

Jean-Michel SALANSKIS

Philosophies et sciences cognitives

Le présent numéro spécial entend marquer l'existence d'un rapport de proximité riche entre les sciences cognitives et la discipline philosophique. Par dessus le marché, il assume l'ambition de montrer que les liens entre les deux domaines ne dépendent pas d'une époque, d'une aire et d'un style de la philosophie, en témoignant de l'intérêt que peut avoir pour les sciences cognitives un dialogue avec la philosophie *continentale*, selon le terme employé aujourd'hui pour l'opposer à la philosophie *analytique*.

Par ce choix, donc, le numéro reflète une position originale et non universelle, qui néanmoins me semble partagée aujourd'hui par un nombre significatif d'acteurs, constitués *de facto* en un courant de réflexion et de recherche particulièrement dynamique. Si la première intention du présent volume est naturellement de faire écho à cette effervescence intellectuelle prometteuse – que ce soit dans ses aspects théoriques ou dans ses aspects méta-théoriques – il a paru également souhaitable qu'un texte d'introduction décrivît de manière synoptique comment les points de vue exprimés ici se situent par rapport à un ensemble de débats et de repères conceptuels caractéristiques de ce qu'a été jusqu'ici – et reste pour une grande part – le lien entre philosophie et sciences cognitives.

Pour évoquer ce contexte, il faut sans nul doute d'abord signaler que le développement philosophique de la problématique cognitive est généralement dominé, notamment dans l'aire anglo-saxonne (qui elle-même est l'aire prépondérante pour les sciences cognitives à tous égards), par quelques discussions exprimées dans le style de la philosophie analytique, ayant le privilège de fasciner préférentiellement les auteurs et leur public.

I. L'état classique du lien

Les carrefours argumentatifs les plus célèbres sont en l'occurrence, me semble-t-il :

— la discussion sur la légitimité du *fonctionnalisme*, ou *représentationnalisme*, dont Putnam peut revendiquer la paternité, mais qui est plutôt rattaché en fin de compte dans la communauté cognitive à l'exposé canonique donné par son élève Fodor dans *The Language of Thought*¹, ou au traité désormais classique de Pylyshyn² ; c'est une discussion sur la valeur de connaissance d'une perspective théorique déterminée sur le “niveau cognitif”.

— la discussion sur la possibilité de restituer tous les effets du sens dans un cadre logico-calculatoire, avec comme points de cristallisation le test de Turing, l'argument de la chambre chinoise de Searle et le théorème de Gödel ; c'est essentiellement une discussion sur l'éventuelle transcendance de la pensée sur le calcul.

— la discussion sur le rapport corps-esprit, et notamment sur l'identité ou la distinction entre les contenus de conscience, notamment les sensations, et les *qualia* neurologiques incarnant ceux-ci ou les énoncés qui les publient (pris avec leur grammaire), discussion relancée et approfondie par maint philosophe analytique (citons par exemple Kripke, Putnam et Rorty).

À beaucoup d'égards, ces trois discussions sont la même, on peut estimer que leur existence autonome correspond simplement à la distinction de trois domaines où une même visée “réductionniste” tout à la fois s'affirme et prête le flanc à la contestation. C'est justement pourquoi il me semble intéressant de les envisager de manière séparée, afin de dégager mieux ce qui fait leur teneur conceptuelle propre, même si dans les écrits effectifs elles sont presque toujours mélangées.

1) La perspective représentationnelle

La discussion sur le représentationnalisme mérite d'être envisagée avant toute autre parce qu'elle absorbe presque intégralement, à elle toute seule, le lien philosophie-sciences cognitives tel qu'il est

¹ Cf. pour une version abrégée Fodor [1981].

² Pylyshyn [1984].

actuellement pratiqué : elle jouit en quelque sorte d'une position comparable à celle de la discussion sur le platonisme en philosophie des mathématiques (soit la discussion sur la légitimité avec laquelle la mathématique moderne peut se présenter comme l'activité de description d'une réalité supra-sensible, immatérielle, notamment d'entités ensemblistes infinitaires).

C'est que le représentationnalisme (ou computationnalisme, ou fonctionnalisme, à ce niveau d'évocation je crois non pertinent de distinguer) reste – à l'instar du platonisme en mathématique pour poursuivre l'analogie – la position de référence, largement dominante, parmi ceux qui pratiquent l'intelligence artificielle aussi bien que parmi ceux qui se situent à l'un des multiples confins du champ cognitif.

Fondamentalement, la thèse fonctionnaliste consiste à poser l'existence d'un niveau “fonctionnel” du cognitif qui recouvre ce qu'on a jusqu'ici pensé comme le “mental”, et que les conceptions behavioristes avaient prétendu évincer, au profit de la pure prise en considération de l'esprit comme une “boîte noire”. On décrira donc le comportement effectif de l'esprit, de la cognition en plaçant dans la boîte noire des *représentations*³ et des règles de manipulation sur ces représentations, qui exprimeront la causalité intra-mentale (mais celle-ci reflète à sa manière la causalité du monde vers le mental et la causalité du mental vers le monde). Les représentations et le mécanisme qui gouverne leurs occurrences dans le champ mental seront supposés “implantés” dans le neurologique comme un software sur un hardware siliceux. En ce sens, les descriptions mentalistes sont donc sauvées sans qu'il soit nécessaire d'avoir recours à une ontologie “transcendante” de l'âme. Le niveau mental est *implanté*, il relève d'une logique et d'un mode de description propres, sans compromettre pour autant la viabilité d'un “matérialisme” pour lequel chacun semble éprouver de grandes craintes.

³ Sur le plan du principe, et c'est un aspect de la discussion, on peut imaginer un fonctionnalisme pour lequel les unités de base de l'intérieur de la boîte noire ne sont pas des représentations, mais des images ou des sous-représentations, ce qui complique le modèle ; nous nous limitons ici à la forme la plus simple, la plus massive de la thèse fonctionnaliste.

On peut faire trois remarques sur la thèse fonctionnaliste :

— elle procure à la démarche cognitive courante, notamment dans la mesure où celle-ci est dynamisée et portée par sa relation à l'expérimentation informatique de l'IA, exactement la légitimation dont elle a besoin ;

— elle fait intervenir la philosophie dans le débat cognitif sur un mode qui réalise la continuité parfaite avec les préoccupations de la philosophie du langage d'obédience analytique, à tel point que celle-ci peut se rebaptiser *philosophie de l'esprit* sans réaménagement notable de sa problématique, sa méthode, son style ;

— elle touche à une question essentielle connue de la tradition continentale, sans en être, semble-t-il, consciente, au moins jusqu'ici et la plupart du temps.

Le premier point ne semble aller de soi. Lorsqu'on s'essaye à écrire des programmes qui accomplissent les comportements passant pour intelligents, on n'a pas d'autre perspective que celle de ramener toutes les propriétés sémantiques des termes à leurs propriétés "causales" d'interaction avec les autres termes semblables, c'est-à-dire en fait de trouver une définition *implicite* du sens des termes par des axiomes qu'ils satisfont ; en fait de "causalité", ce qu'on exhume est alors l'ensemble des liens réguliers et légaux de la syntaxe. La thèse fonctionnaliste dit que la boîte noire mentale est remplie par cela même que l'intelligence artificielle cherche à construire, que le mental a la forme que l'IA cherche à remplir dans chaque cas. Elle justifie donc à la fois les espoirs de cette dernière et l'alignement de l'étude de la cognition humaine (psychologie, linguistique computationnelles) sur ses recherches.

Le second point n'est pas plus problématique. La caractérisation fonctionnelle de la sémantique des représentations présuppose l'*atomicité* constitutive du champ représentationnel, présupposition qui a été explicitée par les premiers maîtres de l'école analytique et qui reste un thème central des travaux qui sont nés au sein du courant. Les études de philosophie du langage qui se sont développées spontanément sous le ciel analytique, indépendamment des sciences cognitives, semblent de prime abord concernées elles aussi par la mise au jour de systèmes de règles restituant la régularité de l'usage des représentations. Dans Fodor [1981], l'auteur évoque le problème de l'intentionnalité tel qu'il se pose à l'hypothèse

fonctionnaliste dans des termes qui sont exactement ceux de la philosophie du langage environnante (problème de l'internalisation logique d'un prédicat de référence, problème des modalités épistémiques et du réseau des énoncés présupposés par les croyances). Il faut d'ailleurs reconnaître que les travaux accomplis dans les années 1955-1980 par des philosophes-logiciens apparentés au courant analytique⁴ sur les logiques modales et leurs modèles ont une résonance, on oserait presque dire une application directe du côté des réseaux sémantiques et autres techniques de représentations des connaissances.

Le troisième point a été souvent mis en avant par Jean Petitot dans toute sa réflexion publiée depuis 1975⁵. Le niveau "fonctionnel" dont parle Fodor n'est pas, nous l'avons déjà indiqué obliquement, un véritable niveau "causal", les liens qu'il exhibe sont ceux d'un système formel, ou d'une "structure" au sens structuraliste. Si ce niveau existe et légitime la recherche cognitive, il ne peut s'agir que d'un niveau *émergent* sur fond d'une réalité dominée par la causalité spatio-temporelle usuelle. D'où la vulnérabilité du fonctionnalisme à certaines critiques philosophiques fondamentales.

Searle, dans Searle [1990], critique par exemple d'un seul mouvement non seulement les thèses fonctionnalistes, mais tout ce qui y ressemble (par exemple, les descriptions générativistes de type chomskien), en tant qu'on y présuppose toujours que le comportement cognitif résulte de l'application inconsciente de règles (elles-mêmes inconscientes), alors que ces concepts sont *contradictoires* : nous ne pouvons appeler de manière conséquente *règle* que ce qui est obéi par un sujet porteur d'intentionnalité. La structure, la régularité formelle indépendante du substrat que l'enquête fonctionnaliste veut mettre en évidence semble ne pouvoir être spécifiée que par l'explicitation des règles que le réel envisagé au niveau fonctionnel paraît obéir. Mais la révélation d'une *organisation* sur ce mode révèle toujours l'organisation comme *fiction*. La cause finale substantivée par le canon des règles n'est là que métaphoriquement, pour répondre aux besoins de l'enquêteur

⁴ Nous pensons à des gens comme Kripke, Montague, Von Wright, etc.

⁵ Cf. par exemple Petitot [1982].

structuraliste ou du cogniticien fonctionnaliste, en particulier les relations que Fodor appelle *causales* en termes desquelles il reconstruit le sens et l'intentionnalité ne sont pas de vraies relations causales, mais des *ersatz* régulatifs renvoyant à un sujet, des relations qui, sur le plan de la seule causalité admise, la spatio-temporelle, la physique, ne se laissent pas construire et attester.

Mais ce dilemme était déjà celui de la troisième critique kantienne, dans sa seconde partie (critique de la faculté de juger téléologique). L'article de Searle est complètement réminiscent de ce texte fameux, et s'achève sur les mêmes conclusions : l'imputation d'organisation, de finalité à telle ou telle portion de la nature reste une façon de parler, une maxime pour la recherche de causalités bon teint, mécanistes-physiques. Il s'en faut de beaucoup, pourtant, que l'universalité et le classicisme de ce problème soient perçus et thématiques au sein du débat classique.

2) *La réduction au logico-calculatoire*

Le deuxième débat-vedette de l'articulation Philosophie-Sciences Cognitives porte au fond sur la réductibilité du comportement intelligent à un algorithme opérant sur des entrées représentationnelles.

Turing, dans un article célèbre⁶, passe pour avoir défini dans les termes suivants l'objectif de l'intelligence artificielle⁷ : il s'agit d'écrire un programme à qui l'on puisse donner en entrée les phrases d'un être humain, et qui sache donner en sortie des phrases s'insérant comme "réponses" dans une conversation, en telle sorte que l'interlocuteur du programme ne perçoive pas de différence avec une conversation humaine ordinaire.

Comme le travail de tout programme effectivement implanté n'excède certainement pas ce dont est capable une machine de Turing, l'objectif semble tenir pour acquise la thèse selon laquelle le concept disons de *fonction récursive* n'explicite pas seulement l'essence du calcul, mais encore celle de la pensée : puisqu'on a du

⁶ Turing [1950].

⁷ Ce qu'il en est réellement de ce que Turing a exprimé, Jean Lassègue essaie, après François Rastier (Rastier [1991], 163-175), de nous l'expliquer dans Lassègue [1993].

mal à imaginer une dimension de la pensée qui ne s'exprimerait pas dans la conversation, une machine subissant avec succès le test de Turing prouverait du même coup la réductibilité de la pensée à l'opération récursive, au calcul.

Une partie de la réflexion philosophique motivée par les Sciences Cognitives consiste à argumenter sur cette réductibilité. L'argument de la chambre chinoise de Searle⁸ prétend prouver que la satisfaction du test de Turing n'est pas un bon critère de réalisation de l'intelligence : il invoque une situation hypothétique où des manipulateurs de symboles satisfont à un test de Turing en chinois sans être sinophones, et conclut que quelque chose comme la *compréhension* ou l'*intentionnalité* peut manquer à un programme qui subirait le test avec succès.

Dans une autre direction, relevant plus de ce qu'on a toujours appelé *rétorsion*, on présente parfois le théorème d'incomplétude de Gödel comme quelque chose qui disqualifie *a priori* la réduction de l'intelligence au calcul : la proposition de Gödel, dit-on, est une proposition vraie dans le modèle standard de l'arithmétique mais qu'une machine de Turing bien équipée ne sait pas atteindre comme la proposition vraie qu'elle est. C'est Penrose qui a le plus récemment enfourché ce cheval de bataille, qu'il attelle avec quelques autres pour proposer ce que d'aucun nomment un paradigme de "l'intelligence artificielle quantique"⁹.

L'ensemble du débat sur la réductibilité de la pensée au calculable, et tout particulièrement sa modalité "gödélienne", sont symptomatiques d'une configuration englobante, et dont il faut réaliser l'importance : celle du rapport privilégié, déterminant quant à leurs interactions possibles, que nouent chacune de leur côté les Sciences Cognitives et la Philosophie avec la Logique.

En effet, la thèse fonctionnaliste, déjà, justifie essentiellement le programme cognitiviste dit des "systèmes experts", qui consiste à intégrer l'information nécessaire aux tâches cognitives dans les systèmes informatiques sous forme de déclarations ou de règles globalement susceptibles d'être ramenées à la forme standard de la logique des prédicats du premier ordre, éventuellement agrémentée

⁸ Cf. Searle [1990].

⁹ Cf. Penrose [1990], pour une version abrégée.

d'opérateurs modaux, et à concevoir le travail intelligent du programme comme une variante de la dérivation ou l'inférence de type logique. La syntaxe logique vulgaire et la démonstration automatique de théorèmes prennent donc la valeur de paradigmes ou de pierres de touche de l'accomplissement du programme cognitiviste. D'où la boutade rapportée par Brian Cantwell Smith, selon laquelle, s'il fallait enseigner à des étudiants très rapidement l'essentiel de la science cognitive moderne, on devrait sans doute choisir de leur exposer le théorème de Herbrand¹⁰.

La rencontre des sciences cognitives avec la philosophie analytique du langage, évoquée plus haut, a lieu elle aussi sous les auspices de ce que Quine appelle “embrigadement” dans Quine [1959] : quelque chose comme une traduction de la langue naturelle dans la logique des prédicats du premier ordre qui vaudrait comme révélation de la subordination profonde de la signification à cette structure. C'est bien évidemment parce que la logique du vingtième siècle est le guide et la référence *aussi bien de la philosophie analytique que des sciences cognitives* qu'elle constitue le médium de leur interaction la plus aisée.

La problématique philosophique de la légitimité de la réduction au calcul ne fait pas exception à cette sujétion à la logique : tout le débat s'organise autour de la détermination technique de la calculabilité par les travaux de Turing, Church et autres, le théorème de Gödel pouvant être vu comme un effet ou un reflet de ces travaux sur et dans la méta-mathématique. Notamment il faut bien voir que même – à vrai dire, *surtout* – dans la mesure où le débat est un débat sur la *limite* de la logique, il est complètement capturé par le style et les attendus de la discipline logique.

Bien entendu, tout à l'autre bout, lorsqu'il s'agit du *design* des machines ou de l'informatique théorique (de la théorie des automates, de la compilation, des langages, des grammaires), le rôle technique de la logique est également absolument central, c'est en fait “naturel” pour le coup.

On ne comprendra pas correctement l'articulation philosophie-sciences cognitives et ses possibilités actuelles de divergence et d'évolution si l'on ne prend pas d'abord la mesure d'un événement

¹⁰ Cf. Smith B.C.[1991].

épistémique plus général et déterminant à tous égards, qui est le triomphe disciplinaire de la logique au cours du siècle.

3) *Le problème corps/esprit*

J'ai signalé une troisième discussion : celle qui porte sur le rapport corps-esprit. Il convient de remarquer que le représentationnalisme, auquel j'ai d'abord fait un sort, se présente lui-même comme une prise de position dans ce troisième débat. Dire qu'il y a un niveau "structuré comme un langage" qui possède sa causalité structurale est une manière d'assigner l'identité du *mental*, ou encore d'envisager l'esprit comme une strate fonctionnelle "au-dessus" du niveau neuro-physiologique, qui relèverait du corps. J'ai néanmoins traité à part la question du représentationnalisme parce que, me semble-t-il, la thèse représentationnaliste se constitue – peut-être dans l'après-coup – essentiellement comme *fondation épistémologique des sciences cognitives*, et qu'il y a donc un débat autonome suscitée par elle, sur elle, qui n'est plus débat sur l'esprit, mais sur le caractère méthodologiquement acceptable du recours à la causalité structurale, débat qui concerne chaque discipline cognitive de manière très immédiate.

En revanche, ce que j'appelle débat sur le problème esprit-corps est le débat proprement philosophique – qui peut se revendiquer d'ailleurs comme la poursuite de la réflexion cartésienne – sur la pertinence ontologique de cette distinction. Ce débat est abordé, même dans l'aire anglo-saxonne, par une voie qu'on peut qualifier de "phénoménologique" au sens large, puisqu'aussi bien la question la plus souvent posée est celle de savoir ce qu'il faut faire des contenus de sensation singuliers en tant que "vécus" par un sujet : sont-ils des réalités immanentes plus certaines que toute autre réalité parce qu'elles sont directement saisies avec évidence par la conscience, et procurent un soubassement irremplaçable à tout un versant de l'activité de l'esprit, ou bien pouvons nous les considérer comme des artefacts éliminables dans un contexte philosophique résolument empiriste? Là encore, la discussion anglo-saxonne, le plus souvent, aura recours à l'instrument logique pour développer le problème : on cherchera à analyser le régime logique – notamment modal – des énoncés véhiculant la sensation, dans son rapport avec les autres énoncés, à visée externe, mondaine, ou dont le sens est

purement du côté de la cohérence. Kripke, Putnam, Rorty, à la suite de Wittgenstein, qu'on peut considérer comme le père de cette attitude, proposent ainsi des conclusions de type ontologique à partir d'un tel examen logico-épistémique de contenus relevant du discours "phénoménologique"¹¹.

L'évocation des trois discussions majeures du champ confirme donc la domination de la perspective logique.

II. Le lien à l'heure post-cognitiviste

Cela dit, ce triomphe, aujourd'hui, semble menacé par un ébranlement d'importance comparable à celui qui, dans presque toutes les sciences, a porté le canon logique et la formalisation au poste de commandement au début du siècle. De même qu'en mathématiques et en physique, de nombreux esprits redécouvrent le caractère fondateur, originaire de l'élément *géométrique*, en tant que ce dernier ne se laisse pas réduire à tel ou tel régime syntaxique – même s'il est localement assumé par le moyen de cette sorte de régime –, de même que la question du *continu mathématique* est abordée à nouveaux frais (que ce soit dans un cadre constructiviste, un cadre gödélien ou un cadre infinitésimaliste), de même que la physique mathématique constate l'importance des modèles continus dans ses théories et la développe (que ce soit en physique quantique ou en physique des phénomènes critiques), de même les sciences cognitives assistent depuis quelques années au développement d'une *modélisation connexionniste*, essayant de se situer sur le même terrain continuiste que la physique mathématique [Grossberg, le groupe PDP, Amit], et d'une *linguistique cognitive*, insistant sur la fonction centrale de l'espace et des schèmes spatio-temporels dans la sémantique [Lakoff, Langacker, Talmy]¹².

Nécessairement, cette mutation a des retentissements sur la vie du lien philosophie-sciences cognitives. Je l'ai dit, la présente livraison de la revue INTELLECTICA voudrait témoigner, justement, de l'ampleur et de la diversité de ces retentissements. Avant de laisser la parole aux textes qui parleront mieux que cette brève synthèse,

¹¹ Cf. Kripke [1982], Putnam [1984], Rorty [1979].

¹² Pour une présentation de la mutation en cours, cf. Andler [1989].

essayons simplement de classer sous quelques rubriques significatives ce qui s'annonce aujourd'hui comme les différents aspects du renouveau du lien philosophie-sciences cognitives :

1) *La redécouverte de la Gestalt*

Une partie du lien consiste dans la redécouverte de textes pré-cognitifs, dans leur majeure partie disciplinairement apparentés à la philosophie, qui ont développé une conception non logiciste, non cognitiviste de l'activité de l'esprit. Le courant gestaltiste, qui s'est développé en Allemagne, en Autriche et en Italie à la fin du siècle précédent et dans la première partie de celui-ci, est la grande référence à laquelle beaucoup de ceux qui réfléchissent sur les sciences cognitives reviennent ainsi aujourd'hui. Qu'il s'agisse de textes de psychologie, volontiers centrés sur le problème de la perception, ou de textes théoriques, développant le concept de *Gestalt* pour lui-même et tentant d'en préciser la portée, les écrits de la *Gestalt* anticipent souvent les conclusions et les préoccupations de la linguistique cognitive ou du connexionnisme¹³. Après qu'Ehrenfels eut donné le coup d'envoi conceptuel de la théorie de la *Gestalt*, avec son article-essai de 1890 "Über 'Gestaltqualitäten'", l'école allemande, brillamment illustrée par les noms de Köhler, Wertheimer et Koffka, a conduit toute une série d'expérimentations psychologiques passionnantes – notamment sur la perception de l'espace – et formulé des hypothèses sur le type de fonctionnement cérébral impliqué dans la cognition dont l'audace et la modernité nous impressionnent aujourd'hui¹⁴.

Le séminaire de philosophie austro-allemande, au plus fort de la domination logiciste, a étudié ce passé centre-européen, mettant en avant notamment le rôle fondateur de Brentano et la crucialité théorique des concepts dégagés par Husserl dans sa troisième recherche logique. Il a aussi montré comment la prise au sérieux de

¹³ L'article de Casati et Smith, dans ce volume, montre dans quelle mesure la *Gestalttheorie* et une certaine variété de la phénoménologie ont anticipé la problématique de la *physique naïve* actuelle.

¹⁴ Pour plus d'informations sur cette histoire, on peut se reporter aux travaux du séminaire de philosophie austro-allemande (Smith B. [1982,1988]) et à l'article de l'*Encyclopædia Universalis* (Thinès [1985]).

toute cette tradition devait reconduire la philosophie de l'esprit à certains thèmes aristotéliens.

Un auteur, en France, a étudié la *Gestalt* avec passion, allant jusqu'à ordonner son discours philosophique autour de l'inspiration qu'il en a reçue : je veux parler de Merleau-Ponty. Il est saisissant de voir à quel point le livre *Phénoménologie de la Perception*¹⁵ est installé dans une problématique cognitive, et de constater que ses thèses principales devancent dans une large mesure la littérature anti-logicienne actuelle. Mais on observe aussi que pour quelqu'un comme Merleau-Ponty, l'insistance philosophique sur le niveau pré-logique de la perception, qu'il impute au corps et comprend comme une donation de sens *antéprédicative*, est liée au thème heideggerien de l'*Être-au-monde*. L'étude fine des modalités originellement créatrices (à valeur *constituante*) de la perception, a lieu chez lui, et semble-t-il de manière *nécessaire*, sous la direction de la notion d'*Être-au-Monde*, qui qualifie, on le sait, l'engagement du sujet dans son environnement mondain à un niveau en quelque sorte "métaphysique" chez Heidegger.

2) *La valeur critique et inspiratrice de la notion d'Être-au-Monde*

Or ce lien, entre l'analyse de la riche structure antéprédicative de la perception (site de l'émergence des configurations de type gestaltique notamment), et le concept d'*Être-au-monde*, ne se noue pas seulement chez Merleau-Ponty, dans la préhistoire des sciences cognitives. Il est aussi attesté de nos jours dans toute une littérature de contestation du cognitivisme, antérieure au développement du connexionnisme et de la linguistique cognitive, et à laquelle les événements actuels donnent donc singulièrement raison : nous voulons parler des réflexions "phénoménologiques" de Dreyfus, Winograd et Florès, ou Maturana et Varela.

Le livre de H. Dreyfus *Intelligence Artificielle : mythes et limites*¹⁶, a été le premier à présenter une argumentation philosophique de fond contre le programme cognitiviste dans la

¹⁵ Merleau-Ponty [1945].

¹⁶ Dreyfus [1979].

ligne du concept d'Être-au-Monde. Cette argumentation consiste essentiellement à donner un sens heideggerien à la thèse wittgensteinienne sur l'opacité ultime de la règle : nantir un logiciel de toutes les règles, accompagnées de l'énumération de leurs contextes d'application, information que requiert l'activité de l'esprit, serait *a priori* impossible à cause de l'ambiguïté *en dernière instance* de la règle, telle que la prouve Wittgenstein ; et la faculté que possède l'homme de se passer de l'explicitation achevée des règles et de leur contexte d'application proviendrait de son *Être-au-monde* au sens heideggerien. Une telle analyse ne va pas de soi, et ne peut être présentée que par le truchement de tout un travail philosophique : c'est un tel travail que Dreyfus fut le premier à présenter. Winograd et Florès ont repris cette analyse à leur compte, dans leur langage propre, et en mettant l'accent sur la signification *herméneutique* de l'Être-au-monde¹⁷. Ils ont aussi montré en quoi le travail de Maturana et de Varela¹⁸ se situait dans la même perspective. F. Varela lui-même, notamment dans sa plus récente publication¹⁹, s'explique sur le lien des idées qu'il a promues avec celles de la tradition phénoménologique.

Le nouvel intérêt pour le thème herméneutique dans les matières cognitives ne se limite d'ailleurs pas aux fondements de l'IA : un travail comme celui de François Rastier, approfondissant dans le domaine de la sémantique linguistique le *paradigme différentiel* hérité du structuralisme, rencontre aussi les concepts et les problèmes de l'herméneutique²⁰.

L'évolution actuelle est que l'invocation de références phénoménologiques persiste à être de mise dans la discussion philosophique sur l'IA, ses fondements et ses possibilités, mais en même temps, me semble-t-il, de très nombreux spécialistes du champ cognitif ont compris que l'Être-au-monde était bel et bien le nom d'une propriété ou modalité fondamentale de la cognition, et s'interrogent de façon technique sur la meilleure manière de la faire intervenir dans leur schéma d'explication scientifique ou dans leur

¹⁷ Cf. Winograd-Florès [1986].

¹⁸ Cf. Varela [1986].

¹⁹ Varela [1993].

²⁰ Cf. Rastier [1987].

projet de simulation (au niveau de l'ontogenèse neurale, au niveau du processus connexionniste de la vision, au niveau du régime temporel de la vie psychologique, ou au niveau de la structuration langagière de la signification, pour énumérer rapidement les principales possibilités) : la discussion née spontanément au sein de la connexion sciences cognitives-philosophie sur le plan critico-fondationnel prend donc, de manière tour à fait imprévue, un tour "positif" et opératoire²¹. Dans la mesure où la critique du cognitivisme au nom de l'Être-au-monde est politiquement en accord avec la mutation continuiste des sciences cognitives – puisqu'elle partage avec cette dernière l'ennemi cognitiviste-symboliste – les récupérations technico-scientifiques de la problématique de l'Être-au-monde sont souvent incorporées à tel ou tel article, telle ou telle étude qui s'inscrit aussi dans le combat pour ladite mutation²². On peut néanmoins se demander si la critique du cognitivisme par l'Être au monde et sa critique par le continu sont dans le fond des choses congruentes et compatibles : c'est une des interrogations que j'aimerais voir prises en charge aujourd'hui.

3) La relance de l'épistémologie autour de la question du continu

Il faut, je crois, souligner un troisième aspect des discussions actuelles : le connexionnisme et la linguistique cognitive apportent avec eux de nouveaux modèles. Dès lors que la jeune science cognitive a désormais connu, ou semble en passe de connaître au moins une révolution méthodologique, elle appelle naturellement une mobilisation plus réelle et plus profonde de l'intelligence épistémologique. Si, à l'époque précédente, le discours sur les fondements des sciences cognitives a pu sembler parfois plutôt une des voix par laquelle parlait un préjugé logiciste ambiant qu'une réflexion, ou une évaluation critique, aujourd'hui, dès lors qu'il y a disponibilité de plus d'un paradigme, une véritable discussion philosophique s'installe. Les articles phares de Smolensky et Fodor-

²¹ Cf. aussi la contribution de T. Winograd dans ce volume.

²² Dreyfus déjà, lorsqu'il critique dans Dreyfus [1979] les postulats biologiques et psychologiques de l'IA symbolique, rencontre le courant "continuiste".

Pylyshyn²³ en portent témoignage. Mais au-delà, comme l'alternative se situe entre des modèles fortement hétérogènes, et qu'elle réveille la question du continu, vieille comme la science et chargée de métaphysique, on peut s'attendre à ce qu'une épistémologie proprement philosophique se saisisse des problèmes que posent les nouveaux modèles et leur conflit avec les anciens.

Sur le plan du strict suivi épistémologique, cette nouvelle réflexion sur les fondements théoriques des sciences cognitives me semble principalement confrontée avec un problème d'*unification*. C'est qu'il y a en quelque sorte trois familles de modèles continuistes dans le champ cognitif :

— les modèles catastrophistes développés à partir des idées de Thom au cours des années 1960-70, et appliqués alors à la psychologie ou la linguistique dans un esprit qui était déjà celui des sciences cognitives bien qu'on l'eût par force ignoré ;

— les modèles connexionnistes qui, malgré la préhistoire et l'histoire secrète qu'on peut revendiquer pour eux, ne sont réellement apparus sur le devant de la scène que dans les années 80 ;

— les “modèles” ou plutôt “schèmes” de la nouvelle linguistique cognitive, dont l'émergence est toute récente.

C'est déjà assumer une responsabilité philosophique que d'essayer de dire le comment et la limite éventuelle du recouvrement mutuel de ces langages ou ces approches. En particulier, ce travail ne peut être séparé d'un travail d'éclaircissement épistémologique concernant les rapports qu'entretiennent les formes ou niveaux organisationnels (de la perception, du langage profond, de l'activité neurale-psychique) visés par ces modélisations.

Par ailleurs, une des tâches les plus évidentes est d'évaluer les apports de chacun de ces continuismes pour ce qui regarde les questions centrales de l'objectivation en sciences cognitives : question du représentationnalisme, de l'essence de la règle ou de l'organisation, de la pertinence et la portée du concept d'émergence. Un tel travail, d'après l'approche partielle de la littérature qui est la mienne, semble bien engagé de part et d'autre de l'atlantique.

Par rapport aux enjeux variés qui viennent d'être nommés, la simple honnêteté commande de dire que le travail de Jean Petitot

²³ Smolensky [1988], Fodor-Pylyshyn [1988].

nous donne l'exemple, depuis de longues années²⁴, et à nouveau dans ce numéro.

Je conclurai en ajoutant que les élucidations épistémologiques évoquées à l'instant appellent également une réflexion plus générale, avec un recul supplémentaire vis-à-vis des modèles, sur le type de légitimité qu'on peut attribuer à ces recours au continu : le continu intervient-il dans ces matières sur un mode analogue à celui qui prévaut en physique, c'est-à-dire en tant qu'*interprétant mathématique de formes d'intuition* (l'espace et le temps)? Y-a-t-il un *continu sémantique a priori* indépendant du continu spatial externe classique ? Et mille autres questions délicates de la même eau.

Jean-Michel SALANSKIS
CNRS - Paris

Bibliographie

- ANDLER D. [1989]—Cognitives (Sciences) in *Encyclopædia Universalis*, Paris : Encyclopædia Britannica 1989, t. VI, p.65-74.
- DREYFUS H. [1979]—*Intelligence Artificielle, Mythes et Limites*, Paris : 1984 Flammarion.
- EHRENFELS C. VON [1890]—Ueber 'Gestaltenqualitäten', *Vierteljahrsschrift für Wissenschaftliche Philosophie*, **14**, 1890, 249-92. Trad. anglaise On Gestalt qualities, in *Foundations of Gestalttheorie*, München : 1988 Philosophia Verlag.
- FODOR J. & PYLYSHIN Z. [1988]—Connectionism and Cognitive architecture : A critical analysis, *Cognition*, **28**, 1/2, 3-71.
- FODOR J. [1981]—Le corps et l'esprit, *Pour la Science Mars* 1981 n°**43**.
- KRIPKE S. [1982]—*Wittgenstein : on rules and private languages*, Oxford : Blackwell.
- LASSEGUE J. [1993]—Le test de Turing et l'énigme de la différence des sexes, in D. Anzieu éd. *Les contenants de pensée*, Paris, Dunod, à paraître.
- MERLEAU-PONTY M. [1945]—*Phénoménologie de la perception*, Paris, Editions Gallimard.
- PENROSE R. [1990]—Précis of The Emperor's New Mind: Concerning computers, minds, and the laws of physics, *Behavioral and Brain Sciences*, **13**, 643-705.
- PETITOT J. [1985]—*Morphogénèses du sens*, Paris : 1985 PUF.

²⁴ Cf. Petitot [1985,1989a,1989b,1991,1992].

- PETITOT J. [1989a]—Forme in *Encyclopædia Universalis*, Paris : Encyclopædia Britannica 1989, t. XI, p.712-728.
- PETITOT J. [1989b]—Hypothèse localiste, Modèles morphodynamiques et Théories cognitives : Remarques sur une note de 1975, *Semiotica*, **77**, 1/3, 65-119.
- PETITOT J. [1991]—Syntaxe topologique et grammaire cognitive, *Langages*, n°?, 97-128.
- PETITOT J. [1992]—*Physique du sens*, paris, Éditions du CNRS.
- PUTNAM H. [1981]—*Raison, Vérité, et Histoire*, Paris : 1984 Minuit.
- PYLYSHYN Z. [1984]—*Computation and cognition*, Cambridge, Masschussets, London, England : MIT Press.
- QUINE W.O. [1959]—*Le mot et la chose*, Paris : 1977 Flammarion.
- RASTIER F. [1987]—*Sémantique interprétative*, Paris : PUF.
- RASTIER F. [1991]—*Sémantique et recherches cognitives*, Paris : PUF.
- RORTY R. [1979]—*L'Homme spéculaire*, Paris, 1990, Editions du Seuil.
- SEARLE J.R. [1990]—Consciousness, explanatory inversion, and cognitive science, *Behavioral and Brain Sciences*, 585-596.
- SMITH B. ED. [1982]—*Parts and Moments*, München : Philosophia Verlag.
- SMITH B. ED. [1988]—*Foundations of Gestalt Theory*, München : Philosophia Verlag.
- SMITH B.C. [1991]—The owl and the electric encyclopædia, *Artificial Intelligence* **47**, 251-288.
- SMOLENSKY P. [1988]—On the proper treatment of connectionism, in *The Behavioral and Brain Sciences*, **11**, p.1-23.
- THINES G. [1985]—Gestaltisme, in *Encyclopædia Universalis*, édition 1985, vol. 5, 560-565.
- TURING A. M. [1950]—Computing Machinery and Intelligence, *Mind*, Vol. LIX, n° **236**.
- VARELA F. [1986]—Experimental Epistemology : Background and Future, *Cahier du C.R.E.A.* n°**9**, 107-121.
- VARELA F. [1993]—*L'inscription corporelle de l'esprit*, Paris, Le Seuil.
- WINOGRAD T. & FLORES F. [1986]—*Understanding Computers and Cognition: A New Foundation for Design*, Ablex, Norwood, NJ.