

Bruno BACHIMONT

Nature, Culture et Artefacture : la place de l'intelligence artificielle dans les sciences cognitives

L'intelligence artificielle (IA) est-elle une science cognitive ? Si elle est souvent nommée parmi les nombreuses disciplines cognitives, il n'en demeure pas moins que l'équivoque, quant à savoir si elle est mentionnée à titre de science à part entière ou de simple technologie, est à chaque fois présente. D'une part, l'IA se voit en effet souvent attribuer un rôle ancillaire de technologie au service d'autres disciplines scientifiques, comme la psychologie cognitive dont elle simule les modèles, ou la linguistique dont elle concrétise les théories. L'IA dans ce cas est une informatique avancée permettant de manipuler des symboles selon des lois mises au point dans le cadre d'autres disciplines scientifiques cognitives. D'autre part, l'IA en tant que science est la plupart du temps définie comme un cognitivisme computationnel¹.

¹ L'IA, appréhendée au sein des sciences cognitives, hérite de leur naturalisme (au sens où l'on veut définir l'esprit à partir de la nature matérielle, que ce soit à partir de son environnement physique ou de sa constitution matérielle) et donc est à ce titre un cognitivisme. Nous ne traitons dans le cadre de cet article que du cognitivisme computationnel tel qu'il est formulé par (Pylyshyn, 1984) et (Newell, 1982, 1990), étant donné que la plupart des travaux de l'IA, notamment ceux en acquisition des connaissances (Aussenac et al., 1992), en reprennent les notions et la terminologie. Nous ne mentionnerons dans ce texte que pour mémoire les théories syntaxiques éliminationnistes (Stich, 1983), l'instrumentalisme de (Dennett, 1990) et le cognitivisme physique, pour lequel la connaissance résulte de l'émergence d'invariants cognitifs de dynamiques physiques sous-jacentes internes. À une vision computationnelle du sens, un tel cognitivisme physique oppose une physique du sens

Mais, à l'analyse, la situation de l'IA au sein des sciences cognitives est à notre sens encore plus complexe. Non seulement l'IA hésite entre deux caractérisations possibles, mais surtout aucune des deux ne lui convient. En effet, l'IA comme science, *i.e.* comme cognitivisme computationnel, est pour nous aporétique car elle applique des méthodes formelles inadéquates à son objet, *i.e.* la connaissance. Il semble donc compromis d'inscrire dans ce cadre les modèles développés en IA. Mais si, pour échapper aux contradictions de l'IA comme science, on en fait une simple technologie informatique, on manque ce qui fait sa spécificité, à savoir le traitement informatique de la connaissance. Rejeter la prétention des modèles de l'IA pour n'en faire que des avatars informatiques, c'est jeter le bébé avec l'eau du bain. Si elle est considérée parmi les sciences cognitives, l'IA est donc prise dans un dilemme épistémologique impossible : elle a en effet le choix de se dissoudre dans une technologie informatique ou de s'abîmer dans une science aux fondements aporétiques.

Nous considérons que l'origine de ce dilemme épistémologique gît dans la volonté d'inscrire l'IA dans les sciences cognitives, *i.e.* au sein des sciences naturalistes de l'esprit qui considèrent la connaissance et le sens comme des phénomènes naturels qui, à ce titre, doivent être étudiés par des méthodes empruntées aux sciences dites de la nature. La thèse naturaliste de l'IA comprise comme science est que la connaissance est le phénomène naturel produit par les mécanismes de manipulation syntaxique des symboles. Les problèmes rencontrés par une telle conception conduisent à rapporter l'IA à l'informatique, perdant alors tout ce qui fait l'intérêt de l'IA. Pour éviter ces apories sans perdre l'IA, il faut à notre sens abandonner une telle conception naturaliste de la connaissance.

Nous proposons de considérer que la connaissance n'est pas la *manifestation phénoménale* de la syntaxe, mais résulte bien plutôt de *l'interprétation sémiotique* des signes constituant la littéralité symbolique de la syntaxe. Cette interprétation sémiotique, quoique rationnelle, n'est pas formelle. Cela signifie qu'il est possible de poser des lois descriptives du sens (aspect rationnel), mais pas des

(Petitot, 1992) qui évite bien des apories du cognitivisme computationnel (Petitot, 1990).

lois prédictives (aspect non formel) : on peut décrire *comment* une représentation a du sens, et non pas *pourquoi*. Pour comprendre comment une syntaxe a du sens, il faut étudier pour elles-mêmes les lois d'attribution du sens qui font qu'une représentation symbolique formelle aura tel sens et non tel autre pour un interprétant humain. Ainsi, un système à base de connaissances possède des connaissances dans la mesure exacte où l'utilisateur considère la manipulation des représentations symboliques effectuées par l'ordinateur *comme* des connaissances sur lesquelles un raisonnement est opéré. Cet utilisateur *interprète* les représentations symboliques et les *comprend* comme des connaissances. Il faut donc abandonner l'idée de reproduire la connaissance dans le système informatique par des mécanismes appropriés pour considérer nos systèmes en fonction du sens que nous leur donnons, que nous pouvons leur attribuer. À l'occasion d'un tel renversement copernicien (!) de perspective où, au lieu de faire graviter nos connaissances autour des mécanismes computationnels, ces derniers sont considérés en fonction des connaissances que nous pouvons leur associer, nous proposons d'intituler l'IA *artefacture*, puisque la sortir des sciences cognitives naturalistes pour l'inscrire dans les sciences sémiotiques revient à en faire un moyen terme entre nature et culture. L'*artefacture* est l'IA comprise comme la technologie informatique devant respecter les lois d'attribution du sens.

Dans le cadre de ce bref article, nous nous proposons d'esquisser les principaux points de notre argumentation, (i) en rappelant le projet cognitiviste computationnel de l'IA, (ii) en insistant sur les problèmes posés par ce projet et finalement, (iii) en montrant que l'IA comme *artefacture* peut être comprise comme une réponse donnée à ces problèmes à partir d'arguments empruntés à la tradition philosophique transcendantale.

I. Le problème de l'intelligence artificielle

1. Du nombre au symbole

Selon le dictionnaire (*e.g.* Le petit Robert), l'informatique est le traitement automatique des informations : les ordinateurs permettent

de véhiculer et transformer des informations (au sens de Shannon, *i.e.* ce sont des codages). Essentiellement ce sont des calculateurs universels : ils sont capables d'exécuter toute fonction récursive. Dans les architectures classiques, les données soumises à l'ordinateur comprennent un code représentant la fonction à utiliser et les données sur lesquelles l'appliquer. Tout traitement d'information correspond alors à l'application d'une fonction récursive sur des données numériques : l'opération fondamentale est le codage numérique. Résoudre informatiquement un problème, c'est concevoir une fonction récursive codant un modèle formel donnant la solution qui sera appliquée sur un codage des données du problème. Selon cette conception, tout se ramène à des nombres et à des calculs : un problème n'est donc traitable automatiquement que s'il possède une modélisation formelle codable numériquement.

Une conception concurrente mais équivalente théoriquement considère les objets informatiques non comme des nombres mais comme des symboles (Newell, 1980). Ce changement de perspective entraîne une nouvelle façon de résoudre informatiquement des problèmes. En effet, pour résoudre un problème, on utilise des connaissances. Comme le rappelle (Ganascia, 1991), une façon habituelle et commode de définir une connaissance consiste à la caractériser comme ce qui relie un sujet à un objet. C'est dans la connaissance que quelque chose peut se *constituer* comme objet pour un sujet. L'ensemble de ces objets constitue une *objectivité*. Il y a autant de types de connaissance que de façons de constituer un objet, *i.e.* que d'objectivités ; réciproquement il y aura autant d'objectivités que de type de connaissances. Ces précisions sont utiles car il existe plusieurs types de connaissances. Un premier type est la connaissance scientifique, dont la physique mathématique est l'archétype². Les théories mathématiques de la physique définissent des modèles de la réalité dont la pertinence est fixée par voie expérimentale. Pour résoudre automatiquement un problème, une méthode peut être alors

² Les activités de recherche ne s'intitulent souvent *science* que dans la mesure où elles empruntent à la physique mathématique son style démonstratif et expérimental. Une connaissance ne sera souvent dite scientifique que dans la mesure où, de formalisme mathématique, elle se rapporte à une objectivité physique.

de disposer d'un tel modèle qui, formulé mathématiquement, peut être rendu effectif et calculable par un codage informatique. Le programme informatique obtenu hérite de la propriété du modèle physico-mathématique de décrire une réalité physique ainsi qu'une validation expérimentale permet de s'en assurer. Pour ce type de connaissances, l'approche informatique de la résolution automatique de problèmes est donc le traitement automatique de codages de modèles physico-mathématiques.

Mais il existe des connaissances autres que les connaissances scientifiques. Ainsi les connaissances que nous appellerons *phénoménologiques* constituent une objectivité phénoménologique composée des objets familiers peuplant la quotidienneté, avec lesquels un être humain sait parfaitement interagir. A cette objectivité appartiennent aussi des objets théoriques non mathématiquement formalisés comme les signes médicaux, les diagnostics, les classes biologiques, etc. Nous distinguons donc *phénoménologique* et *scientifique*. Tandis que le premier désigne un rapport direct au *monde*, tel qu'il se présente à notre perception et notre réflexion, le second renvoie à l'appareillage technique de la physique dite expérimentale assurant la médiation entre le sujet et la *nature* qu'il étudie. Alors que le premier se traduit par des descriptions qualitatives et emprunte la langue naturelle pour s'exprimer, le second se traduit par des mesures quantitatives et emprunte les langages formels scientifiques pour s'exprimer.

Ces deux objectivités, la scientifique et la phénoménologique, ne sont pas identiques. La connaissance scientifique est en effet une théorie formelle mathématique fixée par la mesure expérimentale de la nature et elle construit des objets comme les électrons, les quarks, les cordes, etc. Ces derniers ne sont pas des objets de la phénoménologie, qui, pour sa part, est une description qualitative fixée par la compréhension et l'interprétation du monde, et qui constitue des objets comme les tables, les diagnostics, etc. qui ne sont pas des objets de la physique³. Par exemple, en médecine, on

³ Nous ne pouvons aborder dans le cadre limité de cet article les problèmes profonds de la modélisation physico-mathématique et des rapports complexes entre une écriture symbolique réglée et l'objectivité qu'elle permet d'appréhender. Pour ces questions nous renvoyons à (Bachimont, 1992) et à sa bibliographie où nous

aura des connaissances scientifiques comme les équilibres électrolytiques et des connaissances phénoménologiques comme les corrélations signes / diagnostics. Dans le premier cas on *mesure* la réalité physique des équilibres, dans le second on *interprète* la réalité phénoménologique des signes par un diagnostic.

Dans de nombreux cas, nous ne disposons pas de connaissances scientifiques pour résoudre les problèmes que nous rencontrons et nous nous appuyons plutôt sur nos connaissances phénoménologiques. S'il était possible d'utiliser informatiquement ces connaissances phénoménologiques, c'est toute une nouvelle classe de problèmes, pour lesquels les connaissances de type scientifique manquent, qu'il serait possible de traiter automatiquement. Or, considérer les programmes informatiques non comme un codage mais comme une représentation suscite l'espoir de pouvoir utiliser ces connaissances et de traiter ainsi ces problèmes qui restaient jusque là hors de la portée du traitement informatique.

En effet, les connaissances phénoménologiques portent sur la structure qualitative du monde linguistiquement décrit. En représentant symboliquement ces descriptions, on espère conserver la structure qualitative pertinente pour la réalité décrite et qui permet de considérer la représentation comme une connaissance. Si bien qu'en manipulant les représentations obtenues pour résoudre un problème, on obtient une représentation solution qui sera la connaissance résultat : pour un tableau de signes cliniques, on manipule des représentations jusqu'à obtenir une représentation dont la structure qualitative est telle qu'elle correspond à ce que vise la connaissance phénoménologique qu'est le diagnostic.

De la même manière que le codage conserve le sens mathématique et physique en rendant effectif des lois quantitatives, si bien que le problème est effectivement résolu par l'ordinateur (on *résout numériquement* des équations différentielles par exemple), la représentation symbolique des connaissances phénoménologiques

abordons ces questions depuis l'épistémologie et phénoménologie transcendantales. Nous utilisons ici l'acception habituelle de *modèle* comme représentation d'une objectivité dont la pertinence doit être définie par une confrontation à cette objectivité, expérimentale pour les modèles physico-mathématiques.

doit permettre de conserver leur description qualitative du monde pour que la manipulation des représentations reproduise bien le raisonnement en termes de connaissances et conduise à la solution.

L'informatique devient ici manipulation formelle de représentations symboliques des connaissances phénoménologiques. Ainsi, du point de vue de la résolution de problèmes, alors que l'informatique s'était affirmée comme le traitement automatique des informations, c'est-à-dire du codage des connaissances scientifiques, l'intelligence artificielle peut donc se définir comme le traitement automatique des représentations, c'est-à-dire de la représentation symbolique des connaissances phénoménologiques. Les modèles qu'elle conçoit sont donc des modèles qualitatifs, comme l'a souligné (Clancey, 1989) et portent de façon privilégiée sur les domaines où le savoir se présente sous cette forme, ainsi qu'on peut s'en assurer en considérant ses réalisations ((Aïmeur, 1993) sur la classification en ichtyologie par exemple, ou bien (Volot et al., 1993) en médecine).

2. Le problème de la connaissance

Le codage informatique de modèles physico-mathématiques permet de résoudre un problème dans la mesure où le codage hérite de la pertinence du modèle à l'égard de la réalité qu'il modélise. La connaissance scientifique s'est en effet formée autour de cette relation à la réalité : la connaissance devint scientifique quand elle fut mathématiquement déduite et formulée et expérimentalement confirmée ou infirmée ; tout le problème étant d'articuler la forme du concept et sa valeur objective⁴. On peut donc considérer, au moins

⁴ Quand j'ai la sensation de la vision d'une chaise dans l'esprit, je suis sûr de voir une chaise, mais non qu'il y ait réellement une chaise. Mon idée de chaise a donc un contenu objectif, il s'agit d'une chaise, mais pas forcément de valeur objective, la chaise peut ne pas exister. Descartes montra que les idées empiriques (pour nous, phénoménologiques) étaient dépourvues de valeur objective, et posa métaphysiquement que les idées mathématiques possédaient non seulement un contenu objectif, mais aussi une valeur objective. La physique expérimentale montre comment une telle valeur objective peut être mise en évidence et problématisée sans recours métaphysique pour les objets de l'objectivité scientifique. On a donc des idées mathématiques à valeur objective pour l'objectivité scientifique et des idées

à titre pratique puisque les modèles de la physique semblent fonctionner, que ces modèles possèdent une valeur objective qui se propage à leur codage informatique. C'est ainsi que les problèmes peuvent être effectivement résolus par un traitement informatique.

Mais qu'en est-il lorsqu'il s'agit de représenter des connaissances phénoménologiques pour manipuler ces représentations et résoudre un problème ? Dans quelle mesure hérite-t-on dans les représentations informatiques de la validité phénoménologique qui fait que ces connaissances phénoméno-logiques permettent effectivement aux êtres humains de résoudre leur problème ? La question est donc : comment des expressions formelles peuvent-elles suffire pour la résolution automatique d'un problème nécessitant par ailleurs des connaissances quand il est résolu par des humains ? En clair : en quoi des expressions formelles dans la machine peuvent-elles avoir une valeur objective pour l'objectivité phénoménologique ? En quoi une représentation symbolique formelle permet-elle à une machine de posséder la connaissance phénoménologique que cette représentation est censée modéliser ? Il faut s'attacher à caractériser scientifiquement les connaissances phénoménologiques pour savoir d'où elles procèdent et pouvoir les reproduire par des programmes. Ce sera l'ambition naturaliste du cognitivisme computationnel.

II. Le problème de la signification

1. Fonctionnalisme et systèmes symboliques physiques

La connaissance phénoménologique s'exprime par une description de type linguistique : cela signifie qu'elle n'est pas seulement un *flatus voci*, mais qu'elle est pourvue de sens. C'est parce qu'elle est davantage qu'un signal acoustique ou une suite de traces sur un support qu'une représentation ou description est une connaissance: c'est parce qu'elle est pourvue de sens. Ce sens est

empiriques sans valeur objective pour celle-ci. Le problème est que ces idées empiriques sont néanmoins valables phénoménologiquement (résolution des problèmes que traite l'IA), selon une validité qui échappe à la science. Elles possèdent une valeur objective pour l'objectivité phénoménologique.

de nature linguistique, et en tant qu'il sert à véhiculer des connaissances phénoménologiques, nous parlerons du contenu phénoménologique de la représentation ou description. Une représentation symbolique est une connaissance dans la mesure où elle possède un contenu phénoménologique. On peut donc se demander dans quelle mesure une représentation, pourvue d'un sens ou contenu phénoménologique, peut encore avoir un tel contenu quand elle est insérée dans un programme. La question est: quel est le sens, le contenu, d'un programme ?

Les signes et représentations manipulés par l'ordinateur ne sont tels que pour l'utilisateur qui les considère comme des signes ou représentations et leur associe des significations extérieures à la machine. Cette dernière ne traite les signes et représentations que comme des formes syntaxiques dépourvues de signification. L'informatique est souvent posée comme une *physique* des signes, physique abstraite puisqu'elle ne s'intéresse pas à la matière (le silicium ou l'arsenure de gallium) constituant les représentations, mais physique néanmoins puisqu'elle pose les lois des calculs mécaniques effectifs. Elle ne s'intéresse pas au sens de la syntaxe de ses algorithmes. *A priori*, les programmes sont donc des codes et ne possèdent aucun contenu de signification de quelque nature que ce soit.

Cependant cette vision n'est pas la seule possible. Une autre conception concurrente apparaît quand la signification n'est pas comprise comme association mais comme opération. Le paradigme théorique suivi est le lambda calcul. La sémantique qui en résulte est une *sémantique fonctionnelle ou opérationnelle*. L'idée essentielle est qu'une fonction n'est pas un graphe, comme le veut la sémantique déclarative, mais une opération donnant un résultat. La signification d'une expression est alors le résultat obtenu quand toutes les opérations qu'elle contient sont effectuées. Une expression associée à un évaluateur effectuant ces opérations possède une signification. Le sens de cette expression est dérivé car il repose sur l'évaluateur : une expression ne possède aucun sens en elle-même. Mais l'ensemble < expression + évaluateur > possède lui un sens intrinsèque : le résultat obtenu. Si l'évaluateur peut être incorporé à la machine (*e.g.* Lisp), on obtient alors une machine pouvant associer un sens aux expressions (signes ou

représentations) qu'elle manipule. C'est parce qu'il y a processus qu'il y a sens : la signification est obtenue à la suite d'une opération interne à la machine et non par une correspondance avec un domaine d'interprétation établie de façon externe.

Ainsi une fonction est-elle appliquée non sur les expressions formelles de ses arguments, mais sur les valeurs résultant de l'évaluation de ces expressions. *Une action exprimée par une fonction tient compte du sens de ses arguments et pas seulement de leur expression formelle.* Il est donc possible d'affirmer que les ordinateurs sont le siège d'un véritable pouvoir de désignation, dénotation ou représentation : une entité X représente une entité Y pour un processus P si le comportement de P appliqué à X dépend de Y (Newell, 1980). Si P est une fonction f, et l'entité est $1 + 2$, f appliquée à $1+2$ va dépendre de 3. Donc $1+2$ représente 3 pour la fonction f. $1+2$ tient le rôle fonctionnel de 3 pour la fonction f dans le programme : ce sera le sens de $1+2$ pour le programme.

Selon cette argumentation, les ordinateurs sont des systèmes symboliques physiques (Newell, 1980) dans la mesure où d'une part, ce sont des machines matérielles construites selon les lois de la physique, et d'autre part, leur fonctionnement matériel donne un rôle symbolique et fonctionnel aux codes qu'elles manipulent. Les codes, matériels, se comportent comme des symboles car désignent quelque chose qu'ils ne sont pas eux-mêmes. On a donc une intentionalité originelle dans la mesure où les symboles désignent d'eux mêmes quelque chose d'autre qu'eux-mêmes.

On obtient donc un fonctionnalisme dans la mesure où le sens d'un symbole est le rôle fonctionnel qu'il tient dans le programme le contenant. À un niveau physique (*hardware*) vient donc s'ajouter un niveau symbolique (niveau des programmes) déterminant le sens symbolique des éléments matériels du niveau physique. Un point essentiel est que ce sens symbolique, quoique dépendant pour exister du niveau physique qui l'implante, ne dépend pas de lui dans sa nature. Le sens n'est pas une propriété physique du *hardware*, mais une propriété computationnelle du *software*. Les propriétés informatiques des programmes ne sont pas des propriétés physiques des composants micro-électroniques qui les implantent. Ce fonctionnalisme est un matérialisme (le sens symbolique dépend de la matière physique dans la mesure où chaque symbole est

physiquement réalisé, pas de *software* sans *hardware*) non réductionniste (le niveau des symboles est abstrait — dans le sens où la machine de Turing est abstraite). Les connaissances scientifiques qui étudient la matière physique sont la physique mathématique qui en détermine le contenu physique et celles qui étudient les symboles sont l'informatique, cette physique des signes, qui fixe leur contenu fonctionnel.

Un tel fonctionnalisme, fondé sur la sémantique opérationnelle inhérente aux systèmes symboliques physiques, montre comment les expressions formelles de l'informatique ne sont pas vides de contenu, mais qu'elles possèdent un sens. Nous parlerons donc du *contenu computationnel* ou *contenu fonctionnel* des symboles et des représentations. Si la syntaxe donne les règles pour assembler des signes, des éléments matériels sur un bout de papier ou d'ordinateur, la sémantique opérationnelle nous dit ce que veulent dire ces signes quand ils s'assemblent en un programme tournant sur une machine. Le fonctionnalisme a donc répondu à la moitié de la question, qui est de savoir comment des expressions formelles peuvent avoir du sens pour une machine. En effet, il affirme que ces expressions ont un contenu. Il faut maintenant argumenter que ces contenus sont bien les contenus phénoménologiques qui font qu'une représentation est une connaissance phénoménologique.

Ce fonctionnalisme rejoint celui de Putnam (Lycan, 1990) pour lequel une structure physique neuronale (matière) n'est un état mental (esprit) que si cet état matériel joue un rôle computationnel dans le comportement décrit comme un programme. On sait qu'un état mental correspond à une attitude⁵ comme la croyance, le doute, la certitude etc., en relation avec une description ou proposition (Fodor, 1981). Par exemple, le système cognitif "médecin" aura l'attitude de croyance vis-à-vis de la proposition ou description suivante: "le patient présente un signe d'infarctus". Les états

⁵ Nous considérons une sous-classe des états mentaux traités par Putnam : pour lui la douleur par exemple est aussi un état mental, alors que pour nous non, car elle n'a pas de contenu propositionnel. Les états mentaux sont ici des états intentionnels au sens de Searle et de Husserl (cf. Dreyfus, 1991, note à ce sujet).

mentaux sont donc des connaissances phénoménologiques⁶, *i.e.* des attitudes vis-à-vis de descriptions qualitatives, et c'est dans cette mesure que l'on retrouve pour les traiter la psychologie dite de sens commun où l'on décrit un comportement d'une personne non pas au moyen d'une explication biochimique causale, mais au moyen d'une attitude propositionnelle. Le fonctionnalisme de Putnam affirme donc qu'une attitude propositionnelle est une explication du comportement ssi elle est le nom donné au rôle computationnel d'un état neurophysiologique. Les contenus phénoménologiques sont des contenus computationnels.

On a donc une réponse complète au problème de l'IA : le fonctionnalisme est l'explication scientifique des connaissances phénoménologiques qui peut être reproduite sur l'ordinateur puisque le contenu des connaissances s'avère n'être autre que le sens dont se trouve pourvue toute représentation symbolique manipulée par un programme.

2. *Le niveau des connaissances*

Le fonctionnalisme implique que tout système computationnel est un système cognitif possédant des connaissances, puisqu'une connaissance est un rôle computationnel. Il doit donc être possible d'associer à tout programme les connaissances qu'il contient. Or, cela n'est pas toujours possible (Pylyshyn, 1984). Il ne suffit pas de caractériser les connaissances comme des programmes, il faut en outre que ces programmes respectent des *contraintes intentionnelles*, *i.e.* possèdent des significations, comprises ici comme des états de choses du monde associés. Ces états de choses correspondent à la structure qualitative du monde telle qu'elle est visée par les connaissances phénoménologiques. Cela revient à dire que ces dernières ne sont pas seulement des contenus computationnels, mais qu'elles doivent avoir un contenu phénoménologique ou *contenu intentionnel* distinct des contenus computationnels : non seulement une connaissance est un

⁶ Cela explique que, pour édifier des modèles qualitatifs du monde, l'IA a souvent demandé à la psychologie cognitive de lui fournir une description computationnelle du fonctionnement cognitif humain (Richard, 1990).

programme, mais c'est un programme sémantiquement interprété. Il faut donc caractériser les systèmes cognitifs, *i.e.* les systèmes possédant des connaissances phénoménologiques, non seulement en termes computationnels au niveau des programmes, mais aussi en termes intentionnels au *niveau des connaissances* ou *Knowledge Level* (Newell, 1982).

Le contenu intentionnel ou phénoménologique, visant des états de choses du monde, est compris dans cette théorie comme une sémantique formelle dénotative associée au programme. Les contenus phénoménologiques sont donc compris ici comme des *contenus dénotationnels*, alors que le fonctionnalisme les envisageait comme des contenus fonctionnels. Ces contenus ne sont donc plus intrinsèques comme les contenus fonctionnels, mais extrinsèques, puisqu'ils résultent de l'attribution d'une dénotation par un observateur.

On a un niveau des connaissances dans la mesure où les programmes peuvent être décrits comme des connaissances phénoménologiques en leur associant une structuration qualitative du monde à travers une sémantique dénotationnelle. Les programmes interprétés sont compris comme les connaissances d'un système cognitif ou agent : ce sont les buts qu'il poursuit, les actions qu'il peut entreprendre, et les croyances qu'il a sur le monde. Il se comporte conformément au principe de rationalité selon lequel il accomplit les actions pour atteindre ses buts en fonction de ses croyances. Un tel système se caractérise alors comme un *agent rationnel*.

La théorie du niveau des connaissances affirme que les systèmes physiques sont cognitifs dans la mesure où des descriptions intentionnelles sont nécessaires pour expliquer et comprendre leur comportement⁷. Ces systèmes restent incompréhensibles pour

⁷ Le lecteur aura reconnu la définition même de la stratégie intentionnelle de Dennett (Dennett, 1990), selon laquelle un système est cognitif si le fait de le considérer comme un agent rationnel constitué d'états intentionnels améliore notre potentiel prédictif à son égard, *i.e.* on prévoit et explique mieux son comportement. La différence avec la théorie du niveau des connaissances est que Dennett ne cherche pas à articuler le niveau fonctionnel et le niveau intentionnel alors que cela est crucial pour Newell, comme nous le rappelons dans la suite de notre texte.

l'observateur, inintelligibles, s'il ne leur associe pas une dénotation. (Pylyshyn, 1984) constate donc que, pour étudier scientifiquement ces systèmes, il est nécessaire de le faire au niveau des connaissances. Mais, poursuit-il, pour que cette explication soit valable, il faut qu'elle possède un soubassement computationnel : si tout programme n'est pas un agent cognitif, tout agent cognitif doit se décrire comme un programme. Pourquoi ce recours aux programmes ?

Le niveau des connaissances permet d'expliquer le comportement de l'ordinateur exécutant un programme. Mais qu'entend-on ici par *expliquer* ? On peut opposer une *explication formelle* ou explication qui s'exprime par des connaissances scientifiques (e.g. les lois de la physique) et une *explication interprétée* ou *explicitation intentionnelle* qui s'exprime en connaissances phénoménologiques (e.g. expliquer à quelqu'un), en représentations interprétées. D'après ce que nous avons dit, le niveau des connaissances est un niveau d'explicitation, et le niveau des programmes un niveau d'explication.

Il est nécessaire d'asseoir l'explicitation intentionnelle sur l'explication computationnelle dans la mesure où la première n'est pas une relation causale⁸, si bien qu'elle ne peut prétendre au statut d'explication scientifique : les connaissances scientifiques doivent rendre compte de la causalité. S'il est nécessaire de comprendre le système au niveau des connaissances, en associant au systèmes des états choses qu'il vise, réalise ou croie, il faut l'expliquer causalement au niveau des programmes (puisque les programmes, implantés dans la matière, possèdent une efficace causale). L'explicitation au niveau des connaissances ne sera scientifique que si elle repose sur l'explication causale computationnelle au niveau des programmes⁹. Ainsi, un système cognitif comme l'ordinateur ou

⁸ "The basic problem is that *representing* is a semantic relation, that semantic relations, like logical relations, appear not to be causally *definable* ... though each case of a semantic relation can be accompanied by *some* causal relation, the two are not type equivalent." (in Pylyshyn, 1984, p.42).

⁹ "The reason there must be symbolic codes, of course, is that, because they are instantiated as physical patterns, they can enter into causal relations. Unlike semantic

l'être humain, est incompréhensible au niveau des programmes et inexplicable au niveau des connaissances. Cependant, si à chaque interprétation on fait correspondre un programme, on comprend le système en restant compatible avec le matérialisme nécessaire à une explication causale¹⁰.

On peut alors comprendre la relation unissant l'explicitation intentionnelle à l'explication computationnelle comme une relation de manifestation. C'est par ce qu'on comprend un système cognitif en lui attribuant des connaissances phénoménologiques (manifestation du sens) qu'on déduit une explication computationnelle dont le fonctionnement dans ce système provoquera chez l'observateur l'explicitation intentionnelle qui manifestera cette explication¹¹. La signification, la connaissance associée à un système est la manifestation phénoménale par laquelle il est possible d'accéder à sa réalité physique sous-jacente, le programme. Dans l'investigation de ce qui fait qu'un système physique est un système cognitif, le programme est la *ratio essendi* et la signification la *ratio cognoscendi*.

Le problème de l'IA se résout donc ainsi : le traitement d'un problème en termes de connaissances phénoménologiques est la

contents, tokens of symbols can actually cause the system to behave in certain ways." (in Pylyshyn, 1984, p.29).

¹⁰ "A computational process is one whose behavior is viewed as depending on the representational or semantic content of its states. [...] the formal symbol structures mirror all relevant semantic distinctions to which the system is supposed to respond and continue to do so when certain semantically interpreted rules are applied to them, transforming them into new symbol structures [...] Thus, any semantic feature that conceivably can affect behavior must be syntactically encoded at the level of a formal symbol structure. By this means we arrange for the system's behavior to be describable as responding to the content of its representations — to what is being represented — in a manner perfectly compatible with materialism." Pylyshyn, 1984, p. 74).

¹¹ Salanskis (communication personnelle) écrit en commentant (Pylyshyn, 1984) : "le point important est que la notion de *représentation* aurait en sciences cognitives la valeur d'une dimension intuitive dans laquelle les phénomènes cognitifs se présentent." La notion d'intuition étant prise en un sens kantien, cela signifie explicitement que la représentation interprétée est comprise comme une manifestation phénoménale.

manifestation du fait qu'un programme symbolique est à la base de la résolution. En reproduisant ce programme sur ordinateur, on aura la même manifestation, c'est-à-dire qu'on le comprendra comme un agent rationnel résolvant un problème à l'aide de ses connaissances phénoménologiques. On aura bien une résolution automatique du problème puisque c'est ainsi que l'on comprendra ou interprétera le comportement du programme.

3. *Dualisme et naturalisme*

Le cognitivisme computationnel reconnaît donc qu'il existe des phénomènes particuliers, les explicitations intentionnelles, qui peuvent être mis en évidence comme l'explicitation du comportement de systèmes vus comme des agents rationnels. Mais cette explicitation, sans efficace causale, doit se muer en une explication, fondée sur une causalité physique mathématiquement légalisée. L'explication physique (abstraite) de l'explicitation intentionnelle est l'objectivité computationnelle.

Or, la distinction effectuée par le fonctionnalisme entre le programme et son implantation fait que l'on obtient un dualisme ontologique. En effet, d'une part, la matière physique, constituant les substrats matériels des programmes, obéit aux lois de la physique et possède son propre ordre de manifestation phénoménale ou phénoménalité, la mesure expérimentale. D'autre part, une matière symbolique (au sens des systèmes symboliques physiques), implantée dans la matière physique, obéit aux lois de la computation et possède sa propre phénoménalité, l'interprétation intentionnelle. Le cognitivisme computationnel (Fodor et McLaughlin, 1990) affirme que la phénoménalité de la signification ne peut en aucun cas émerger de la matière physique, mais seulement de la matière symbolique. On a donc bien un dualisme de l'objectivité physique matérielle, construite à travers les lois de la physique mathématique, se manifestant *via* l'expérimentation, et l'objectivité symbolique ou computationnelle (objectivité physique abstraite), construite à travers les lois de l'informatique, se manifestant *via* les explicitations intentionnelles ou connaissances phénoménologiques.

En conclusion, le naturalisme qui veut voir dans la connaissance et la signification des phénomènes naturels qui doivent à ce titre être étudiés par les méthodes des sciences de la nature conduit le cognitivisme computationnel à dégager un nouvel ordre du réel, l'objectivité computationnelle, et à le poser comme le fondement des significations, qui s'avèrent d'autant plus réelles qu'elles sont indispensables pour comprendre et expliciter les systèmes dits cognitifs. Le naturalisme conduit donc au dualisme ontologique de la matière et du symbole.

III. L'artefacture

En assimilant l'esprit à un ordinateur, le cognitivisme aboutit à un dualisme ontologique, qui contredit directement le naturalisme à la base de l'entreprise cognitive, qui est un monisme matérialiste. L'IA comme science ne peut exister sur de telles bases. Mais il est impossible de la décrire comme un simple technologie informatique sous peine de ne pas comprendre l'intérêt de ses modèles. Il faut sortir de cette impasse. Le problème vient du statut accordé à la signification, *i.e.* aux contenus phénoménologiques. En se démarquant du fonctionnalisme, le cognitivisme affirme que les significations ne sont pas des illusions subjectives perturbant la science cognitive mais sont véritablement une phénoménalité dont il faut rendre compte de manière computationnelle. Mais, et c'est là le cœur du problème, il faudrait être sûr que les significations fussent réellement des phénomènes naturels à l'instar de nos bons vieux phénomènes tels que la physique les étudie.

1. Objets et phénomènes

Il convient donc de préciser quelque peu les notions clefs de phénomènes et d'objets. Etymologiquement, un phénomène est ce qui apparaît. En s'appuyant sur une terminologie kantienne, on dira que les phénomènes sont du divers spatio-temporel, c'est-à-dire des contenus inscrits dans l'espace et le temps qui sont les conditions *a priori* de toute perception. Ces phénomènes sont des apparitions et non des apparences qu'il faudrait récuser pour dévoiler une réalité sous-jacente (Deleuze, 1963). Ces apparitions sont déterminées les

unes par rapport aux autres par des lois de type mathématique. Les apparitions et leur relations mathématiques constituent des objets qui ressortissent à l'objectivité scientifique. Ainsi, les phénomènes correspondent à ce que les capteurs permettent de mesurer en physique. Les lois sont les lois mathématiques reliant ces mesures et les déterminant. Enfin, des objets (*e.g.* électrons, quark, mouvement) peuvent être posés, dont l'objectivité réside uniquement dans les lois qui leur sont associées et dans les phénomènes correspondant à ces lois. Ainsi, le concept d'électron aura un contenu physique uniquement déterminé par les lois de la physique, indépendamment de toute autre signification que l'on voudrait lui associer. On dit alors que ces lois modélisent les phénomènes et schématisent les objets. Or, notre problème est, rappelons-le, de savoir si le cognitivisme peut légitimement poser que les connaissances phénoménologiques sont des phénomènes manifestant une objectivité dont les lois sont de type computationnel. Notre réponse tient en deux temps.

Premièrement, les connaissances phénoménologiques ne sont pas une phénoménalité que l'on doit traiter par des lois formelles prédictives mais une phénoménologie que l'on doit traiter par des lois rationnelles non formelles descriptives ; cela revient à dire que les contenus phénoménologiques n'obéissent pas aux lois des contenus fonctionnels ou dénotationnels associés aux programmes. Il n'est pas possible de rendre compte des connaissances phénoménologiques en les associant dénotationnellement à des connaissances scientifiques car les premières n'obéissent pas aux lois formelles gouvernant les secondes, Deuxièmement, les connaissances phénoménologiques ont des contenus sémiotiques dont l'étude ne ressortit pas à l'informatique mais aux sciences sociales de type sémiotique. Cette irréductibilité des contenus phénoménologiques aux contenus dénotationnels ou fonctionnels, niée par le cognitivisme, a été soulignée au début de ce siècle par Husserl¹².

¹² Les sémantiques dénotationnelles et fonctionnelles sont postérieures à Husserl si bien qu'on ne saurait l'y opposer directement sans anachronisme. Cependant, son opposition à la réduction du sens phénoménologique au sens mathématique permet d'affirmer cette thèse.

2. Phénoménologie et éidétique descriptive

La phénoménologie transcendantale husserlienne (e.g. *Recherches Logiques* (Husserl, 1959) et des *Ideen I* (Husserl, 1950)) s'est présentée comme un retour aux choses mêmes : *Nach den Sachen selbst* ! Ce retour se traduit par la description des choses telles qu'elles se présentent à la conscience. C'est ainsi que la phénoménologie montre comment les objets du sens commun se constituent dans la conscience à partir d'une synthèse continue d'actes ou vécus intentionnels visant un objet : en voyant une maison, je ne perçois en fait que la face visible. En tournant autour, ma perception se modifie, cependant que j'ai conscience de voir la même maison. Cette dernière se constitue comme la synthèse unitaire de ces perceptions partielles. Ce n'est donc pas un objet forcément existant du monde, puisque sa réalité, constituée dans la conscience, est simplement présomptive. La maison, comprise dans ce sens d'unité présomptive de mes vécus, est le *sens* ou *signification* de ma perception, puisque mes sensations sont *appréhendées, interprétées, comprises*, comme le contenu perceptif de la vision d'une maison. Par ailleurs, les significations sont idéales, ce sont des *eidos* : d'une part, elles sont partagées par plusieurs vécus et d'autre part, elles ne sont pas des composantes réelles, effectives de ces vécus, puisqu'elles ne sont que présomptives¹³.

Le but de la phénoménologie est de décrire quelles sont les relations idéales entre les significations, comment elles se constituent et se combinent. La phénoménologie se fait donc éidétique descriptive. Les connaissances phénoménologiques sont des entités idéales dont la phénoménologie étudie les lois descriptives, et elles ne sauraient être réduites à des faits naturels étudiés par des lois formelles prédictives de type scientifique. La méthode de description des significations ou *eidos* proposée par Husserl, la

¹³ La sémantique n'est plus dénotationnelle (association d'un objet déjà là à un signe), mais constituée (constitution, à travers l'intentionnalité de la conscience, d'un objet visé). Alors qu'auparavant, l'intentionnalité était seconde et l'objet premier, à présent, l'intentionnalité est première et l'objet second.

variation éidétique par substitution et différence, est très proche de la méthode différentielle pratiquée en sémantique et sémiotique (Rastier, 1987, Bachimont, 1992). On peut poser alors que la description phénoménologique caractérise les contenus sémiotiques des connaissances phénoménologiques. Les contenus phénoménologiques sont des contenus sémiotiques.

La phénoménologie est donc fondamentalement descriptive¹⁴, ne peut que l'être¹⁵ et ressortit aux sciences sociales sémiotiques. Cela signifie que les contenus sémiotiques dégagés par la description phénoménologique ne sont pas et ne peuvent être des contenus formels, *i.e.* physiques, computationnels ou dénotationnels. Cela signifie donc que les significations ou contenus sémiotiques non seulement constituent une phénoménologie et ne sont pas une phénoménalité, mais ne *peuvent* constituer une phénoménalité. Cette opposition marquée par Husserl rejoint celle qu'il effectue dans les *Ideen II* entre les lois de la *motivation* décrivant le sens du monde sans en poser l'existence et les lois de *l'explication* recherchant les causes à partir de la présupposition du monde. La phénoménologie des significations décrit leurs relations non causales de motivation alors que la science de la phénoménalité des significations rechercherait plutôt à les articuler selon des lois causales explicatives; ce que fait le cognitivisme.

3. *Le constitutif et le régulateur*

¹⁴ Husserl s'intéresse à la définition " d'une *théorie objective de la connaissance*, et s'y rattachant intimement, d'une *phénoménologie pure des vécus de la pensée et de la connaissance*. Celle-ci, comme la *phénoménologie pure des vécus en général* qui l'englobe, a exclusivement affaire, dans une généralité d'essence pure, aux vécus qu'on peut appréhender dans l'intuition et analyser [phénoménologie], mais non aux vécus aperçus empiriquement comme des faits réels, en tant que vécus d'hommes ou d'animaux les vivant dans le monde des phénomènes, [phénoménalité] posée comme fait d'expérience." *in Recherche Logique I* p. [2].

¹⁵ " Il faut bien voir en outre que les *sciences exactes* [phénoménalité] et les *sciences purement descriptives* [phénoménologie] ont bien entre elles une liaison, mais qu'elles ne peuvent jamais être prises l'une pour l'autre et que quel que soit le développement d'une science exacte, c'est-à-dire opérant avec des infrastructures idéales, elle ne peut résoudre les tâches originelles et autorisées d'une description pure." *in Ideen I*, p. [139].

En refusant de tenir compte de l'impossibilité soulignée par Husserl d'appréhender par les méthodes des sciences exactes de la nature ces objets dégagés par l'analyse éidétique descriptive que sont les significations, le cognitivisme doit affronter une antinomie au sens kantien du mot : la première thèse de cette antinomie est que tous les états mentaux doivent être expliqués causalement en termes de contenus computationnels. La seconde thèse est que les états mentaux doivent être explicités de façon non causale en termes de significations ou contenus sémiotiques.

Stich dans Stich (1983), constate que si l'on considère rigoureusement les significations, alors il s'avère impossible de les ramener à des contenus computationnels; il prône une théorie purement syntaxique de l'esprit et élimine toute interprétation sémantique. Mais, comme le constate Pylyshyn, un tel éliminationisme des contenus ne rend pas compte du fait que la meilleure façon de comprendre et d'utiliser certains systèmes est de les considérer comme des agents rationnels (cf. *supra*). Les explicitations intentionnelles sont de véritables apports scientifiques dont il faut conserver les contenus sémiotiques. Mais ceux-ci ne sont pas soumis à la causalité à laquelle obéissent pour leur part les contenus computationnels. De manière analogue à Kant dans Kant (1986)¹⁶, on est donc confronté à une théorie formelle (le fonctionnalisme) insuffisante et impuissante à rendre compte de la réalité considérée (le comportement des systèmes cognitifs).

Le cognitivisme affirme donc qu'il faut utiliser la compréhension que l'on a du système, son explicitation intentionnelle, nos connaissances phénoménologiques à son sujet, pour l'expliquer

¹⁶ Kant constate que les causes efficientes mathématiquement légalisées ne permettent de construire qu'une nature en général, et non la nature particulière qui nous entoure. Il faut alors pour l'expliquer poser des principes régulateurs, correspondant aux causes finales. Si ces dernières sont considérées au même niveau que les lois mathématiques de la nature (attitude métaphysique), on obtient l'antinomie selon laquelle la nature obéit à la fois aux causes finales et aux causes efficientes. Résoudre cette antinomie consiste à poser que la nature obéit aux causes efficientes et que nous la pensons selon les causes finales. Les premières sont constitutives (construire une théorie de la nature), les secondes régulatrices (penser cette théorie comme si elle se conformait à des causes finales).

causalement en termes de programmes. Or, expliciter intentionnellement un système, c'est l'expliquer par les causes finales (buts et croyances de l'agent). On recourt donc aux causes finales (explicitation) pour en déduire les causes efficientes (calcul symbolique) en supposant que les lois auxquelles obéissent les premières sont les mêmes que celles gouvernant les secondes. Mais, la légalité formelle et scientifique expliquant ces causes finales ne peut être issue des lois de la matière physique (cf. *supra*), si bien qu'en réduisant le contenu phénoménologique au contenu computationnel, on résout l'antinomie opposant les calculs comme causes efficientes aux significations comme causes finales par le dualisme ontologique de la matière physique et du symbole. On traite alors comme une phénoménalité de la computation ce qui n'est que le résultat d'une description éidétique phénoménologique de notre compréhension de la computation, alors que les contenus sémiotiques sont irréductibles aux contenus computationnels, même *via* une sémantique formelle.

Pour sortir d'une telle antinomie, il faut, à la suite de Kant, utiliser une thèse (la première) sur un mode déterminant et constitutif, et l'autre (la seconde) sur un mode réfléchissant et régulateur. Ne pouvant comprendre que les contenus sémiotiques, et ne sachant construire que des contenus computationnels, il nous faut donc renverser notre conception des systèmes dits à base de connaissances. Le cognitivisme veut voir dans les contenus computationnels construits par ces systèmes les causes naturelles de la manifestation des significations par lesquelles nous explicitons le comportement du système, et tombe dans une antinomie qu'il résout par un dualisme ontologique. Il faut plutôt voir dans ces systèmes des constructions informatiques qui, à travers l'utilisation que nous en faisons, vont se voir comprises par nous. Cette compréhension doit guider la construction du système, à titre régulateur, et non être prise comme une phénoménalité. Les significations, indispensables pour comprendre nos systèmes, pour les penser, ne peuvent nous fournir les principes constructifs de ces systèmes. À une conception naturaliste des significations qui en fait la *manifestation* des contenus computationnels se substitue une conception sémiotique et phénoménologique qui en fait *l'interprétation* des représentations formelles. Alors que ces dernières obéissent aux lois informatiques,

les significations sont décrites par les lois de la sémiotique non réductibles aux lois formelles scientifiques en général et informatiques en particulier.

4. Méthodologie de l'IA comme artefacture

Ainsi se dessine pour l'IA une méthodologie fort différente de celle préconisée par le cognitivisme. Puisque l'IA veut utiliser les connaissances phénoménologiques que tout être humain possède en quantité, et qu'elle ne sait faire que de l'informatique, il faut d'une part, dégager la description phénoménologique de ces connaissances en montrant comment elles visent qualitativement le monde comme une objectivité phénoménologique, et d'autre part, étudier comment l'objectivité computationnelle contruite par l'informatique et le fonctionnalisme permet de la simuler, non en la reproduisant, mais en l'approximant. À une étape de modélisation sémiotique ou de modélisation au niveau des connaissances s'ajoute donc une étape de modélisation computationnelle. La première caractérise ce qu'il faut faire, ce qu'il faut comprendre, la seconde comment on peut le faire.

Quand le système exécute le modèle computationnel du modèle sémiotique, l'utilisateur considère le modèle computationnel dans les termes du modèle sémiotique. Il interprète le comportement du système selon ses connaissances, analysées dans le modèle sémiotique. Il peut alors comprendre ce que fait le système, et évaluer s'il résout correctement le problème ou non. En effet, puisque, par hypothèse, l'utilisateur ne sait résoudre le problème qu'en termes de connaissances phénoménologiques, il ne pourra utiliser et évaluer le système que si ce dernier paraît manipuler ces dernières. Mais puisque l'ordre des significations n'est que régulateur, la construction informatique ne peut être validée comme une expérience des sciences de la nature. Il ne peut s'agir que d'une compréhension, d'une interprétation du comportement du programme où l'utilisateur associe aux représentations symboliques formelles du programme, non un contenu computationnel, mais un contenu sémiotique. On ne sait que l'on a correctement modélisé une connaissance que si les contenus computationnels dégagés lors du fonctionnement de la machine peuvent être compris, interprétés

comme des connaissances, comme des contenus sémiotiques. Il faut que l'explication computationnelle serve de support à l'explicitation sémiotique alors que le rapport qui les unit est celui d'interprétation et non de manifestation. Cela revient à une forme modifiée du test de Turing (Turing, 1990) où ce dernier n'est plus un test d'intelligence, mais une évaluation sémiotique de la pertinence du modèle informatique. Peu importe que le système soit idiot ou intelligent : ce qui importe, c'est que je puisse reconnaître dans son fonctionnement les connaissances phénoménologiques qui me permettent de résoudre mon problème.

5. Dualisme revisité

Le naturalisme de principe de l'IA cognitive en fait un dualisme ontologique de la matière physique et du symbole contredisant son monisme de principe. La confusion du cognitivisme provient du fait qu'il veut considérer le niveau des connaissances comme un niveau objectif de description scientifique des systèmes, alors qu'il n'en est que l'interprétation. Pour faire de l'explicitation une explication scientifique, il lui a fallu l'assimiler à une phénoménalité. Le cognitivisme a eu le mérite de reconnaître qu'il n'était pas possible de faire l'économie de l'explicitation. Son erreur fut d'y appliquer une méthode inadéquate. Depuis Dilthey, on oppose en effet *l'expliquer* des sciences exactes au *comprendre* des sciences humaines. Or, le niveau des connaissances, l'explicitation des systèmes, ressortit, avons-nous dit, à la description phénoménologique et sémiotique, où notre *compréhension* du système est analysée. Cette description sémiotique est de l'ordre du comprendre¹⁷ des sciences humaines, et non de l'ordre de

¹⁷ Husserl définit la phénoménologie dans l'ordre du comprendre, car " Elle ne veut pas *expliquer*, au sens psychologique ou psychophysique, la connaissance, l'événement *de fait* dans la nature objective, mais élucider *l'idée* de la connaissance d'après ses éléments constitutifs ou encore d'après ses lois ; ce ne sont pas les relations réelles de coexistence et de succession dans lesquelles sont insérés les actes de connaissance accomplis en fait qu'elle cherche à atteindre, mais ce qu'elle veut *comprendre*, c'est le *sens idéal* des relations *spécifiques* dans lesquelles

l'expliquer des sciences exactes. L'IA, selon cette conception, est donc à la jointure de ces deux domaines de la pensée : elle doit inscrire la construction computationnelle et sa littéralité symbolique dans les registres du sens et de la signification. À un dualisme ontologique, nous substituons un pluralisme épistémologique du comprendre et de l'expliquer qu'il faut faire communiquer à travers les artefacts informatiques: l'IA devient alors une artefacture dans la mesure où elle doit étudier comment un artefact informatique peut se trouver interprété et compris comme un agent ayant des connaissances phénoménologiques décrites par les lois sémiotiques alors qu'il n'obéit qu'aux lois computationnelles. Elle a la charge de trouver en quoi le sens est plurivoque (sens/contenu sémiotique — comprendre — irréductible au sens/contenu formel —expliquer), mais cependant unique: l'artefact est le point de jonction où il procède à la fois du formel et du sémiotique; l'artefact nous révèle comment le formel suscite une compréhension sémiotique en étant davantage une machine donnant à penser qu'une machine qui pense.

Conclusion

Nous avons voulu esquisser dans cet article une argumentation selon laquelle le naturalisme de sciences cognitives mettait l'IA dans une situation impossible. Si l'IA est une science, alors elle se définit comme un cognitivisme computationnel. Ceci la conduit à postuler un dualisme ontologique entre le physique et le symbolique, et à considérer les significations ou contenus sémiotiques comme phénoménalité manifestant une objectivité fondée sur des contenus computationnels, alors qu'elles ne constituent qu'une phénoménologie. Elle aboutit à une antinomie provenant de l'impossibilité théorique de confondre contenus sémiotiques et contenus computationnels. Par ailleurs, si l'IA est une technique informatique, on ne considère alors que les contenus computationnels des représentations et on manque le fait que l'IA veut appréhender le contenu sémiotique de ces représentations. En proposant de considérer les contenus sémiotiques seulement à titre

l'objectivité de la connaissance trouve sa légitimité” in *Recherche Logique I* p.[21].

régulateur et non à titre déterminant, nous pensons résoudre l'antinomie du cognitivisme sans perdre la spécificité de l'IA, à savoir la prise en compte des contenus sémiotiques des représentations symboliques qu'elle manipule. Mais ce passage du constitutif au régulateur implique que l'on abandonne le naturalisme des sciences cognitives pour faire de l'IA une *artefacture*, i.e. une technologie informatique (usage déterminant) à valeur sémiotique (usage régulateur) pour l'utilisateur. Si l'IA veut conserver la technologie informatique et traiter de la connaissance, il lui faut perdre ses rêves démiurgiques de machine qui pense (sans préjuger des autres approches). L'artefacture s'inscrit à la jointure des *Geisteswissenschaften* et des *Naturwissenschaften*, puisque le seul moyen de surmonter leur opposition est de prendre en compte leur irréductibilité réciproque.

Bruno BACHIMONT
INSERM U. 194 - Paris

Bibliographie

- Aïmeur, E. (1993) Acquisition d'objets structurés par caractérisation. *Quatrième journées d'Acquisition des connaissances*, St Raphaël.
- Aussenac, N., Krivine, J.-P. et Sallantin, J. (1992) Éditorial, Numéro spécial Acquisition des Connaissances, *Revue d'Intelligence Artificielle* (6) 1-2.
- Bachimont, B. (1992) *Le contrôle dans les systèmes à base de connaissances, Contribution à l'épistémologie de l'intelligence artificielle*. Paris, Hermès.
- Clancey, W.J. (1989) Viewing Knowledge Bases as Qualitative Models. *IEEE Expert*, Summer 89:9-23.
- Dennett, D. (1990) *La stratégie de l'interprète*. Gallimard. (*The Intentional Stance*, The MIT Press, 1987).
- Deleuze, G. (1963) *La philosophie critique de Kant*. Paris, PUF.
- Dreyfus, H. (1991) Husserl et les sciences cognitives. *Les études philosophiques*, 1:1-30.
- Fodor, J. (1981) Methodological solipsism considered as a research strategy in cognitive psychology. In Haugeland J., ed, *Mind Design*, pp. 307-338, Cambridge (Mass.), The MIT Press.
- Fodor, J., McLaughlin, B. (1990) Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work, *Cognition*, 35:183-204.

- Ganascia, J.G. (1991) L'hypothèse du "Knowledge Level" : théorie et pratique. Rapport LAFORIA 20/91.
- Husserl, E. (1950) *Idées directrices pour une phénoménologie (Ideen I)*. Paris, Gallimard.
- Husserl, E. (1959) *Recherches Logiques*. Paris, PUF, Epiméthée.
- Husserl, E. (1982) *Idées directrices pour une phénoménologie et une philosophie phénoménologique pures : Recherches phénoménologiques pour la constitution (Ideen II)*. Paris, PUF, Epiméthée.
- Kant, E. (1986) *Critique de la faculté de juger*. Paris, Vrin.
- Lycan, W.G. (éd.) (1990) *Mind and Cognition, A Reader*. Cambridge (Mass.), Blackwell.
- Newell, A. (1980) Physical Symbol Systems, *Cognitive Science*, 4:135-183.
- Newell, A. (1982) The Knowledge Level. *Artificial Intelligence*, 18:87-127.
- Newell, A. (1990) *Unified Theories of Cognition*. Cambridge (Mass.), Harvard University Press.
- Petitot, J. (1990) *Le physique, le morphologique, le symbolique ; remarques sur la vision*. Revue de synthèse : IV° S., N° 1-2.
- Petitot, J. (1992) *Physique du sens*. Paris, Editions du CNRS.
- Rastier, F. (1987) *Sémantique interprétative*. Paris, PUF, Formes sémiotiques.
- Richard, J.-F. (1990) *Les activités mentales*. Paris, Armand Colin.
- Pylyshyn, Z. (1984) *Computation and Cognition*, Cambridge (Mass.), The MIT Press.
- Stich, S. (1983) *From Folk Psychology to Cognitive Science, the Case against Beliefs*. Cambridge (Mass.), The MIT Press.
- Turing, A. (1990) Computing Machinery and Intelligence, In M. Boden, éd., *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press.
- Volot, F., Zweigenbaum, P., Bachimont, B., Ben Saïd, M., Bouaud, J., Fieschi, M., Boisvieux, J.-F. (1993) Structuration and Acquisition of Medical Knowledge Using UMLS in the Conceptual Graph Formalism. *Symposium on Computer Applications in Medical Care 93*, à paraître.