignés des habituels robots ménagers. Mais les coûts dans l'avenir en diminueront très vite.

À partir de ce corps, il faudra concevoir un système émotionnel qui lui permettra d'acquérir la notion de corps et ne pas se contenter d'être un assemblage de circuits et de composants électroniques sans unité. Ce sera un système générateur de ce que Damasio (Antonio Damasio, *Le sentiment même de soi*,



éditions O. Jacob, 1999) op.cit.) appelle le proto-Soi. Enfin, on posera d'emblée la question pratique de la faisabilité, parce qu'une fois ce système conçu, il faudra pour le réaliser vérifier qu'il est effectivement implémentable ou portable de manière distribuée sur des grappes de processeurs liés aux systèmes de contrôle/commande du robot.

Avant d'en venir là, précisons quelques points.

- Le système ne sera pas seulement réactif. Il ne fera pas que réagir et optimiser une trajectoire, comme un missile guidé. Il faudra qu'il prenne des initiatives lui appartenant en propre et qui ne seront pas seulement le fait du hasard ni d'un déterminisme préprogrammé. Il aura des émotions et éprouvera des sensations, des sensations ressenties pour son propre compte. Il aura une certaine structure et une certaine organisation, un peu semblable à la nôtre, mais pour lui artificielle et calculable. Ceci veut dire que faire, en informatique, des systèmes organisés dépassant le niveau de la réactivité, ne sera pas du même ordre que faire des systèmes de type contrôle/commande pour des suivis de trajectoires, systèmes totalement rigides, prédéterminés, entièrement prédictibles. Dans le cas de notre système, il faudra sortir de ce cadre et aborder une sorte d'informatique encore peu répandue, typiquement adaptative et évolutive. Heureusement, les réalisations actuelles en Intelligence Artificielle évolutionnaire et autonome ont bien éclairé la voie.

- Quels seront les mouvements et les comportements motivés d'une telle architecture fortement plastique et évolutive lorsqu'elle sera activée ? Quelle relation pourra-t-on établir entre son comportement, ses pensées et la formulation langagière grâce à laquelle elle communiquera éventuellement avec nous ? Un tel robot ne devra pas seulement produire, pour celui qui l'observe, l'impression qu'il marche, qu'il fait certaines choses, qu'il a envie d'aller ici ou là. Non. Il devra véritablement disposer d'une intention d'agir avant que d'agir. Pour cela, il devra disposer d'une intentionnalité, c'est-à-dire générer des faits de conscience avec intention, attribuer des significations à propos de choses et d'événements distingués par lui dans la réalité de son monde et avoir envie de s'interroger sur elles et eux.

Ceci posé, la conception du système fera appel à des bases concrètes que la robotique peut d'ores et déjà produire (fut-ce à des coûts encore élevés). On s'appuiera entièrement sur la corporéité d'un robot. Idéalement, nous l'avons dit, il faudrait posséder un système aussi complexe que possible, disposant de milliers de capteurs, de senseurs et d'effecteurs. Il faudra aussi recueillir beaucoup d'informations venant de tous ces capteurs et évaluer leurs variations continues, pour pouvoir contrôler un système qui va introduire de multiples variations d'informations, lesquelles vont, par fusions, déclencher des émotions, des sensations en activant de multiples procédures. En pratique, pour des raisons d'économie, on pourra rechercher des solutions faisant appel à des systèmes

distribués sur des grappes de processeurs, sur le mode du "grid" ou réseau de calculateurs. Ces problèmes, sans être triviaux, sont aujourd'hui solubles.

Il faudra ensuite concevoir l'architecture logicielle générant les émotions. Pour cela il faudra se représenter clairement ce qu'est une émotion artificielle générée par une architecture logicielle organiquement liée à l'ensemble des capteurs d'un corps de robot. Une telle émotion artificielle ne pourra être définie que par ce qui la constitue et son concept, comme dans tout organisme. Une émotion est, au sens d'une entité concevable et calculable, un mouvement organisationnel dans un système dynamique complexe, qui a une certaine durée, une certaine intensité, une certaine fréquence et surtout une certaine régularité. Il s'agit d'un mouvement réorganisationnel d'entités. Cela peut se représenter par une courbe d'effet dans un espace.

Pour comprendre, existentiellement, ce qu'est une émotion et ne pas se contenter de l'effet qu'elle produit, il faut abstraire. Il faut représenter par une fonction, à découvrir, dans un espace, également à découvrir, ce qui donne l'existence au robot. La focalisation sur ce qui permet l'existence, plutôt que sur ce qui produit des effets, doit pouvoir fournir la solution au problème de la génération d'émotions et de faits de conscience.

Mais pour cela, il faut de plus préciser ce qu'est l'intention. Contrairement à l'émotion, qui est une réaction systématique à un stimulus, l'intention signifie "vouloir penser à quelque chose et s'engager vers la formulation de cette pensée". La question est toujours la même à travers les siècles, de Parménide à Martin Heidegger : "qu'est-ce qui nous amène à penser ce que nous pensons lorsque nous le pensons?". Transposer cette question dans le domaine du calculable entraîne une conséquence majeure : il s'agit de savoir si la notion de pensée est conceptuellement calculable ou non.

### **Le système générateur d'émotions**

Pour répondre à cette question, demandons-nous ce que l'on trouvera à la source d'un système générateur d'émotions artificielles tel que nous le concevons ? Ce ne sera pas un système constitué de neurones formels, car il faudrait lui donner une taille telle qu'il ne serait pas envisageable de le construire effectivement. De la même manière, il semble bien qu'il soit impossible de résoudre le problème de la cause permettant au cerveau de produire de la pensée en se focalisant sur le niveau moléculaire (celui des synapses du neurone vivant et des neurotransmetteurs par exemple). Pour tout problème, il faut trouver le bon niveau (le bon grain et le bon modèle) permettant sa résolution.

En fait, pour trouver l'élément minimal dans le système, il ne faut pas se placer au niveau de la structure physique. Au delà de la matérialité du corps, il faudra créer une architecture plastique faite d'entités virtuelles, d'entités informatiques autonomes, communicantes et fortement agrégatives, dont l'organisation permettra la génération des émotions et des pensées. Ceci permettra au système artificiel de s'organiser, de s'auto-organiser, d'évoluer.

L'élément de base qui nous nous proposons de retenir est assez récemment apparu dans la culture informatique. C'est l'agent logiciel. Un agent logiciel n'a strictement rien à voir avec un agent humain (ni d'ailleurs avec l'agent intelligent généralement évoqué par l'Intelligence Artificielle classique). C'est une entité d'action au niveau informatique, qui est capable d'agir sur son environnement (également logiciel). C'est-à-dire que c'est un certain type de programme et rien d'autre. Il est composé d'un ensemble d'instructions et de structures de données compliquées, mais c'est quand même un programme. Et c'est un programme qui est proactif, c'est-à-dire qui travaille pour son compte, pour atteindre des buts qu'il peut réviser mais qu'il s'efforcera toujours d'atteindre. C'est donc un programme particulier qui, contrairement à ce que fait un programme classique, fera des actions pour son compte, avec des objectifs rationnels et qui aura la possibilité de s'associer à d'autres programmes semblables. Il pourra évoluer en modifiant sa structure, en communiquant avec d'autres agents, en utilisant éventuellement des informations provenant des capteurs du robot ou en envoyant d'autres informations aux effecteurs.

### **Les agents aspectuels**

L'entité logicielle de conception, pour un système de ce type, se place donc à un niveau qui est très supérieur à celui de l'instruction ou de la variable numérique. On peut, si l'on veut, dire que cette entité est du même ordre que les groupes de neurones du cerveau exerçant des fonctionnalités précises. C'est en fait un petit système autonome et communicant, qui pourra s'exécuter sur un ou des processeurs, à un certain moment et pour la simple raison qu'il doit atteindre l'un de ses buts. Dans la terminologie choisie par Alain Cardon, cet agent a été nommé agent aspectuel (1), pour bien faire remarquer qu'il désigne des aspects symboliques très locaux de l'activité du système. Cet agent est en effet doté d'un caractère symbolique précis. Il exprime ainsi ce qu'il fait, et cette action est quelque chose que nous pouvons interpréter, qui nous permet de le suivre et qui, surtout, correspond à ce qu'il va faire effectivement dans le système. Cet agent est donc aussi une forme d'action, une véritable activité parmi toutes les autres activités des autres agents. Nous ne la comprenons pas obligatoirement tout de suite, mais nous pourrons la découvrir après coup.

Il faut bien distinguer l' «agent », ainsi conçu, de l' «objet », entité utilisée depuis longtemps en informatique. L' « objet » est une entité définie de manière rigide, qui ne change pas de comportement et dont l'autonomie est insignifiante. L'objet est structurellement lié à tous les autres de façon fixe dès la conception. L'agent, par contre, est une entité d'action disposant d'une signification propre et d'un très fort degré de liberté allant jusqu'à la capacité d'évolution structurelle et de reproduction.

Alors, en utilisant de très nombreux agents, on obtient une autre entité logicielle, de niveau organisationnellement supérieur, et que l'on appelle une organisation d'agents. Dans cette organisation, les agents peuvent socialement s'activer, communiquer, s'agréger, se séparer, s'annihiler, se commander, se multiplier, se structurer, se hiérarchiser, fusionner, se reproduire, évoluer ... Chaque organisation d'agents présentera ainsi des caractères comportementaux spécifiques, propres à elle et finalement définissant tout en la reflétant l'action de cette organisation dans son milieu. Cette organisation ressemble, si l'on veut, à un très vaste ensemble de traits conceptuels, ayant de multiples possibilités de réorganisation. Pour qu'un tel ensemble dégage une cohérence, il faudra qu'il se structure, qu'il s'organise afin que sa forme ait un sens. Et il devra évidemment se réorganiser de lui-même, sans supervision.

L'agent aspectuel est un programme qui dispose de connaissances sous forme de règles du type « si...alors » : "si telle chose est dans tel contexte, alors réaliser tel effet à tel endroit à la suite de cet état". L'effet ou l'action de l'agent prend la forme de communications et d'actions avec des objets de l'environnement ou avec d'autres agents. L'agent fait montre d'un comportement strictement rationnel, qui lui est donné par un automate comportemental à états lors de la construction. Mais cet automate peut évidemment évoluer ...

En fait, cet agent n'a rien de mystérieux. Il s'agit d'un programme, sophistiqué mais équivalent à une machine de Turing, c'est-à-dire qu'il est un élément faisant des calculs. Ce n'est certainement pas seulement avec un tel agent que l'on pourra construire une machine pensante.

Alors en quoi une organisation d'agents de ce genre pourrait-elle représenter une émotion ? Pour représenter une émotion, il faut faire appel à des processus en cours et pas seulement aux effets de ces processus. Il faut représenter de l'action en train de se dérouler et pas seulement s'intéresser aux résultats, ce que l'informatique traditionnelle ne sait pas bien faire. L'informatique réalise habituellement des traitements visant essentiellement à produire des résultats de calculs, pour que ceux-ci soient utilisés es qualité. Mais le calcul, en ce qui concerne une émotion, est en train de se faire. C'est l'action même de calculer qui sera l'émotion artificielle que l'on souhaitera représenter. Une émotion est un processus complexe, qui a une

certaine ampleur, une certaine durée, une certaine modulation. Comment procéder en ce cas ?

Il faut dans le modèle représenter des processus en cours d'exécution, c'est-à-dire des agents logiciels en train d'opérer, d'agir, de communiquer. On retrouve, ce qui est significatif au demeurant, les travaux de la neurologie moderne montrant, grâce aux images de la caméra à positrons, les aires neuronales du cerveau en train de s'activer et de communiquer entre elles. Dans le cas décrit ici, il faudra représenter une fonction en cours de calcul et pas seulement le résultat de la fonction calculée : une fonction qui se calcule, pendant qu'elle calcule ! Ceci n'avait jamais été abordé jusqu'ici, et constitue l'originalité du système (2).

Pour représenter à la fois les agents et les actions organisationnelles des agents, on devra de nouveau faire appel à une organisation d'agents logiciels. Il faudra en fait définir une structure d'auto-observation des calculs en train d'être opérés par les agents aspectuels, une structure capable de suivre l'état courant des multiples calculs.

Résumons. Nous disposons d'une organisation d'agents dits aspectuels correspondant aux fonctions typiques du robot en train de capter des informations, par exemple recevoir telle information sonar, de percevoir et analyser tels mots entendus, d'opérer sur des faits, des informations et des connaissances. Tous ces calculs sont effectués par des dizaines de milliers d'agents aspectuels qui sont distribués sur des grappes de processeurs. Mais avec cela, on ne fait pas grand-chose au niveau émotionnel. Même si le système produit quelques bons résultats, il ne nous informe pas de la façon dont il les produit, il ne nous précise pas le pourquoi des calculs qu'il réalise et la raison de la variation qui produit l'intensité de l'émotion ? Or nous devons le savoir pour jouer de l'état organisationnel des agents aspectuels et moduler le déroulement d'une émotion.

L'exigence que nous exprimons est nouvelle en informatique. Elle découle des théories de la morphogenèse émises par le regretté René Thom il y a maintenant quelques années : observer les formes comportementales d'une organisation d'entités proactives et exprimer la signification des déformations de ces formes. Face à une organisation d'agents qui s'activent, il faudra observer les comportements de chaque agent, ceux des groupes et ceux de l'ensemble afin de faire apparaître la forme de cette organisation. En d'autres termes, il faudra faire apparaître ce qu'Alain Cardon appelle la morphologie d'un paysage d'agents en cours de réorganisation.

## **Les agents morphologiques**

Pour cela, on organisera l'autocontrôle d'une organisation massive de centaines de milliers d'agents par une autre organisation, par des agents que nous appellerons agents morphologiques, analysant en ligne les caractères organisationnels du premier ensemble d'agents en train de se réorganiser. On retrouve ici la notion de forme d'une organisation en train de se mouvoir, en train de s'activer. On est en présence d'un système double ou, si l'on veut, à double face. Il s'agit en fait de deux systèmes fortement couplés : celui qui fait les calculs et celui qui observe les calculs en train de se faire, qui en tire des conclusions et qui ce faisant communique avec le premier.

Qu'est-ce qu'une organisation d'agents logiciels en train de s'activer, en train de faire des calculs ? Imaginons la souplesse d'une machine massivement parallèle dotée de processeurs multiples occupée à observer les activités et les inactivités de ses différents processeurs et processus. Elle fera apparaître par exemple l'émergence d'un comportement et la vitesse de son développement, comment ce comportement évolue depuis l'éveil jusqu'à l'action en coopération, comment il atteint ses buts. On observera aussi l'activité sociale de l'agent. Est-il seul ou non, trouve-t-il ou non des alliés au sein du système ? Que représente-t-il de singulier, de non partagé et qui n'aura aucun effet dans l'activité des autres agents ? On observera le poids et la domination d'un agent par rapport aux autres : en quoi un agent est-il leader ou soumis... ?

On se trouve en fait devant une population d'agents qui évoluent comme des êtres vivants sur le mode classique de la compétition darwinienne. Ils doivent notamment, pour survivre, s'agréger afin de dégager, par leur regroupement, un ensemble de caractères nouveaux. Mais parce que ce sont des agents actifs, proactifs, communicants, autonomes et qui peuvent agir, ils présentent un certain nombre de caractères précis. Ceux-ci définissent les dimensions d'un certain espace que nous avons appelé espace morphologique, caractérisant l'organisation en train de s'activer.

On obtient ainsi une organisation d'agents aspectuels qui agissent en constituant un système lui-même capable de mesurer continuellement sa propre activité organisationnelle. Ceci pourrait être comparé à une société d'individus organisée pour permettre à un formidable Big Brother de la regarder en train de s'activer, point par point mais aussi globalement, afin d'en estimer les tendances, la cohérence, les agrégations et les ruptures... Mais il s'agit d'une société d'agents et les agents aspectuels sont munis de sondes pour pouvoir être observés !

Les livres précités d'Alain Cardon définissent précisément ce qu'est la détermination d'une morphologie : une organisation d'agents avec des caractères qui se distinguent, que l'on peut projeter sur un espace à plusieurs dimensions. Le schéma suivant donne une image simple de ce que l'on peut obtenir par projection dans un espace de dimension 2.

On voit visiblement qu'il se passe un certain nombre de choses dans l'organisation, qui peuvent la caractériser comme étant cohérente ou non cohérente, agrégative, se coupant en morceaux... On obtient là une représentation géométrique de l'évolution de l'organisation aspectuelle, c'est-à-dire finalement de la façon dont le robot va réagir globalement aux événements qu'il est en train d'affronter. Remarquons que le terme "géométrique" veut dire "en dehors de toute sémantique". Les agents aspectuels ont de la signification de par leurs buts et comportements, mais leur organisation est observée de manière essentiellement géométrique, en y recherchant des formes.

Nous avons donc conçu un système composé de plusieurs organisations d'agents, toutes couplées par des structures de couplage particulières qui seront aussi des agents :

- des agents d'interface, qui vont agir sur les effecteurs puis sur les senseurs du robot,
- des agents aspectuels, qui vont donner de la signification aux activations,
- des agents de morphologie, qui vont regarder comment évolue l'organisation des agents aspectuels,
- des agents de synchronisation, qui vont permettre les couplages entre les différentes parties des organisations précédentes.

Nous allons alors établir un bouclage fort entre l'organisation des agents aspectuels qui agissent et l'organisation morphologique qui observe cette activité. De plus, cette organisation morphologique, observant l'organisation aspectuelle, va pouvoir modifier l'action de cette dernière : le processus de couplage fort qui va s'établir entre les deux organisations permettra une rétroaction débouchant sur un certain état d'équilibre entre les deux organisations. Nous définissons plus précisément un couplage de groupe à groupe entre ces deux organisations.

Ainsi sera produit l'équivalent d'une carte émotionnelle selon Damasio, une carte émotionnelle artificielle composée d'une organisation d'agents aspectuels, d'une organisation morphologique qui la contrôle, avec un processus de couplage fort entre ces deux organisations réalisé par une organisation d'agents de synchronisation. Ceci formera un processus miroir qui fera que lorsque l'organisation aspectuelle commencera à s'activer à la suite de la réception d'un stimulus, l'organisation morphologique détectera ce mouvement et éventuellement pourra l'empêcher, ou au contraire l'amplifier, ou encore, dans une autre partie du système, pourra faire agir des modules de reconnaissance de formes ou de toutes autres structures. Cela permettra un auto-contrôle qui partira de l'organisation aspectuelle par l'évaluation de sa conformation, pour la faire aller dans certaines

directions, pour donner des tendances précises à ce qu'engage le comportement aspectuel. Il s'agira donc bien d'un processus d'auto-contrôle et de co-activation régulant de lui-même l'activation des agents aspectuels. On peut retrouver là un des processus de base supposés de la conscience humaine générée par des neurones réagissant à des informations sensorielles et réagissant en retour sur ces mêmes neurones ou sur d'autres pour influencer leur comportement.

Mais il faut bien voir que l'organisation double [aspectuelle/morphologique] ainsi décrite ne sera qu'une carte particulière dans le système. Dans la machine pensante, il y en aura des centaines sinon des milliers analogues. De ce fait, cette carte émotionnelle, ce composant adaptatif, auto-contrôlé, centré sur un rôle (que l'on a défini comme une même catégorie dans les agents aspectuels), coopératif et synchronisable avec d'autres composants similaires, participera à l'adaptation d'ensemble du système.

Passons maintenant à la structure générative des émotions. Nous disposons d'un système comportant les capteurs et les effecteurs du robot, des éléments d'interface, c'est-à-dire des agents spécifiques qui prennent sans cesse les changements d'informations venant de l'extérieur, et un ensemble de cartes émotionnelles artificielles, synchronisables par des agents de synchronisation. Ces éléments vont avoir un rôle considérable dans le système : ils vont faire basculer celui-ci, en quelque sorte, au-devant de lui-même dans son fonctionnement. Ils vont permettre son action anticipatrice.

Les agents de synchronisation vont en fait prendre les décisions suivantes : ces cartes qui agissent et qu'ils observent peuvent avoir la même morphologie; ces groupes morphologiques peuvent être semblables, cette carte-là présente un aspect organisationnel semblable à celui de cette autre, bien que les sémantiques données par les agents aspectuels soient totalement différentes. Les éléments de synchronisation pourront opérer des mises en conformité, des associations de cartes par unification de leurs rythmes d'activation. Par exemple, ils coupleront systématiquement des cartes représentant du son et de l'image, ou des cartes représentant un mouvement repéré avec une information relative à la distance nécessaire à la fuite...

### **Le système générateur de pensée**

Le système générateur d'émotions étant défini, on peut s'intéresser au système générateur de pensées. Ce système est beaucoup plus compliqué que le système générateur d'émotions mais l'enjeu est considérable : définir un Soi artificiel et inciter un système à penser de lui-même aux choses du monde qu'il peut se représenter. Dans le système générateur de pensées, le problème de la conception



est particulièrement difficile à résoudre. Nous avons retenu une architecture inspirée de celle du système générateur d'émotions, mais il y a beaucoup plus dans le Soi que dans le proto-Soi, et l'architecture devra être différente. Il faudra y introduire les concepts calculables qui correspondent à la solution de problèmes difficiles comme par exemple ce qui engage effectivement un système à réorganiser ses composants pour générer une représentation intentionnelle d'une chose de son monde.

Et il faudra donc répondre à de nombreuses questions comme celles-ci :

- Comment se représente effectivement ce que l'on va appeler la pensée artificielle ? Qu'est-ce qu'une pensée artificielle ? Il s'agira d'une manière particulière d'organiser des entités, mais laquelle, comment et pouvant produire quel état ?

- Qu'est-ce qui peut conduire un certain organisme artificiel à penser ici et maintenant à telle chose ? Penser ce sera, pour le système, organiser des ensembles d'agents, des structures à double face organisation aspectuelle / organisation morphologique, synchroniser les organisations aspectuelles avec les organisations morphologiques, et ceci d'une façon très fine. Quand il s'agit d'une émotion, nous l'avons dit, le processus est assez clair. On reçoit un stimulus, on l'interprète et on provoque une réaction spatiale et temporelle dans l'espace morphologique des organisations d'agents. Ici, en ce qui concerne la pensée artificielle, il s'agit de provoquer le comportement des organisations d'agents, de préciser pourquoi il est amené à s'organiser de telle ou telle façon, c'est-à-dire à penser.

- Comment se représente et agit la mémoire des choses. Il faut construire une mémoire à la fois incrémentielle (capable d'acquérir en permanence de nouvelles informations) et qui soit aussi sélective, en permettant l'oubli ?

- Quel est le lien entre la structure représentant des émotions et celle générant des pensées ? Autrement dit, comment pourra-t-on aboutir à des sensations suffisamment fortes pour que le système sache vraiment, et de lui-même, qu'il les éprouve ? Les émotions ne sont que des réactions non perçues, alors que les sensations sont un ressenti effectif, perçu, représenté par quelque chose d'interne que l'organisme appréhende et qu'il peut utiliser pour générer des pensées.

- Et comment se génère et se représente le langage, comment, d'où et pourquoi viennent les mots perçus ou prononcés ? Comment telle formulation langagière s'attache-t-elle à une situation particulière ? Il faudra engendrer des mots à partir de formes de pensées artificielles et permettre au système de comprendre des sons tels des mots dotés de significations.

Pour toutes ces questions, il faudra pouvoir répondre en proposant des structures calculables, sinon on se bornera, comme d'ailleurs nous le faisons en posant ces questions, à proposer des métaphores. Mais les métaphores ne conduisent pas à grand-chose en informatique.

Les réponses à ces questions figurent dans le modèle longuement développé par Alain Cardon et détaillé dans l'ouvrage sur la conception d'une machine pensante. Il s'agira d'une architecture basée sur la notion de composant adaptatif (qui va ressembler beaucoup à la conception des cartes émotionnelles), qui sera aussi proactif, auto-observateur et évolutif. On la nommera "structure de signification".

Le système capable à générer des pensées artificielles sera formé d'un certain ensemble de ces structures de signification synchronisées par des agents de synchronisation. Il sera de plus construit en co-construction avec le corps du robot, pour suivre le paradigme de la biologie du développement.

Ce sera donc un ensemble qui comportant d'un côté la carte corporelle avec des éléments d'interface, le système de prise d'informations, des agents aspectuels spécifiques chargés de suivre ces informations, et des agents de morphologie fortement couplés qui observeront sans cesse comment les agents aspectuels agissent, ce qu'ils font, ce qu'ils deviennent. C'est exactement de cette façon, notons-le au passage, que les décideurs politiques gèrent des villes ou des Etats : ils regardent ce que font les citoyens, analysent leurs activités et les mesurent pour prévoir et contrôler les mouvements globaux...On aura là la composante émotionnelle du système.

De l'autre côté, le système disposera d'une composante de signification. Celle-ci comportera des éléments qui ressembleront structurellement à la carte corporelle. Ce seront les structures de signification qui permettront de représenter de l'action signifiante pour le système, et ceci avec l'intention à procéder à la génération d'une telle action. La structure de signification est plus compliquée que la carte émotionnelle car elle utilise des éléments de mémoire.

Cette mémoire, distribuée dans des agents particuliers et dans les différentes structures de signification, va opérer la mise à distance (la prise de recul) du système par rapport à l'action immédiate dans son environnement, afin de le détacher de l'émotion réactive. Elle va pour cela retrouver et transformer en agents actifs des traces des scènes anciennes qui ont été vécues ou conçues.

On trouvera enfin dans ce schéma, ce qui est le point fondamental, le facteur ou la raison qui amènera la machine pensante à générer des pensées. Il s'agit des éléments dits d'anticipation, qui ont le rôle d'anticiper l'action des différentes organisations d'agents par rapport au mouvement global en cours. Pour un informaticien, faire penser un système consiste à l'obliger à produire un signe qui l'y

engage, qui l'amène à générer cette forme de pensée. Il ne s'agit pas de viser un but prédéfini, mais d'engager le système à adopter, dans son mouvement organisationnel, une posture particulière. L'intention à penser n'a pas de but initial à atteindre et ceci illustre bien la notion si forte de liberté à penser.

Cette anticipation, le système ira la chercher dans les morphologies précédemment présentées, dans les formes géométriques des structures de signification, ceci sans recours à aucune sémantique. Le système générera donc des signes, des signes au sens de la sémiotique de C.S. Peirce, qui donneront aux agents morphologiques des structures de significations pour que le système soit globalement incité à organiser son activité autour de cette signification géométriquement indiquée. On retrouve ainsi la définition de la visée, inspirée des travaux de Martin Heidegger ou de Paul Ricœur. L'activation, le fonctionnement de ce système, à partir de cette visée, constituera sa façon de penser intentionnellement.

Les éléments de synchronisation - ceux qui synchronisent les cartes corporelles par leurs relations - seront connectés au corps de la machine. Sinon, on se bornerait à réaliser une certaine machine de Turing, certes compliquée, mais machine de Turing cependant dont la fonction de calcul serait déjà complètement programmée. Or pour arriver à générer un système qui pense, qui pense pour lui-même en produisant des formes idéelles chaque fois nouvelles, il faudra que les éléments de synchronisation soient aussi capables d'interrompre drastiquement l'activation des structures de signification, afin de réaliser la synchronisation du système générateur de pensées avec le temps réel d'activation du robot, avec la réalité corporelle du corps de la machine située dans son monde et affrontant une suite ininterrompue d'évènements.

C'est ici que la corporéité est nécessaire pour produire la pensée, et qu'il est impossible de concevoir un système artificiellement conscient, avec des émotions et des sensations, sans corps. Sans corps, la pensée n'a pas de sens dans le monde réel et n'est pas de la pensée telle que nous l'entendons habituellement. Ceci dit, le corps n'est pas ici nécessairement fait de composants physiques localisés dans un endroit précis. Il peut être réparti dans différents composants, il peut par exemple être distribué et accessible via les nœuds d'un réseau. Mais le corps doit exister, avec la totalité de ses attributs corporels, senseurs et effecteurs notamment.

Le système proposé, bien que différent d'une machine de Turing, le sera partiellement cependant, puisqu'il s'agira en dernier ressort d'un système informatique et donc essentiellement calculable. Tous les éléments, les agents, seront des fonctions calculables qui seront calculées, même si l'on essaie de les faire muter, évoluer, cloner, se reproduire, selon les principes de la programmation évolutive ou génétique d'agents (qui met en compétition des algorithmes, les algorithmes évolutionnaires, afin de faire émerger le plus efficace de ceux-ci dans

une situation donnée). Ce seront des fonctions parfaitement définies, telles des machines de Turing. Mais l'ensemble, précisément, ne sera que partiellement une machine de Turing. Cela veut dire qu'il sera non prédictible, pas simplement non déterministe, mais non défini à l'avance dans ses perspectives calculatoires elles-mêmes. Le système calcule des fonctions en continu, qui sont des machines de Turing. Il produit sans cesse de nouvelles machines de Turing.

Le système dépendra du corps du robot, des effecteurs et des capteurs, de sa situation ici et maintenant dans le monde. Tous ces éléments n'appartiennent pas au domaine du calculable mais de l'automatique et de l'électronique, même s'ils induisent la génération, à certains endroits, dans certains agents, de nouvelles lignes de code non prévues. Ainsi se marquera l'ouverture du système sur la réalité temporelle du monde, et aussi ce que nous avons appelé plus haut, son autonomie.

Le système présentera un état fluctuant, c'est-à-dire qu'il y aura une relation systématique entre le système émotionnel et le système générateur de pensées : il éprouvera des sensations en continu même en générant des pensées très rationnelles. Il aura la sensation de son organisation, une espèce de sentiment de Soi puisqu'il pourra en jouer. Il ressentira globalement son organisation interne d'une façon analogue à celle dont localement les capteurs seront perçus et ressentis par les organisations d'agents des cartes corporelles qui s'activent sur les entrées d'information.

D'autres points ont été étudiés, dont on ne parlera pas ici, et qui font l'objet de publications dans l'ouvrage d'Alain Cardon: il s'agit notamment des prégnances, c'est-à-dire des raisons organiques qui conduisent le système à pouvoir penser de certaines manières : pulsions, motivations et tendances fondamentales. Là, on ne se place pas forcément dans le même registre que l'humain. A priori il n'y a pas de pulsions sexuelles chez un robot, mais il y a quand même des pulsions, qui sont représentées par des aspects morphologiques d'organisations d'agents...

## **Construire**

Pratiquement, pour construire une machine pensante ainsi définie, il faudra disposer d'un ensemble de très nombreux processeurs en grappe fortement couplés, formant un tout pouvant être embarqué et donc aussi miniaturisé que possible. Cela existe aujourd'hui, mais il n'y en a pas encore en France. On trouve ce que l'on appelle des clusters, mais pas de grappes de processeurs avec des liaisons à très haut débit implantées dans des cartes. Le tout devra fonctionner dans un parallélisme lui-même à très haut débit. Il faudra aussi calibrer les cartes émotionnelles selon les fréquences d'acquisition des informations, et calibrer les structures de signification, ce qui supposera l'appel à des ingénieurs spécialisés.

Mais plus généralement, il faudra rompre intellectuellement avec l'informatique traditionnelle qui vise toujours l'optimisation des procédures et la distribution contrôlée d'informations à ses utilisateurs humains. Il ne s'agit plus, nous l'avons vu, de faire des systèmes de guidage pour missiles de croisière qui trouveront leur cible à quelques mètres près après cinq mille kilomètres de course. Il s'agit ici d'une autre définition de la calculabilité consistant à représenter l'acte de penser sous une forme calculable, mais dont la calculabilité n'apparaîtra que par constat, une fois les calculs effectués ! Au départ, on ne saura pas ce qui va vraiment être calculé mais on disposera d'une visée: puis le système calculera, il sera modifié, les agents évolueront en permanence, se multiplieront, se cloneront, certains vivront, d'autres mourront. Les composants, cartes ou agents, s'altéreront, se modifieront automatiquement quand ils seront activés. On obtiendra donc un système qui, dès sa conception, ne sera plus complètement contrôlé. Ce système pourra même se révéler défaillant, comme certains êtres vivants. On pourra donc constater, après coup, que l'on a calculé telle ou telle chose, que l'on a produit telle ou telle forme organisationnelle correspondant à telle ou telle pensée artificielle, mais on ne pourra pas à l'avance annoncer ce que l'on va calculer. C'est ce qui se produit dans les processus de pensée que nous connaissons. On ne peut pas dire : j'ai l'intention de penser à cela que je ne pense pas encore. La pensée n'est pas prédictible, non plus que la raison de penser.

La réalisation concrète d'un tel système, dans ces conditions, supposera un travail pluridisciplinaire, associant informaticiens, roboticiens mais aussi neurobiologistes, psychologues, voire philosophes. Cette collaboration permettra de ne jamais faire le choix de la facilité, en privilégiant tel développement et en perdant de vue les autres. Plus généralement, vu les ambitions du projet, il ne serait pas concevable que le travail soit confié aux seuls informaticiens, en négligeant les apports des sciences humaines.

Terminons ce survol de la conception d'une machine pensante en rappelant ce que nous avons plusieurs fois souligné mais qui est facilement oublié vu l'aspect contre-intuitif d'une telle notion : à la différence de la pensée humaine (au moins quand elle s'incarne dans des individus) les systèmes de ce type peuvent être multi-corps. Ils peuvent avoir un système émotionnel et de conscience loin du système physique avec lequel ils sont pourtant en liaison étroite. Ce sont donc des structures distribuées qui, mal employées, pourraient être dangereuses, car il n'existe pas encore de comité d'éthique en informatique et les perspectives d'applications militaires de tels systèmes sont déjà très nombreuses et inquiétantes.

Nous avons vu cependant que les besoins civils et pacifiques sont potentiellement plus considérables encore. L'avenir de tels développements devrait être d'une richesse aujourd'hui inimaginable, par exemple en jetant les bases d'une conscience artificielle de type humain projetée hors de l'environnement terrestre,

entre satellites robotisés notamment. Il ne faudrait donc pas décréter d'absurdes moratoires aux recherches. Il faudra par contre, c'est l'un des buts de ce livre, que les citoyens s'intéressent à ces questions et participent à leur développement.

## **Discussion**

L'exposé que nous venons de présenter pourra faire l'objet de critiques ou tout au moins de demandes d'éclaircissements. Voyons certaines d'entre-elles.

- Les spécifications fonctionnelles proposées ne sont-elles pas déduites, idéalement, de ce que l'on sait ou croit savoir aujourd'hui du fonctionnement du cerveau. Dans ce cas, la conscience artificielle esquissée ici devrait donner satisfaction puisqu'elle serait censée reproduire ce qui marche bien au niveau de la conscience humaine ; par exemple être capable d'observer le comportement du corps ou ceux de modules partiels de conscience afin d'en tirer un diagnostic global correspondant à la conscience de soi à tel moment. Ce diagnostic permettra en retour d'émettre des ordres destinés au corps afin d'améliorer la situation du sujet, sur le moment ou pour le futur. Mais le problème ne consiste pas à se représenter ce que fait, plus ou moins bien, le cerveau humain conscient. Il est de réaliser les dispositifs informatiques capables de produire les mêmes résultats en s'adressant, non à un corps biologique, mais à un corps de robot. Or là nous n'avons aucune indication concernant la faisabilité ou l'efficacité des dispositifs calculatoires informatiques évoqués.

On répondra à cela que si le système conçu par Alain Cardon n'a pas été encore programmé dans sa totalité et ne peut donc pas faire l'objet d'un démonstrateur, il a été testé en partie et s'est comporté conformément à ce que l'on en attendait. Le fait qu'aucun bailleur de fonds français n'ait voulu entreprendre la réalisation complète du projet ne signifie pas que celui-ci ne soit pas viable. Il est déjà intéressant de disposer d'un modèle théorique complètement développé qui permette aux personnes désireuses de se représenter ce que pourrait être une conscience artificielle d'en avoir une vue détaillée. Pour nous, nous estimons que ce modèle représente un complément indispensable à ce que nous avons écrit précédemment dans ce livre concernant la conscience humaine. Pour le reste, on ne peut demander à un inventeur soumis à un fort espionnage économique d'exposer sans précautions toutes ses solutions.

- N'obtiendra-t-on pas, si les spécifications fonctionnelles résumées ici étaient complètement développées au sein d'un système en vraie grandeur, un robot non maîtrisable par ceux qui voudraient l'utiliser. Alain Cardon insiste en effet sur un point capital en ce qui concerne une pensée consciente, quelle qu'elle soit. C'est

qu'elle ne peut se voir commander ce à quoi elle pensera ici et maintenant. Elle surprendra toujours, aussi bien l'entourage que l'entité qui en est doté. On peut concevoir alors que des industriels ou des militaires souhaitant disposer de robots asservis ou autonomes dans d'étroites fourchettes seulement, ne veuillent pas développer de tels robots. Ils en auraient plus de soucis que de satisfactions. Dans le même esprit d'ailleurs un chef d'entreprise ou un officier préférera souvent avoir des subordonnés contrôlables que capables de trop d'autonomie, fut-elle intelligente. Pour tirer parti d'un robot ou de systèmes de robots pleinement autonomes, il faudrait en effet réorganiser complètement les chaînes de production ou de commandement, comme plus généralement l'interface homme-robot. Mais alors, si cet effort était entrepris, les bénéfices en retour pourraient être considérables.

1 : Aspectuel : l'agent exprime des « aspects » significatifs du système. On pourrait dire, par transposition, que les mimiques inconscientes du corps humain sont des agents aspectuels révélant certains faits de conscience : la timidité, la colère, etc.

2 : Pour prendre une comparaison simple, nous dirions qu'il s'agirait de modéliser le fonctionnement d'un coureur en train de battre un record, au lieu de se borner à enregistrer le résultat final qui est le record battu.

## Chapitre 4, Section 2 :

### Les recherches de Sony-CSL Paris

Les psychologues et linguistes évolutionnaires considèrent généralement que c'est en développant des échanges symboliques codés, sous forme de langages ou proto-langages, que les individus ont progressivement acquis la conscience de soi. Le phénomène s'est développé sous des formes variées au sein de très nombreuses espèces animales pour qui la vie en petits groupes constituait la forme habituelle d'occupation des territoires. Mais il n'a pris de forme systématique que chez les hominiens, pour des raisons dont la cause première est encore objet de discussions entre paléo-anthropologues. On admet cependant que l'aptitude du langage parlé (au-delà des gestes et mimiques porteurs de significations codées) à favoriser la cohésion et la survie adaptatives des hominiens a été telle qu'elle a entraîné un accroissement considérable de la masse cérébrale et des aires dédiées au traitement de l'information symbolique.

Dans la suite du langage se sont très vraisemblablement installés les premiers échanges en miroir permettant aux individus de s'identifier comme semblables à leurs vis-à-vis, c'est-à-dire dotés d'autonomie et de l'aptitude à se représenter eux-mêmes à eux-mêmes. Les bases neurologiques de cette première proto-conscience de soi sont encore en discussion. On pense que les neurones-miroirs (Sur les neurones-miroirs, voir chapitre 3) considérés il y a quelques années comme ouvrant des perspectives très importantes dans la compréhension des mécanismes de la conscience, jouent effectivement un rôle majeur. Mais ils ne sont certainement pas les seuls. Pour ce qui nous concerne, présentant les hypothèses matérialistes relatives à l'apparition de la conscience chez l'homme et de toutes les formes culturelles associées à celle-ci, il est en tous cas intéressant pour nous de mieux connaître le rôle du langage.

Malheureusement peu de traces demeurent aujourd'hui des conditions dans lesquelles se sont développées les cultures langagières humaines, bien plus complexes que ne le sont celles associées aux langages animaux. On ne voit pas nettement, notamment, les modifications anatomiques ou comportementales précises en ayant résulté. C'est pourquoi il convient d'attacher une grande importance aux travaux contemporains concernant l'apparition du langage chez les robots.

Ces travaux, malheureusement encore trop rares et mal connus, soulèvent il est vrai un grand scepticisme. Comment des machines, aussi perfectionnées soient-elles, peuvent-elles acquérir des rudiments de langage et par conséquent sans doute



des éléments de conscience de soi, qui en feraient véritablement nos semblables ? On s'imagine généralement que si elles le font, c'est parce qu'elles sont programmées par leurs concepteurs à cette fin, comme le sont d'innombrables dispositifs technologiques capables d'identifier des sons et de synthétiser des réponses vocales. Mais il ne s'agit absolument pas de cela, comme nous allons le voir. Les robots participant à de telles expériences acquièrent progressivement, en détournant (exaptant) des modes d'entrée-sortie qui n'étaient pas prévus pour cela, la capacité de communiquer et de se doter de formes culturelles exploitant ces possibilités de communication.

Certes, les robots utilisés pour ces échanges sont très bien dotés en programmes internes implantés par leurs concepteurs. Le chien Aibo de Sony, nous l'avons dit, utilisé dans certaines de ces expériences, est, comme ses homologues provenant d'autres laboratoires, une machine très complexe, aux nombreux sous-programmes. Cependant les chercheurs de Sony participant au programme ECAgents ont dépouillé le chien Aibo (qui n'est d'ailleurs plus commercialisé, nous l'avons dit) de tous ses programmes commerciaux. Il constitue cependant encore une plateforme robotique d'une grande richesse, aux nombreuses « boîtes » susceptibles d'interagir. De ce fait, mis dans une situation nouvelle, il est capable de comportements par essais et erreurs lui donnant des possibilités d'adaptation innovantes importantes. Mais encore une fois, répétons que dans de tels cas, le comportement adaptatif nouveau qui est observé – quand il apparaît un comportement nouveau, ce qui n'est pas toujours le cas - n'est pas programmé a priori par les responsables de l'expérimentation. Il est seulement constatée et si possible expliqué a posteriori.

Sur le plan informatique, il faut préciser que l'architecture hardware de ces robots fait un large appel aux réseaux neuronaux, de même que la programmation utilisée recourt à la sélection de type génétique des instructions les plus efficaces. Il s'agit de deux modes plus ou moins inspirés de la nature, qui ouvrent aux robots des possibilités d'évolution non négligeables – bien qu'encore limitées par le manque de ressources informatiques générales. On notera que la machine consciente proposée par Alain Cardon (section précédente) ne fait pas appel aux réseaux neuronaux ni à la programmation génétique, mais aux agents logiciels. Il s'agit d'un système massivement multi-agents, dont les capacités d'apprentissage et d'interaction sont bien plus importantes. Mais c'est une solution « luxueuse » et encore mal maîtrisée, que la plupart des laboratoires robotiques ne souhaitent pas utiliser à ce jour.

## **Des bactéries aux robots : de la coopération aux langages.**

Le laboratoire Sony-CSL de Paris, sous la direction de Luc Steels, a été pionnier dans l'expérimentation des modalités d'acquisition du langage chez les robots. Ce laboratoire, bien que dépendant d'une entreprise industrielle, est dédié à la recherche fondamentale. Et les objectifs ont été dès le début d'une grande portée épistémologique. Il s'agissait de comprendre comment est apparue et s'organise la communication humaine, non pas par des études sur l'origine des langages dans les sociétés animales et pré-humaines, mais en construisant des robots qui interagissent ensemble et avec l'homme. Ces robots servent de modèles pour la compréhension des mécanismes qui ont été et demeurent à l'œuvre dans l'apparition et l'apprentissage du langage dans les sociétés humaines. On n'affirmera pas que ces modèles représentent exactement ce qui a pu se passer aux origines de la communication langagière, mais on obtiendra de fortes présomptions débouchant sur des hypothèses susceptibles d'être reprises et testées par les chercheurs en sciences humaines et les épistémologues. Les robots modernes sont à la fois suffisamment proches et suffisamment différents des hommes pour que ces hypothèses puissent être fortement suggestives et constructives.

Les travaux menés chez Sony-CSL Paris ne sont pas les seuls à explorer le thème de l'émergence du langage dans des populations de robots interagissant en communautés. Ils participent d'un bouquet de recherches associant plusieurs laboratoires européens et un japonais, au sein du projet ECAGents, financé par le programme européen Technologies futures et émergentes (IST-1940) (1). L'objectif en est de développer une nouvelle génération d'agents incorporés (embodied, c'est-à-dire dotés de corps robotiques) qui puissent interagir sans intervention humaine avec le monde physique et communiquer spontanément, soit entre eux, soit avec des humains. Le site du projet indique que les recherches feront appel dans un premier temps aux fonctionnalités permises par les technologies existantes (téléphone mobile, WI-FI, robots) avant de proposer de nouvelles fonctionnalités et de nouvelles solutions technologiques. Sur le plan méthodologique, on visera aussi le développement de nouveaux concepts, outils, modèles, algorithmes, méthodes d'évaluation s'appliquant à des populations évoluant grâce à l'interaction et à la communication. Le projet cherchera à faire apparaître les propriétés de base des différents systèmes d'échange, depuis ceux des animaux jusqu'aux langages humains complexes utilisant les médias technologiques. Il s'agira de mettre en évidence la nature des systèmes actuels de communication et de proposer des hypothèses selon lesquelles concevoir de nouvelles technologies basées sur des systèmes artificiels incorporés, c'est-à-dire des robots.

Il n'est pas besoin de longs discours pour démontrer à la fois l'ambition et l'intérêt considérable de telles recherches. Malheureusement, la communication s'établit encore mal entre les travaux de la robotique moderne et ceux des autres

sciences, qu'il s'agisse des sciences du vivant ou des sciences sociales et humaines. Il en résulte que les avancées de tout ce qui concerne la vie artificielle et l'intelligence artificielle évolutionnaires restent peu connues des autres disciplines. Au sein même de l'intelligence artificielle, du fait de l'excessif cloisonnement des communautés dédiées en France à l'intelligence artificielle «classique», la linguistique évolutionnaire, pour reprendre le terme proposé, n'a pas encore trouvé selon nous la place qu'elle mériterait. Il nous paraît donc important d'y faire référence ici. Nous sommes là en effet au centre d'un problème important: montrer comment la robotique et l'intelligence artificielle peuvent non seulement simuler des processus analogues à ceux ayant permis l'apparition de la vie et des sociétés intelligentes, mais envisager les directions selon lesquelles les systèmes biologiques et sociaux intelligents (dont l'homme) pourraient co-évoluer avec des entités artificielles interagissant avec eux.

### **La coopération chez les bactéries**

Mais si l'on veut aborder cette ambitieuse « vision » avec le recul nécessaire, il nous semble indispensable de remonter au plus haut de l'apparition de la communication dans les systèmes vivants. C'est semble-t-il dans l'archéologie de la communication biologique que l'on pourra identifier des mécanismes génériques susceptibles de s'appliquer, sous des formes différentes, à toutes les échelles du vivant. On fera sans doute alors une constatation intéressante, à savoir que ces mécanismes génériques paraissent sinon les mêmes, du moins très proches de ceux qui permettent l'émergence de la communication langagière au sein des groupes d'automates. Les imaginatifs iront encore plus loin. Ils feront l'hypothèse que ces mêmes mécanismes se retrouvent à la source de toutes les formes d'évolution construisant des représentations symboliques de notre univers.

Il n'est pas question ici d'aborder tous ces thèmes à la fois. Bornons-nous, avant de revenir aux travaux de Sony-CSL et aux recherches menées dans le projet ECAgents, d'évoquer ce que nous apprennent les travaux récents relatifs à l'émergence de l'association chez les organismes monocellulaires.

Pourquoi les organismes vivants ont-ils développés au cours de leur évolution des langages de plus en plus organisés ? Compte tenu des dépenses en temps et en énergie que l'émergence de ces langages leur imposait, il a bien fallu qu'ils y trouvent avantage. La réponse classique est que le langage permet d'améliorer la coopération entre les individus au sein d'un groupe, coopération rendant ce groupe plus efficace dans sa compétition darwinienne avec les autres. L'efficacité du groupe s'étendant à celle de l'espèce, les langages sont devenus des attributs essentiels de la compétition entre des espèces appelées à affronter des milieux changeants et pauvres en ressources. Soit. Mais en ce cas, avant le langage,

le facteur qu'il convient d'étudier en premier lieu est la coopération... Vient alors la question de savoir comment et pourquoi la coopération s'est installée ? Remarquons que le terme de coopération peut désigner nombre de choses... Sans remonter à l'apparition des premières associations de protéines auto-répliquantes, nous pouvons considérer que les formes les plus visibles de coopération sont apparues quand les bactéries unicellulaires se sont associées pour former des organismes multicellulaires.

Un article d'Emily Singers résume les recherches actuelles des biologistes évolutionnaires sur les origines des organismes multicellulaires (NewScientist, 4 décembre 2004, p. 46. Life force). Par rapport aux recherches traditionnelles sur le terrain, elles montrent comment l'étude des processus de l'évolution peut être aujourd'hui considérablement améliorée grâce aux techniques qui permettent d'obliger diverses espèces de bactéries à s'adapter *in vitro* à des changements de milieu ou de type de compétition artificiellement provoqués par les chercheurs. Ceux-ci en déduisent comment les souches de bactéries évoluent face à ces changements. Les bases mêmes de l'algorithme darwinien « mutation-sélection-ampliation » ne sont pas modifiées, mais ces expériences montrent les nombreuses façons selon lesquelles elles se conjuguent en fonction des circonstances, si bien qu'il est impossible de prédire les chances évolutives d'une souche mutante particulière, qu'elle soit actuellement dominante ou, au contraire, apparemment menacée de disparition. Ces recherches sont évidemment particulièrement utiles pour l'étude de la résistance des bactéries aux antibiotiques.

Or la coopération est pour les bactéries une forme omniprésente d'adaptation. Elle prend souvent la forme de l'émission d'agents chimiques. On a pu parler de webs bactériens, sans aller jusqu'à dire que les substances émises par les microbes pour se reconnaître, telles les «quorum sensors», constituent des précurseurs des langages (2). Mais la coopération devient encore plus évidente quand elle aboutit à créer des organismes multicellulaires, associant les ressources de bactéries différentes dans des entités dont le potentiel total est plus important que la somme des potentialités des associés. Or comment, si l'on peut dire, l'idée de la pluri-cellularité peut-elle venir à des bactéries monocellulaires qui pouvaient survivre sans cela ? Cette question, on le voit, est proche de celle que l'on pose concernant les origines du langage. Comment l'idée d'échanger des signaux à signification convenue vient-elle à des organismes qui jusqu'alors ne parlaient pas et s'en trouvaient apparemment très bien ?

L'article d'Emily Singer est très explicite sur ce sujet de la coopération entre bactéries. Elle cite les propos du biologiste Paul Rainey de l'Oxford Centre for Environmental Biotechnology - voir : <http://www.eng.ox.ac.uk/oceb/biog.html> - . «Il est coûteux de coopérer, donc il doit y avoir un bénéfice à se mettre en groupe». Mais quel est ce bénéfice ? Les expériences en laboratoire montrent que les bactéries obéissent à une tendance très forte à la coopération, comme si celle-ci représentait une valeur qui leur était imposée de l'extérieur.

Il n'est pas question de dire qu'ainsi se manifeste le doigt de Dieu en faveur de l'altruisme dans la nature. Les expériences montrent en revanche que dès que des bactéries sont soumises à des conditions nouvelles nécessitant une adaptation, elles font différentes tentatives, faisant appel à des solutions parfois éloignées, pour s'associer afin de trouver la meilleure formule de survie collective. Elles ne créent pas encore à ce stade d'organismes multicellulaires, cependant elles sont en bonne voie pour le faire. Elles procèdent alors par essais et erreurs, sur le mode darwinien de la mutation. Tout se passe comme si elles avaient acquis un gène de l'association qui les prédisposerait à s'associer... gène que l'on retrouverait dans les organismes plus complexes.

Mais parler d'un gène de l'association relèverait d'un darwinisme naïf. Si les bactéries, comme tous autres organismes plus complexes, sont poussées à s'associer, c'est sans doute pour des raisons plus simples. Elles réagiraient par exemple à des signaux physiques découlant de leur proximité géographique. Si deux organismes semblables sont suffisamment proches pour que les substances qu'ils émettent dans le milieu puissent éventuellement s'additionner, il peut s'établir une association de fait. Si l'association permet aux associés d'accéder à des sources de nutriments inaccessibles autrement, elle se perpétue, y compris sur le mode génétique. Dans l'article cité, Paul Rainey mentionne le cas de la bactérie *Pseudomonas fluorescens* dont les individus se lient par un mucus pour former des radeaux. Grâce à cette formule, elles peuvent flotter à la surface d'un liquide nutritif et profiter des ressources en oxygène offertes par l'air, ressources hors de portée des bactéries restant immergées dans le liquide. Le radeau pourrait être considéré comme le prototype d'un organisme multicellulaire dont les composantes pourraient ensuite se spécialiser dans des fonctions différentes.

Les premières *Pseudomonas* ayant «inventé» ce dispositif ne l'ont pas fait intentionnellement. Elles disposaient déjà de la faculté de sécréter du mucus et cette faculté a été exploitée par la présence d'un nombre suffisant de bactéries dans un même lieu. La fabrication du radeau peut alors être considérée comme une fonction exaptée (pour reprendre le terme de Stephen Jay Gould) à partir d'une fonction antérieure toute différente, sécréter du mucus. Certaines expériences, citées également par l'article, montrent que si l'expérimentateur empêchait une souche de bactéries de faire appel à telle propriété lui ayant permis de créer une association, d'autres propriétés jusque là inutilisées pour l'association pourraient donner naissance à une coopération reposant sur d'autres facteurs.

On observera par ailleurs qu'à l'intérieur des contraintes de sélection imposées par l'environnement, ce sont les modes d'association les moins exigeants en ressources qui sont sélectionnés les premiers. On retrouve là une des grandes règles de la morphogenèse : la priorité donnée aux processus les moins gourmands en énergie (Cf. Chapitre 2).

En résumé, que conclure de ce qui précède ? L'association entre organismes jusque là indépendants obéit à la règle générale du hasard et de la nécessité. Elle ne se produit que si d'une part certaines propriétés de ces organismes, jusque là liées à des fonctions vitales n'ayant rien à voir avec l'association, leur permettent dans telles circonstances particulières de se lier entre eux et si, d'autre part, cette liaison ou coopération se révèle favorable à la survie des organismes associés. On obtient alors des populations d'agents agissant en groupe (le radeau de bactéries flottant à la surface d'un liquide, par exemple) ou de véritables symbioses aboutissant à créer de nouveaux organismes (monocellulaires donnant naissance à des pluricellulaires). Dans l'un comme dans l'autre cas, les propriétés exaptées ayant permis de fonder une association durable se trouvent renforcées et structurées par l'évolution. On pourra les qualifier de moyens de communication spécifiques (par exemple des émissions-réceptions de médiateurs chimiques) ou de langages plus polyvalents, selon leur complexité.

Enfin, pour que l'association se produise et soit durable, il faut que les organismes soient suffisamment proches les uns des autres pour pouvoir s'influencer, il faut que les moyens exaptés soient les plus économiques possibles en ressources et enfin, il faut que l'association apporte des bénéfices. Ces règles semblent valables quelle que soit la taille des organismes considérés : monocellulaires ou pluricellulaires plus ou moins complexes, humains, robots.... On retiendra aussi que l'association obéissant au principe darwinien du hasard et de la nécessité, ou si l'on préfère de la mutation-sélection, les associations réussies sont en très petit nombre par rapport à toutes les associations qui pourrait théoriquement se produire si les organismes actuellement séparés se trouvaient rapprochés et bénéficieraient de conditions favorables. Bien qu'en petit nombre par rapport à la gamme des possibles, elles ont quand même donné lieu à l'infinie variété des formes vivantes et des cultures.

### **La coopération chez les robots**

Si l'on change brutalement de niveau dans l'échelle des complexités et si l'on s'intéresse au langage, on peut considérer que le même besoin générique de s'associer et de faire pour cela appel en premier lieu aux processus les moins gourmands en énergie, s'est traduit par la sélection de certaines propriétés jusque là non utilisées pour communiquer de façon symbolique. Ces propriétés ont été exaptées au cours de processus d'essais et erreurs plus ou moins longs (évidemment involontaires, c'est-à-dire sur le mode du hasard et de la nécessité). Apportant des bénéfices compétitifs, ces exaptations ont donné naissance à des échanges de messages symboliques, proto-langages puis langages, au sein des sociétés animales et humaines. Ceci s'est produit tout autant au plan des mécanismes physiques (type de gestes ou de sons) qu'en ce qui concerne la sélection des contenus sémantiques ou de signification des échanges. Mais pour tester ces hypothèses, il n'est plus

possible de générer une évolution artificielle au sein de tubes à essais ou de boîtes de Pétri. Les processus évolutifs ayant donné naissance aux langages symboliques se sont en effet déroulés sur des millions d'années. Sauf dans de très rares cas, tel celui cité par Frédéric Kaplan dans un livre remarqué (Kaplan, F. La naissance d'une langue chez les robots, Hermès Science 2001, op.cit) (un groupe de sourds-muets ayant redécouvert un proto langage naturel) on ne peut pas encore expérimenter au sein d'espèces vivantes complexes les hypothèses relatives à cette science nouvelle que l'on appelle la linguistique évolutionnaire.

Heureusement, les populations de robots offrent désormais la possibilité d'étudier un grand nombre d'hypothèses relatives aux conditions d'émergence de la coopération entre individus, y compris dans ce domaine très sophistiqué qu'est la coopération langagière. Des agents logiciels et mieux encore des robots incorporés (dotés de corps robotisés et de systèmes d'intelligence artificielle) peuvent être mis en présence les uns les autres dans des conditions très variées. Si on les soumet à des contraintes les obligeant à évoluer pour survivre, on s'aperçoit qu'ils découvrent l'intérêt de la coopération et donc de la communication. Certaines des propriétés de leurs organes d'entrée-sortie robotiques, jusqu'ici non dédiées à la communication, peuvent trouver de nouveaux emplois en générant des échanges. Si ces échanges sont profitables, ces propriétés se transforment progressivement en moyens de communication de plus en plus spécifiques et efficaces. Comme on pouvait s'y attendre, ce sont les moyens les plus économes en énergie qui sont progressivement sélectionnés. Ceci explique pourquoi les langages entre robots retrouvent certaines des caractéristiques des langages dans la nature, aussi bien en ce qui concerne les appareils utilisés que les contenus échangés. Le langage est coûteux. On ne parle pas pour ne rien dire ou, plus exactement, si dans certains cas on parle pour ne rien dire, c'est parce qu'à terme ceci se révèle plus profitable que, précisément, ne rien dire.

Donnons un exemple simple du processus d'exaptation évoqué ci-dessus. Supposons un robot doté d'un capteur à ultrasons lui permettant d'éviter les obstacles. Si plusieurs robots dotés de tels capteurs se trouvent réunis et doivent pour des raisons vitales se regrouper (par exemple lutter contre un prédateur), il pourra arriver (pas nécessairement) que ces capteurs leur servent à s'identifier les uns les autres et se regrouper. Le système à ultrasons sera devenu un langage primitif.

On dira que les robots, aussi perfectionnés soient-ils, n'ont pas encore atteint la complexité d'un animal ni même d'une bactérie, et que les expériences susceptibles d'être conduites avec eux incorporent beaucoup de connaissances humaines préalables. On ne se trouve donc pas à l'origine des processus d'émergence du langage dans le monde biologique. Mais l'objection n'est que partiellement valable si l'on considère que, pour reprendre le terme employé plus haut, la coopération, qu'elle prenne ou non le biais du langage, obéit à des

contraintes génériques extrêmement simples, se retrouvant sans grands changements à tous les niveaux de complexité des organismes physiques, biologiques ou artificiels constituant notre univers.

La linguistique évolutionnaire nous intéresse particulièrement ici par un aspect qui reste encore peu étudié pour le moment par les laboratoires de robotique. Il s'agit de la façon dont elle contribue à l'émergence chez les robots d'une proto-conscience de soi, amorce très probable de formes de conscience de soi mieux affirmées. Nous avons en effet indiqué plus haut que c'est sans doute dans l'échange langagier avec leurs homologues que les premiers hominiens, comme aujourd'hui les bébés, ont pris conscience de leur autonomie en tant qu'individu et en ont tiré toutes conséquences utiles à leur propre survie. Il est très probable que les robots, soumis à des contraintes de même nature, nous donneront l'exemple de telles « prises de conscience »

### **Les recherches de Sony-CSL Paris**

Depuis 1998/99 jusqu'à aujourd'hui, le laboratoire a conduit une série d'expérimentations qui, conjuguées, permettent de construire une théorie générale de l'apparition du langage aux résultats impressionnants. Le domaine de recherche fait partie, comme nous l'avons dit, de ce que l'on commence à nommer la linguistique évolutionnaire. On y étudie la façon selon laquelle des agents artificiels en interaction peuvent se doter de langages possédant certaines des propriétés des langues naturelles et comment les significations ainsi échangées peuvent évoluer afin de s'adapter aux besoins et aux possibilités des agents. Citons les principaux domaines étudiés :

- Le Naming Game ou jeu de l'attribution d'un nom. L'expérience a utilisé des agents logiciels représentant des communautés d'utilisateurs dotés de langage afin d'explorer la façon dont des lexiques partagés apparaissent dans une population. Les agents interagissent d'une façon organisée. Ces interactions conduisent à l'apparition d'un répertoire commun de mots permettant de désigner des objets.

- Les Têtes-Parlantes 1999. Cette expérience a donné lieu au livre de Frédéric Kaplan précité. Il s'est agi de la première démonstration faisant appel à de véritables robots. Elle a montré, comme la précédente, comment un lexique et un système de concepts partagés pouvaient s'organiser dans une société de robots au cours de leurs « échanges culturels ». Ainsi le langage peut être considéré comme un système adaptatif complexe comprenant des structures globales résultant d'interactions sociales locales. Les robots ont développé leur vocabulaire en observant une scène à travers des caméras numériques et en communiquant sur ce qu'ils ont vu. L'expérience a été complétée par l'intervention d'observateurs humains qui



pouvaient créer des agents logiciels et les faire interagir avec les robots. Elle visait à proposer des réponses à trois questions d'une grande portée philosophique : comment les concepts acquièrent-ils leurs significations ? Une intelligence artificielle est-elle possible ? Comment les machines interagissent-elles avec les humains ?

La recherche a été reprise dans le projet Talking Aibo, définissant un cadre dans lequel un humain peut enseigner à un robot, en l'espèce le chien Aibo, comment nommer les objets de son environnement.

- Maïdo et Gurby 2001. Cette nouvelle expérience cherchait à montrer, non pas comment émergeaient les contenus conceptuels des langages, mais comment une population pouvait en interagissant se mettre d'accord sur un système de sons partagés, c'est-à-dire sur des véhicules communs permettant l'échange langagier. Dans cette expérience ont aussi été développées des technologies de synthèse vocale émotionnelle et de reconnaissance des émotions dans la voix humaine.

- Aujourd'hui, les recherches de Sony CSL se développent dans plusieurs directions. Un des enjeux est de montrer que les dynamiques collectives peuvent être étendues pour rendre compte de l'émergence de formes grammaticales. Pour cela un nouveau cadre formel a été développé, appelé les « grammaires constructionnelles fluides ». L'autre direction consiste à étudier les pré-requis nécessaires à l'émergence de la communication. Il s'agit de montrer comment des mécanismes programmés dans les expériences précédentes (Têtes Parlantes, Maïdo et Gurby) peuvent être développées de façon autonome par les robots. Ainsi de l'"attention partagée". Un autre exemple porte sur les « origines auto-organisées des systèmes de vocalisations ». Le mécanisme essentiel de la « curiosité » est étudié dans une expérience dite du « tapis d'éveil », Playground experiment(2) dans lequel un robot muni d'un système de curiosité développe des activités de plus en plus complexes sur un tapis d'éveil de bébé.

## **Le projet ECAgents**

(Ces informations sont résumées de celles fournies par le site ECAgents : <http://ecagents.istc.cnr.it/index.php?tmva=7> Les lecteurs souhaitant approfondir ces questions ont tout intérêt à se reporter aux originaux)

Le projet se développe sur plusieurs années à partir de 2003/2004. Il rassemble des partenaires ayant l'expérience des systèmes énumérés ci-dessus. La France y est représentée par le laboratoire de Sony, mais malheureusement pas par un laboratoire universitaire.

Les fondations scientifiques du projet s'inspirent des méthodes et des techniques découlant des recherches sur les systèmes complexes. Un système évolutionnaire de communication sera considéré comme un système complexe adaptatif, dont l'étude s'appuiera sur les théories de l'évolution, de l'information, des jeux, des réseaux, des systèmes dynamiques. Il existe encore aujourd'hui de grandes lacunes entre les recherches sur les systèmes complexes et les technologies de l'information, mais le projet vise à combler ces lacunes, ce qui représentera un objectif d'une très grande importance, susceptibles de nombreuses applications, en robotique auto-évolutionnaire, pour le web sémantique ou pour les technologies de communication sans fil.

### **Plus spécifiquement, le projet vise à :**

- Développer de nouvelles générations de robots capables d'évoluer de façon autonome, de s'organiser et d'opérer de façon efficace dans un environnement dynamique,

- Définir les conditions permettant à une population de robots de développer un langage commun de communication et de partager des connaissances,

- Identifier de nouvelles méthodes et nouveaux algorithmes permettant d'obtenir ces propriétés émergentes.

Les choix méthodologiques du projet sont les suivantes:

- Constituer des populations d'agents: il s'agit d'agents qui en interagissant acquièrent de nouvelles capacités qu'aucun agent ne pourrait manifester seul.

- Les agents sont dotés de corps et sont physiquement situés. Ce sont donc des agents physiques interagissant entre eux et avec un environnement physique, d'une façon non symbolique, mais directe.

- Le système de communication des agents n'est pas défini à l'avance. Il émerge spontanément des interactions des agents entre eux et avec leur environnement physique. Il s'y adapte en permanence.

- L'ensemble s'auto-organise et évolue en fonction des changements des populations d'agents, des médias de communication qu'ils utilisent, de leur environnement et de leurs sujets d'intérêts.

- La stratégie de recherche ne consiste pas à étudier, comme on le fait traditionnellement, les modes de communication d'agents existant déjà dans la nature, mais à construire des populations d'agents artificiels à partir desquelles on conduira des expériences et on formulera des hypothèses.

- Au-delà, on étudiera les propriétés plus générales et plus abstraites de la communication au sein de grandes collections d'agents interagissant, par exemple le rôle de la topologie des réseaux d'interaction et de communication, les propriétés abstraites des systèmes de communication (contenu en information des termes utilisés, systèmes combinatoires contre systèmes non combinatoires, systèmes grammaticaux contre systèmes non grammaticaux, etc.), le rôle de l'interaction dynamique (théorie des jeux).

- A partir de ces recherches, le projet suggérera de nouveaux systèmes technologiques capables d'interagir entre eux et avec l'environnement en utilisant les méthodes étudiées (robots, appareils portables, calcul réparti, etc.).

## **Premiers résultats**

On trouve sur le site du projet plusieurs documents du plus haut intérêt précisant les méthodes retenues par les diverses équipes de chercheurs associées dans le projet, ainsi que les objectifs restant à atteindre.

On consultera notamment (à la date de juillet 2006), les rapports suivants :

Livre Blanc version 1 : Problèmes à résoudre et grands défis (anglais) 224 pages <http://ecagents.istc.cnr.it/imgs/whitepaper.pdf>

Livre Bleu : une feuille de route pour la recherche (anglais) 43 pages [http://ecagents.istc.cnr.it/imgs/blu\\_paper.pdf](http://ecagents.istc.cnr.it/imgs/blu_paper.pdf)

Nous nous bornerons ici à résumer le contenu du Livre Blanc, à regret car le Livre Bleu est tout aussi, voire davantage, intéressant concernant l'avenir de la robotique autonome.

Ce Livre Blanc, comme le Livre Bleu, est d'une grande richesse. Malheureusement, il exige du lecteur un certain nombre de connaissances en linguistique et en théorie de l'information qui ne le rendent pas d'une approche aisée. On verra pourtant, même en se limitant à une lecture rapide, que les objectifs des chercheurs visent bien à permettre aux robots avec lesquels ils travaillent d'acquérir des capacités linguistiques et d'auto-représentation bien supérieures à celles dont l'évolution a doté les animaux. Il s'agit en fait de viser à un niveau de performance mettant le robot au niveau de l'homme. Rappelons que, ce faisant, les

chercheurs ne visent pas à programmer eux-mêmes les capacités qu'ils attendent des robots, mais à mettre ceux-ci dans des conditions de contrainte de développement leur permettant d'acquérir eux-mêmes ces capacités par apprentissage. Evidemment, les robots laissés à eux-mêmes sans interventions humaines n'auraient la possibilité d'acquérir ces compétences. Ils doivent être mis en condition favorable au départ. Mais il en est un peu de même dans la nature. On sait qu'un enfant se développant sans relations avec ses semblables et sans interactions avec un milieu stimulant risque de rester inapte à toute activité intellectuelle.

### **Les équipes signataires du Livre Blanc procèdent en plusieurs étapes :**

1. définir les conditions de la communications entre animaux et robots animaloïdes dits animats. On sait qu'une voie d'expérimentation très fructueuse consiste à immerger des robots dans des sociétés animales, y compris dans des sociétés d'insectes, en dotant ces robots de moyens d'interface avec ces animaux qui soient reconnus par eux (par exemple des phéromones dans une expérience conduite par Jean-Louis Deneubourg dans le programme européen Leurre (3)). On peut ainsi infléchir le comportement d'un groupe, mais aussi, plus surprenant, conduire le mini-robot à se comporter lui-même comme un des membres de ce groupe. Ainsi devient-il possible d'élucider certains des moyens et certaines des fonctions de la vie sociale de l'espèce étudiée.

2. Définir les conditions de la communication de type humain (human-like communication) afin d'induire les robots à les implémenter dans leurs propres échanges. La difficulté est infiniment plus grande, du fait de la complexité des langages humains et de leurs modes d'utilisation. Pour cela, on définira les qualités propres au langage humain par rapport à celles des langages animaux, quelles que soient les espèces considérées. On observera ensuite si des robots mis en situation d'échange compétitif développent ou non ces qualités, et à quelles conditions.

Le principal responsable de ces études, Domenico Parisi, (Conseil National de la Recherche, Rome) propose ainsi d'identifier 8 caractères principaux caractérisant le langage humain. Le langage humain dispose de syntaxes permettant d'articuler entre eux des signes plus élémentaires, il est culturellement transmis et augmenté, il sert à communiquer avec soi-même (langage intérieur) et pas seulement avec les autres, il est capable de transmettre des informations sophistiquées sur le milieu extérieur, il peut utiliser des signaux non physiquement liés à l'émetteur (displaced signals) , il transmet des intentions, il est le produit d'un système nerveux complexe et finalement, il influence massivement la cognition.

Si l'on veut que des robots se dotent de tels langages, il faut les mettre dans des situations constituant des pré-requis qu'ils doivent apprendre à maîtriser. On peut en identifier deux : l'« attention partagée » qui s'obtient chez les humains par différents mécanismes tels que le geste de désignation d'un objet ; et le partage d'un moyen physique de communication suffisamment complexe pour s'adapter à des besoins d'échanges très différents. Sur ce point, les technologies modernes permettent aux robots de communiquer par des moyens bien supérieurs à ceux dont disposaient les premiers hominiens.

3. Organiser une suite d'étapes permettant aux robots d'accroître progressivement les capacités de leurs systèmes linguistiques d'échange, en franchissant des seuils significatifs : par exemple création d'une syntaxe, d'une méta-grammaire... Luc Steels, directeur de recherche chez Sony-CSL, qui a coordonné les contributions sur ce point, propose six étapes au travers desquelles le robot devrait se retrouver doté d'une véritable langue de type humain ou, en tous cas, capable d'échanges avec les langues humaines.

Au-delà de ces premiers objectifs, le projet vise à répondre à deux autres enjeux majeurs. Le premier est d'ordre neurobiologique. Il s'agira de transformer les modèles computationnels du langage chez les robots en modèles neuraux susceptibles d'être utilisés pour comprendre l'acquisition et l'emploi du langage chez l'homme – ce qui est loin d'être encore clairement perçu aujourd'hui. Le second enjeu est d'ordre théorique. Il intéresse l'étude des systèmes complexes, en utilisant les acquis des expérimentations précédemment énumérées : comment par exemple un système macroscopique ressent les modes d'évolution des unités microscopiques dont il est constitué, comment les populations d'agents évoluent en fonction de tel ou tel changement global.

Le projet propose enfin diverses applications pratiques, d'intérêt industriel, sociétal ou commercial, intéressant les différents systèmes et matériels susceptibles d'incorporer les agents communicants étudiés précédemment sur un plan théorique.

1 : Nous ne pouvons que renvoyer le lecteur au site web du projet : <http://www.ecagents.org/index.php?tmvb=0|&tmva=21> On y trouvera notamment la liste des laboratoires associés et de leurs principaux travaux, qui font autorité dans le domaine de la robotique autonome.

2 : Par le terme de "quorum sensing", on désigne la propriété qu'ont certaines bactéries pathogènes de rester dormantes dans l'organisme envahi, tant que leur nombre n'est pas suffisant pour leur permettre de passer victorieusement à l'attaque. Elles émettent une substance dont la concentration, fonction du nombre des bactéries présentes dans l'organisme, leur indique le moment favorable pour déclencher l'infection.

3 : Jean-Louis Deneubourg est le très réputé directeur du service d'écologie sociale de l'Université Libre de Bruxelles, et spécialiste de la communication chez les insectes sociaux.

## Chapitre 4, Section 3 :

### Réalité virtuelle, vie artificielle, automates cellulaires

Dans ce bref panorama des systèmes informatiques et robotiques, nous n'avons pas mentionné la réalité virtuelle et la vie artificielle, deux mondes qui sont les produits de l'augmentation de puissance des ordinateurs et des réseaux. Il s'agit de domaines de recherche qui, bien que distincts, posent tous deux des questions épistémologiques qui nous mettent au cœur de ce que David Deutsch a nommé " l'Etoffe de la réalité " . Nous dirons également quelques mots d'un mode de calcul et de modélisation informatique peu connu du grand public, qui intrigue pourtant par sa puissance et ses mystères de nombreux scientifiques, les Automates cellulaires (AC)

#### La réalité virtuelle

La réalité virtuelle est assez populaire aujourd'hui, vu l'usage intensif qui en est fait dans les films et dans les jeux électroniques. Il s'agit de créer, en utilisant les ressources de plus en plus riches des générateurs informatiques de sons, d'images et autres perceptions sensorielles, des mondes proches ou différents du nôtre, dans lesquels les spectateurs-acteurs que nous sommes pourront être plongés, grâce à des interfaces spécifiques (casques, salles équipées, etc.). La réalité virtuelle n'est pas seulement un jeu ou un instrument de formation (par exemples les simulateurs de pilotage). L'objectif, selon les termes de Denis Berthier, dans son livre Méditations sur le réel et le virtuel est « de construire un monde virtuel et y interfacer un être humain en lui donnant l'impression qu'il y perçoit et agit de manière naturelle : perception en trois dimensions, immersion sensori-motrice, interaction en temps réel, etc. ».

Denis Berthier Méditations sur le réel et le virtuel, Collection Impact des nouvelles technologies L'Harmattan, 2004 . Voir l'article

<http://www.automatesintelligents.com/biblionet/2004/juin/berthier.html>

-

Ce qui en découle est très important, non seulement au plan pratique mais au plan philosophique : il sera de plus en plus difficile de distinguer le réel et le virtuel, dans la mesure où nos perceptions s'organisent dans le cerveau de la même façon quelle que soit la source dont elles proviennent. Il s'en suit qu'en principe, avec des ordinateurs d'une puissance considérablement augmentée et la possibilité d'y connecter nos cerveaux dans les deux sens de l'entrée et de la sortie (perspectives

qui ne sont qu'une question de temps), rien n'empêcherait de faire vivre des humains, toute leur existence durant, dans des mondes artificiels ressemblant au nôtre ou au contraire radicalement différents. Ce que nous appelons la réalité, de ce point de vue et comme le suggère David Deutsch, pourrait résulter d'une création permanente engendrée par un univers profond fonctionnant sur le mode computationnel. Nous avons déjà évoqué cette possibilité dans le chapitre 1, à propos des enseignements de la physique quantique.

A titre de méditation scientifique, on peut évoquer l'hypothèse de Nick Bostrom de l'Université d'Oxford (Voir l'article de Jean-Paul Delahaye dans *Pour la Science*, août 2004, p. 94). Nick Bostrom est par ailleurs un de spères du mouvement transhumaniste (Chapitre 6) ) selon laquelle nous pourrions tous être des créatures virtuelles créées par des civilisations infiniment plus avancées que nous technologiquement. Cette hypothèse repose sur plusieurs postulats :

1 - il existe des civilisations assez avancées (post-humaines) pour a) simuler complètement le cerveau humain et b) décider de le faire ;

2 - compte tenu du principe dit d'indépendance du substrat, il est impossible de distinguer le contenu d'un cerveau humain de celui d'un cerveau artificiel ;

3 - compte tenu du très grand nombre de créatures simulées que pourrait produire et disperser dans l'univers une civilisation capable de simuler le cerveau humain, nous avons une très grande probabilité d'être une telle créature plutôt qu'un être " naturel ". Certains objectent que nulle simulation, aussi parfaite soit-elle, ne pourrait éviter les bugs divers, bugs que, si nous étions simulés, nous remarquerions sous forme de défauts de fonctionnement. Mais à cela on répond que notre société et nous-mêmes sommes suffisamment porteurs d'incohérences internes (sans parler du fait que nous sommes encore incapables de résoudre ce que nous nommons des mystères scientifiques) pour que ceci traduise précisément l'existence de bugs dans les logiciels produits par ceux qui simulent notre fonctionnement.

On s'étonnera peut-être d'apprendre que l'argument résumé ci-dessus, dit de la simulation, fait l'objet de discussions intenses parmi les scientifiques et les philosophes s'intéressant à l'univers calculable. Nous ne sommes pas obligés d'adopter des visions aussi extrêmes concernant la puissance de la simulation. Les citoyens que nous sommes auront déjà beaucoup de mal, s'ils veulent comprendre les influences qui s'exercent sur eux, à distinguer les situations simulées des situations simplement observées. Les médias, pour diverses raisons, visant notamment à augmenter leur audience par la recherche de sensationnel, auront de plus en plus tendance à mêler les différents messages.



En ce qui concerne la recherche scientifique, par contre, la simulation, conjuguée avec la vie artificielle (voir ci-dessous) constitue un extraordinaire outil permettant de créer des univers représentant fidèlement des situations non observables directement. Bien sûr, la fiabilité de telles constructions dépend de la qualité scientifique des informations utilisées au départ, mais la possibilité offerte à l'observateur de visualiser des phénomènes qui sans simulation demeureraient abstraits augmente considérablement le champ de l'imagination créatrice. C'est aussi le cas pour des professions chez qui le regard joue un rôle essentiel, comme en ce qui concerne les chirurgiens. Permettre à un chirurgien de se plonger au cœur d'un organe reconstitué par simulation l'aidera le moment venu à utiliser au mieux les images fournies par les micro-robots d'intervention.

### **La vie artificielle**

La vie artificielle, selon les termes de Jean-Philippe Rennard, auteur d'un excellent livre du même nom sur la question (1) a pour objet de mieux comprendre ce qu'est la vie en recherchant et en tentant de reproduire les processus généraux qui la gouvernent. Mais elle cherche aussi à transposer les mécanismes du vivant au sein d'algorithmes et d'artefacts spécifiques (biomimétiques) à l'efficacité souvent surprenante ”.

Les plus utilisés des outils employés par les chercheurs en vie artificielle sont les automates cellulaires. Un automate cellulaire (AC) se représente sur un écran d'ordinateur, mais le principe peut en être décrit sans faire appel à l'informatique. Les premiers AC furent d'ailleurs réalisés sur papier. Il suffit d'imaginer une rangée de cellules dont les unes sont noires et les autres blanches. On applique à cette rangée une règle aussi simple que possible : par exemple, à partir de la rangée initiale, qui ne comporte qu'une cellule noire, construire une seconde rangée qui se superposera à la première et qui respectera la règle suivante (n'importe quelle règle étant évidemment possible) : toute cellule blanche voisine d'une noire deviendra noire. On construira successivement, par application de cette même règle, d'autres rangées qui elles aussi se superposeront aux deux précédentes. Au bout d'un certain nombre de passes, on verra se dessiner un motif en forme de triangle noir constitué de l'ensemble des rangées superposées, la dernière rangée formant la base du triangle étant entièrement noire. Il s'agit d'un motif que rien ne laissait prévoir au vu de la règle et de la disposition initiale des cellules. Avec des règles un peu différentes, on fera apparaître des motifs tous différents et beaucoup plus complexes. En utilisant un ordinateur, on peut visualiser le processus, ce qui fait apparaître des figures mouvantes analogues à des colonies d'insectes, à des croissances végétales et autres formes vivantes.

Les techniques de la vie artificielle offrent aujourd'hui de très nombreux programmes pour simuler des processus physiques dynamiques ou des processus vivants. Elles servent aussi à la création artistique, dans un domaine généralement qualifié de Computer Art, dont l'artiste français Bernard Caillaud, avait fait une de ses spécialités (Bernard Caillaud, *La création numérique visuelle*. Voir Chapitre 7.) .

Les applications ludiques ou artistiques qui en sont données ne doivent pas faire perdre de vue que les techniques de la vie artificielle sont d'abord de puissants outils scientifiques. Citons le Jeu de la vie proposé par John Conway, les automates cellulaires, la programmation génétique et les algorithmes génétiques, les animats ou animaux artificiels.

L'objectif de la vie artificielle est l'étude de propriétés émergentes à partir d'entités et de propriétés définies au niveau local. Ainsi, les observations portent sur des populations plutôt que sur des individus seuls. Il peut donc émerger de ces groupes des comportements globaux inédits n'étant pas commandés par des propriétés individuelles. Nous avons vu précédemment que cette façon d'observer est de plus en plus utilisée par les sciences de la vie et les sciences humaines pour comprendre les comportements de groupe.

L'évolution des populations artificielles fait appel aux modèles darwiniens reposant sur l'«algorithme» reproduction, mutation, sélection, ampliation. On y retrouve en les transposant les concepts de génotype, phénotype, mutation et espèces. On y retrouve aussi les notions devenues courantes en biologie : l'apparition de niches écologiques, de comportements coopératifs et plus généralement de propriétés globales des groupes.

La question est souvent posée de savoir qui s'inspire de qui : les biologistes ou les programmeurs de la vie artificielle ? Comme toujours dans les sciences nouvelles qui reposent sur une forte coopération entre disciplines de la vie et disciplines de l'intelligence artificielle, l'interfécondation est désormais la règle. Chaque fois qu'elle est possible (Chaque fois notamment que les frontières entre formations et carrières le permettent...), la constitution d'équipes composées de chercheurs appartenant à ces deux séries de disciplines est un gage de succès.

On distinguera la vie artificielle, qui met en œuvre des agents informatiques, et la biologie artificielle ou de synthèse, qui fabrique des objets vivants (par exemple des virus très simplifiés), en assemblant des éléments à base d'acides aminés et autres molécules biologiques. Nous avons examiné cette direction de recherche, en plein développement, dans le chapitre 2 consacré à l'évolution.

## Les automates cellulaires de Stephen Wolfram

L'importance des automates cellulaires est démontrée par un ouvrage très ambitieux, *A New Kind of Science* (NKS), dû au mathématicien et informaticien Stephen Wolfram (2). Ses propositions véritablement stupéfiantes sont encore loin d'être exploitées, bien qu'il ne soit plus possible aujourd'hui de les ignorer.

Si les automates cellulaires (AC) ne sont pas une nouveauté, comme nous l'avons vu plus haut, jamais personne ne les avait encore présentés dans une perspective si riche, traçant un chemin des plus originaux pouvant mener, selon l'auteur (qui n'a en principe rien d'un illuminé) à une nouvelle connaissance du monde et de nous-mêmes.

Dans cet ouvrage, Wolfram démolit impitoyablement la place éminente jusqu'ici donnée en sciences aux mathématiques. Il veut montrer que, dans toutes les disciplines scientifiques, l'emploi judicieux de la modélisation et simulation par AC suffit à la recherche. Les AC s'inscrivent dans le développement du calculateur universel, dont les principes ont été posés par la machine de Turing. Il s'agit de travailler avec des entités discrètes, pas à pas, en principe en langage binaire. Ceci postule que le continu qui relèverait de calculateurs analogiques peut être réduit au discret. Le traitement pas à pas est d'autant plus efficace que les algorithmes utilisés sont simples. La règle ou le programme simples sont vraiment la fondation de l'approche de Wolfram, qu'il retrouve partout à l'œuvre dans la nature.

Ceci conduit directement aux AC, qui sont l'outil utilisé par l'auteur pour formaliser et illustrer toutes ces hypothèses. Les premiers chercheurs en intelligence artificielle en avaient fait un instrument essentiel pour la construction de modèles simulant l'évolution des systèmes complexes, notamment en biologie. Certains les emploient toujours, tels Thomas Schelling. Mais ils furent abandonnés (sauf exception) face au développement des modèles mathématiques complexes eux-mêmes supportés par les programmes informatiques lourds permis par les ordinateurs modernes.

La redécouverte des AC (ou plutôt l'approfondissement du domaine) par Stephen Wolfram s'appuie et s'appuiera d'ailleurs de plus en plus sur les énormes ressources permises par les stations de travail moderne, le travail en réseau de type Grid et les progrès prévisibles des calculateurs numériques. On peut penser que les AC proposeront également des algorithmes bien adaptés aux futurs calculateurs quantiques. On se retrouvera en effet là dans le discret microscopique, c'est-à-dire ce qui se passe au niveau de la particule physique ou de la molécule biologique. Avec les AC on pourra faire le lien avec les possibilités, théoriquement infinies, de calcul multi-directionnel (multiways) permis à la particule quantique. Autrement dit,

les AC pourraient explorer des hypothèses relevant du multivers, c'est-à-dire des univers parallèles.

Avec un minimum de culture mathématique (3), on peut à la lecture de l'ouvrage de Wolfram se rendre compte que la modélisation des problèmes les plus complexes, ceux de la physique quantique, ceux de la relativité générale et même ceux de la théorie des cordes ou autres théories intéressant la gravitation quantique encore en devenir, peuvent être envisagées à partir d'AC judicieusement choisis – ce que les mathématiques actuelles sont bien en peine de faire. Ainsi Wolfram montre comment on peut représenter un univers où le temps est quantifié sous forme de particules, ou bien un autre univers n'ayant plus de références temporelles et spatiales tout en conservant une structure informationnelle – ce qui correspond peut-être à ce qui se passe au cœur des trous noirs. Bien sûr, un AC ne peut prétendre décrire vraiment le cœur des phénomènes cosmologiques, mais il peut nous en donner une image, nous convaincre que certaines hypothèses théoriques seraient physiquement possibles – en général à partir de l'utilisation de règles initiales très simples. De telles démonstrations, si les physiciens les admettaient comme valides, pourraient les conduire à bâtir de nouvelles hypothèses pour de futures vérifications expérimentales. On se trouve ainsi, vis-à-vis du cosmos, dans la situation où étaient les premiers voyageurs utilisant des cartes. Même si celles-ci n'étaient en rien capables de reproduire pour eux les détails utiles des continents, elles pouvaient leur suggérer des idées pour aller sur le terrain découvrir ce qu'il en était, plutôt que se fier à une symbolique poétique, telle celle d'Homère dans l'Odyssée.

Le point essentiel est, nous l'avons rappelé, que l'évolution des AC dirigée par des instructions ou règles ne peut être définie ni prévue à l'avance. La grande découverte de Wolfram est de montrer que les AC, à partir de règles simples judicieusement choisies, peuvent dans certains cas produire ce qui pour nous sera du complexe et de l'indécidable sans limites prévisibles, pour peu que les ordinateurs en charge disposent du temps et des ressources informatiques en quantité suffisante. L'intelligence artificielle utilisant les AC pourra produire des artefacts de systèmes naturels en bien plus grand nombre que ce que l'ensemble des humains pourraient tester et implémenter dans des systèmes physiques et biologiques du monde réel. Il s'agira de ce que l'auteur a nommé des “ complexités intrinsèques ” dont nous pourrions constater l'émergence mais dont notre entendement sera incapable de comprendre les processus de génération – de même sans doute que nous sommes, pour le moment encore, et sans doute définitivement, incapables de comprendre les processus intrinsèques ayant permis la naissance de notre univers. (4)

Stephen Wolfram prend soin de nous mettre en garde. Les AC ne sont que des modèles informatiques des processus s'exerçant dans la nature, que ce soit dans le monde physique, ou biologique. Ils ne peuvent nous renseigner sur ce qui se

passer au niveau des arrangements atomiques et moléculaires des entités naturelles. La croissance d'un flocon de neige simulé par un AC ne nous renseigne pas sur la façon dont les tensions entre molécules d'eau déterminent celle d'un vrai flocon. Mais les AC ont l'avantage irremplaçable de montrer que ce seraient chaque fois – sauf exception, évidemment – des règles simples identiques ou comparables qui s'appliqueraient. Ils nous donnent ainsi un outil indispensable pour démêler derrière la complexité apparente les règles simples qui sont à l'œuvre. Ceci permet alors de les étudier au cas par cas.

Ceci ne veut pas dire que les AC ne nous aideraient pas à traiter les problèmes posés par la recherche fondamentale : physique quantique, astrophysique, biologie, etc. Selon Wolfram, convenablement choisis et mis en oeuvre, utilisant les ressources des nanotechnologies et - si on le réalise - de l'ordinateur quantique, les AC (ou plus exactement des programmes ou algorithmes simples fonctionnant comme eux) devraient alors nous placer au cœur de l'évolution des systèmes physiques et biologiques de notre univers. Permettront-ils de développer l'analyse des systèmes physiques et naturels que jusqu'à présent ni les technologies ni les mathématiques n'avaient permis de modéliser; comme par exemple ce qu'est la vie, ou la conscience ? Permettront-ils aussi en amont de remonter aux lois fondamentales réglant l'évolution de l'univers, soit sur le plan cosmique macroscopique, soit sur celui, quantique, de la particule élémentaire ? Il s'agirait alors non plus de construire des modèles de la réalité mais de reconstruire la réalité avec les mêmes méthodes et moyens qu'elle a utilisés pour construire notre univers jusqu'à ce jour. Dans cette direction, pourrait-on construire aussi des réalités différentes, telles que celles existant peut-être dans d'autres univers ? Stephen Wolfram pour sa part ne renonce pas à découvrir l'équation du Tout dont rêvent beaucoup de cosmologistes, tels Stephen Hawking. Mais il s'agira alors ici d'une équation simple reposant sur le modèle d'un AC universel obéissant à des règles adéquates elles-mêmes simples, qui restent évidemment à trouver.

Inutile de dire que le concert des physiciens a déjà commencé à s'élever contre de telles perspectives, criant à l'ubris onirique. Certains parlent même d'une escroquerie intellectuelle. Mais eux-mêmes, ou plutôt les plus audacieux d'entre eux, tel David Deutsch précité ou Lee Smolin, un des pères de la gravitation quantique envisagent actuellement des perspectives tout aussi bouleversantes. Nous y avons fait allusion dans le chapitre 1. Bornons-nous à retenir que ces théories existent et sont admises par des hommes dont on pourrait dire sans exagération qu'ils sont les plus intelligents de la génération actuelle. Si les AC permettent de rendre ces théories plus intelligibles que ne le font les formalismes mathématiques, alors il faut les utiliser.

Mais les AC ne sont pas seulement des outils commodes pour procéder à des modélisations ou à des calculs. Il semble bien qu'ils nous introduisent au cœur même d'un autre secret de l'univers profond, s'ajoutant à celui de la génération de

réalités virtuelles – ou plutôt découlant de propriétés identiques : la capacité à générer de la complexité à partir de règles simples, selon des processus qui ne sont pas compréhensibles par l'esprit humain, et par conséquent ni prévisibles ni programmables.

Nous avons déjà évoqué ce phénomène à propos du monde quantique, dont peuvent émerger des phénomènes que l'esprit humain est capable de constater mais qu'il est incapable d'expliquer. De plus en plus de cognitivistes en concluent que le cerveau humain, bien adapté au monde macroscopique, est impuissant à comprendre et peut-être même à voir des phénomènes relevant d'un monde sous-jacent dont la connaissance aux « âges farouches » (5) n'était pas nécessaire à la survie.

Cette impuissance sera-t-elle définitive ou disparaîtra-t-elle au fur et à mesure que le cerveau sera confronté pour survivre à des phénomènes que le développement spontané des sciences et des technologies aura fait émerger sans que nul, évidemment, ne l'ait voulu ? Il s'agit là d'un des enjeux de la robotique autonome et de la vie artificielle.

1 : Jean-Philippe Rennard, La vie artificielle. Où la biologie rencontre l'informatique, Vuibert Informatique, 2002. Les lecteurs pourront se reporter au site du livre ( <http://www.rennard.org/iva/> ) où ils trouveront de nombreuses démonstrations faisant appel à la programmation Java. Voir aussi l'article <http://www.automatesintelligents.com/biblionet/2002/dec/rennard.html>

2 : Wolfram Editions 2002. Le livre a été récemment traduit en français par Bernard François, aux Editions Wolfram. Il n'est pas encore publié. On trouvera toutes informations et démonstrations utiles concernant NKSSur le site <http://www.wolframscience.com/> . Le livre peut être téléchargé gratuitement à partir de l'adresse <http://www.wolframscience.com/nksonline/toc.html>

3 : Précisons que Wolfram, bon mathématicien lui-même, ne propose pas de remplacer les mathématiques par des formules magiques. Il veut seulement les dépasser, là où l'état actuel de développement des techniques mathématiques ne permet pas d'aller suffisamment loin dans la modélisation.

4 : Stephen Wolfram emploie constamment le terme d'aléatoire (random), pour qualifier le degré de complexité supérieur produit par les AC. Mais rien ne nous oblige à penser qu'il s'agisse d'un aléatoire réel, si ce terme pouvait avoir un sens. Il s'agit seulement de constructions dont ni nos sens ni notre esprit ne peuvent suivre les procédures. Il n'y aurait d'aléatoire authentique que dans le monde quantique, et encore cette hypothèse est-elle actuellement fortement

discutée. Voir le dossier Randomness, the Last Superstition, dans le NewScientist 2466 du 25 septembre 2004.

5 : Espérons que nos lecteurs se souviennent encore de Rahan, fils des âges farouches, créé par la bande dessinée de Lécureux et Chéret, responsable semble-t-il de nombreuses vocations scientifiques.