

INTRODUCTION

COGNITIVISME

ET SCIENCES COGNITIVES

Pierre STEINER*
psteiner@up.univ-mrs.fr

Dans les limites d'une introduction, j'expose les différentes assumptions (naturalisme, fonctionnalisme) et les principaux axiomes (représentationnalisme, computationnalisme) du paradigme cognitiviste. J'évoque également les critiques d'obédience neuroscientifique de ce programme, avant de présenter les théories qui critiquent son oubli des propriétés conscientes, somatiques, temporelles, environnementales, dynamiques et sociales de la cognition.

« On parle beaucoup de sciences cognitives ces derniers temps. Il paraît cependant nécessaire de distinguer l'édification d'une théorie scientifique ou la croissance d'une discipline scientifique avec des questions bien définies de la prolifération d'idées pour des théories possibles que personne ne sait, même en principe, comment réaliser. »

Hilary PUTNAM¹

I. Introduction

Le cognitivisme est le paradigme principal des sciences cognitives. Il fut le premier, et n'est que depuis peu soumis à un paradigme concurrent (connexionnisme) et à des critiques systématiques, entre autres

* Allocataire-moniteur au Département de philosophie de l'université de Provence. Affilié au laboratoire Ceperc, il prépare une thèse sous la direction de J.-P. Cometti consacrée à une analyse des dimensions d'*Arrière-Plan* de l'esprit, notamment à partir d'une étude des niveaux *explicites*, *implicites* et *tacites* de représentation (cette dernière étant considérée dans une optique interprétationniste) et d'attitudes cognitives. Cette analyse s'effectue à la lumière des travaux de philosophes pragmatistes (Peirce, James, Dewey, Brandom) dont il tente d'intégrer les réflexions dans certains débats actuels en philosophie de l'esprit et des sciences cognitives.

1. *The Threefold Cord: Mind, Body, and World*, New York, Columbia University Press, 1999, p. 35. Ma traduction.

inspirées par la phénoménologie ou motivées par les neurosciences. Avant de présenter le cognitivisme, il s'agit de définir ce que sont les sciences cognitives. Pour reprendre une expression de J. Petitot, on peut considérer les sciences cognitives comme une alliance de disciplines visant à constituer une *science naturelle de l'esprit*. Par « esprit », on entend simplement ici l'ensemble des capacités mentales propres au système nerveux central. Plus précisément, les sciences cognitives sont des *sciences de la cognition* : il s'agit d'étudier les capacités et processus mentaux – mais néanmoins naturels – qui, au moyen d'un traitement (aussi bien sélectif que productif) de l'information, engendrent, transmettent, modifient, utilisent, conservent ou consistent en de la connaissance : sensori-motricité, perception, mémoire, compréhension et production langagière (et plus largement symbolique), représentation des connaissances, ou encore raisonnement. La cognition est à la connaissance ce que la volition est à la volonté : une fonction de production et de réalisation. C'est cette fonction qui amène l'*état* de connaissance (cette fonction est présente chez un adulte, un bébé, ou un chien ; et aussi, de manière plus problématique, chez une huître, un ordinateur ou un thermostat). On étudie donc ce qui rend causalement possible la connaissance, et non pas cette dernière en elle-même, dotée de propriétés – différentielles – culturelles ou normatives.

La période de naissance des sciences cognitives se confond avec celle de l'émergence du paradigme cognitiviste : les années 1950-1960. Durant cette période, plusieurs programmes de recherche se développent : la psycholinguistique, la cybernétique et les sciences de l'informatique et de l'information, qui peu à peu ont convergé vers le même objet d'étude, l'esprit, considéré comme système cognitif. Les travaux de Chomsky en psycholinguistique ont constitué l'un des socles conceptuels du cognitivisme. Dans sa critique dirimante de la théorie behavioriste du langage proposée par Skinner (1957), Chomsky souhaitait montrer que, pour expliquer l'apprentissage et le comportement langagier d'un être humain, nous ne pouvions nous limiter à élaborer un schéma *stimulus-réponse*. Il nous faut postuler l'existence d'une vie mentale interne afin d'expliquer les transitions entre les entrées et les sorties du système (ou même, plus élémentairement, l'interprétation des entrées), ce que le béhaviorisme refusait catégoriquement. Cognitivement, nous ne sommes pas des êtres-réflexes. Nos comportements sont « *stimulus free* » : ce que font les gens dépend de ce qu'ils

croient, désirent, et *de la manière* dont ils interprètent les paroles, les objets ou encore l'environnement. On ne peut pas uniquement comprendre le comportement humain à partir des propriétés objectives du monde ; il faut aussi prendre en compte la manière dont le monde est perçu et représenté dans l'esprit : nous sommes en effet sensibles à l'information que contient un stimulus, et non pas au simple format physique de celui-ci². Il est nécessaire de postuler des représentations comme intermédiaires entre l'*input* sensoriel et l'*output* comportemental. À la base, le cognitivisme est donc antiréductionniste, proposant de laisser une place à l'interprétation, aux représentations internes du système dans une explication comportementale. Mais, afin d'éviter la postulation d'un « fantôme dans la machine » (Ryle), évanescents et incertains, les cognitivistes veulent montrer qu'il est possible de faire de ces représentations internes un *système* fermé qu'il est possible d'étudier *scientifiquement*, en formulant les lois de son fonctionnement et de ses productions. « Un réductionnisme en chasse un autre », diront certains...

Il faut donc considérer l'esprit humain comme étant un *système de traitement de l'information* ; les êtres vivants sont, pour reprendre l'expression de G. Miller, *informavores*. De là découle également l'idée qu'un système cognitif est un système d'inférence (étant donné qu'il *traite* de l'information et la *transforme* d'*input* à *output* suivant des *règles* – voir *infra*). À cette conception de l'esprit se conjugue l'idée que l'on pourrait davantage comprendre le fonctionnement de ce système inférentiel en le comparant avec l'ordinateur, qui est lui aussi un système de traitement de l'information (mais sont-ils de la même espèce ? Là est toute la question, qui engendre évidemment bon nombre de critiques).

Les sciences cognitives, dès leur début, ont rejeté le béhaviorisme (ou du moins une version poussée de ce dernier³), et, tout en refusant

2. Voir Zenon Pylyshyn, *Computation and Cognition. Toward a Foundation for Cognitive Science*, Cambridge (Mass.)/London, MIT Press, 1984, 320 p. et Kim Sterelny, *The Representational Theory of Mind*, Oxford, Blackwell, 1990, 268 p.

3. Pour certains, en effet, en mettant tout à fait de côté la question de la *conscience* ou du *vécu*, les sciences cognitives – à tout le moins dans leur orientation cognitiviste – ne sont qu'une forme de behaviorisme, et sont donc principalement erronées car gravement incomplètes. Pour ce genre d'argument, voir entre autres les travaux de Galen Strawson, John Searle, et Francisco Varela (voir références de la bibliographie).

l'introspectionnisme (déclaré peu objectif) ont adopté un mentalisme centré autour de l'« affirmation de l'importance de processus strictement internes, dotés de réalité et d'autonomie⁴ ». Ce qu'il faut souligner, c'est qu'elles rejettent aussi tout réductionnisme : nos états mentaux ne sont pas que des phénomènes bio-chimico-physiques. En tant que phénomènes cognitifs, nos états mentaux (que l'on suppose facilement identifiables et localisables) représentent⁵ quelque chose ; ils véhiculent une ou des informations portant sur le monde extérieur. Cette information peut être par exemple codée sous une forme linguistique (digitale) ou picturale (codage analogique). Les transformations qui affectent les états cérébraux ne peuvent donc être uniquement décrites en termes physico-chimiques, car ces transformations sont aussi des calculs sur les représentations/informations véhiculées par ces états. Du point de vue cognitiviste, le mental possède donc une nature matérielle (le cognitivisme refuse tout dualisme « ontologique »), mais également une autonomie conceptuelle ou logique ; cette autonomie permettrait alors de l'étudier à un niveau de description portant uniquement sur la fonction des entités mentales⁶. Il faut donc (relativement) distinguer une investigation sur l'architecture du cerveau d'une enquête sur l'architecture des fonctions cognitives, cette dernière enquête pouvant alors prendre l'ordinateur comme *modèle* pour expliquer l'économie de notre système cognitif. L'ordinateur peut *simuler* le fonctionnement de notre vie mentale. Mais « simuler » peut aussi bien signifier « faire comme » que « remplacer ». Si l'on s'en tient au premier sens, il est clair que les modèles *computationnels* présentent une valeur heuristique certaine pour comprendre le fonctionnement de l'esprit.

L'esprit tel qu'il est conçu par le paradigme cognitiviste est assez particulier, et correspond peu aux conceptions du sens commun. Il est par exemple essentiellement inconscient ou encore, selon l'expression

4. Daniel Andler, « Cognitives (Sciences) », dans *Encyclopaedia Universalis*, tome 6, 1992, p. 68.

5. Précisons ici qu'il existe un grand nombre de conceptions de ce qu'est ou de ce que ne peut être une représentation mentale – ou plutôt de ce que représente une telle entité, et de ses relations avec la notion d'*information*. Pour certains, il n'y a pas de représentations (Varela, Dreyfus) ; pour d'autres, elles ne représentent rien (elles n'ont pas de contenu sémantique intrinsèque) ; elles sont des « façons de parler » utiles, des heuristiques (Chomsky, Jackendoff, Stich, Dennett, Cummins, Haugeland) ; pour d'autres encore, elles existent vraiment et possèdent intrinsèquement une valeur sémantique (Fodor). Cette question est l'une des plus débattues en philosophie de l'esprit et des sciences cognitives.

6. Daniel Andler, « Processus cognitifs », dans D. Andler, A. Fagot-Largeault et B. Saint-Sernin, *Philosophie des sciences I*, Paris, Gallimard/Folio (essais), 2002, p. 267.

de Dennett, *subpersonnel* : les opérations cognitives qui prennent place en nous sont inconscientes, et sont effectuées par des « modules » de notre esprit auxquels nous ne pouvons accéder comme nous pouvons accéder à certains de nos souvenirs ou connaissances. La thèse de la modularité est importante : elle consiste à dire qu'une tâche cognitive se réalise principalement à partir de plusieurs sous-systèmes *fonctionnellement* distincts et indépendants (« cloisonnés ») : ainsi, la perception linguistique serait par exemple rendue possible, entre autres, par les modules (anciennement « facultés ») de traitement de l'information visuelle ou auditive, de connaissance phonologique et de connaissance syntaxique. Certains défendent aussi aujourd'hui une thèse de spécialisation par domaine : il y aurait ainsi un module non pas seulement pour chaque mécanisme (sous-modules), mais aussi pour chaque tâche cognitive (langage, reconnaissance de visage, empathie...) ; certains modules seraient d'ailleurs innés (nativisme). Pour certains (Fodor), il n'y a que les opérations modulaires qui puissent faire l'objet d'une science « exacte » ; les processus centraux (inférences, pensées conscientes) seraient en effet trop holistes pour cela.

La pensée qui prend place en nous n'est donc pas le résultat d'une activité *personnelle*, rationnelle et consciente. On explique le fonctionnement de l'esprit en se référant à des états et processus mécaniques ou causaux qui n'ont aucun *équivalent* avec des états personnels, même lorsque ces derniers sont inconscients. Bien que rejetant le béhaviorisme et le physicalisme fort, le cognitivisme présuppose que l'esprit et ses capacités (langage, perception,...) est *interne, dedans*, c'est-à-dire qu'il s'arrête là où commence le monde extérieur, biologique ou culturel. Ce monde existe, mais il est actif sur nous en étant médiatisé par les représentations que nous nous en faisons. D'où par exemple l'idée d'étudier un phénomène culturel à partir des représentations que les gens peuvent s'en faire (anthropologie cognitive).

Les disciplines constituant aujourd'hui les sciences cognitives sont, principalement, la psychologie cognitive, la philosophie, la linguistique, l'intelligence artificielle et les neurosciences (neuro-anatomie, neurophysiologie, neurobiologie,...). La logique, les mathématiques, l'informatique et la physique ont aussi un rôle à jouer. La cohabitation entre ces disciplines n'est pas aisée : chaque discipline conçoit l'esprit d'une certaine manière et à un certain niveau (du neurone au « système central »), et aura parfois tendance à considérer comme non pertinentes

ou non fondées les théories développées par d'autres disciplines. On peut dire qu'une sorte de division (non étanche) du travail s'est tacitement imposée: la psychologie, la linguistique, la philosophie et la logique tentent plutôt de caractériser les aptitudes cognitives humaines à un niveau fonctionnel, indépendamment de leur réalisation cérébrale. Ces disciplines s'intéressent donc en particulier aux processus cognitifs dits supérieurs, dont l'inférence est l'armature, ainsi qu'à leurs résultats: la connaissance, le savoir⁷. Le pôle des neurosciences, des mathématiques et de la physique privilégie les processus de traitement de l'information dits « inférieurs », en particulier la perception et la motricité, où intervient d'ailleurs un *savoir faire* partagé avec les animaux. Les informaticiens et les spécialistes d'intelligence artificielle s'occupent plutôt des processus de simulation des fonctions cognitives, qui n'ont pas, nous le verrons, un rôle négligeable à jouer. Évidemment, un travail interactif est plus que nécessaire à propos de certains phénomènes et capacités.

II. Caractéristiques du cognitivisme : représentationnalisme et computationnalisme

Le cognitivisme, avons-nous dit, fut le paradigme fondateur et encore dominant des sciences cognitives. Jusqu'à peu, l'une de ses forces résidait dans le fait qu'il était *the only game in town*, pour reprendre l'expression de l'un de ses plus ardents défenseurs, Jerry Fodor.

Avant de détailler ses deux hypothèses fondamentales (représentationnalisme et computationnalisme), évoquons rapidement les deux intuitions méthodologiques et philosophiques sur lesquelles se base le cognitivisme: le *naturalisme* et le *fonctionnalisme*. L'intuition naturaliste, plus largement à la base de l'idée de science cognitive, est difficile à circonscrire. Bornons-nous ici à remarquer que le naturalisme est avant tout une position *méthodologique* (ou encore *épistémologique*) pour laquelle l'appareil explicatif des sciences naturelles (physique, biologie et, dans une moindre mesure, chimie) *doit* constituer, en dernière instance, le cadre conceptuel auquel toute théorie (en sciences naturelles évidemment, mais aussi et surtout en sciences humaines et sociales) doit se réduire, ou, en tout cas (*naturalisme modéré*) se

7. Daniel Andler, « Cognitives (Sciences) », *op. cit.*, p. 66 et suivante.

conformer. Bref, toute théorie et tout vocabulaire, peu importe la discipline, devraient finalement pouvoir être expliqués (ou traduits) par l'appareil des sciences naturelles, ou en tout cas ne pas être en désaccord avec lui⁸. La pensée, par exemple, doit donc être *mise en causes* : elle doit s'inclure dans un monde composé uniquement de relations causales et de lois naturelles. Le pendant *ontologique* du naturalisme est donc bien souvent le *matérialisme*, (souvent assimilé au *physicalisme*⁹) : la seule chose qui existe, ultimement, c'est la matière (et l'énergie) ; tous les phénomènes qui peuvent exister (propriétés, états, événements, processus, relations) doivent donc ultimement se ramener à des phénomènes explicables par la physique contemporaine. Cette position ontologique est motivée par la peur d'un *spectre* : celui du dualisme cartésien, incapable d'expliquer l'interaction entre l'esprit (ou encore la non matière) et la matière (si l'esprit peut faire bouger le corps, alors il *doit* être physique car tout *effet* physique n'est rendu possible que par une *cause* physique).

Le fonctionnalisme cognitiviste est cette position qui défend l'idée que ce qui constitue la nature ou l'identité d'un état mental, ce n'est pas sa composition ou sa matière cérébrale, mais plutôt son rôle (sa *fonction*) dans le système cognitif. Ce ne sont pas alors les propriétés neurologiques des états et processus mentaux qui importent dans une étude de la cognition (elles seraient entre autres trop complexes et contingentes pour cela), mais plutôt leurs propriétés fonctionnelles, leurs rapports causaux avec les entrées, sorties et autres états/processus du système. Les propriétés et états mentaux doivent être identifiés avec des états ou propriétés structurelles, organisationnelles (la fonction d'un état mental particulier n'est pourtant pas rigoureusement prédéfinie comme c'est le cas pour les entités des théories fonctionnalistes en anthropologie par exemple). Comme ma croyance qu'il pleut

8. Plus précisément, les disciplines ayant pour objet d'étude la pensée ou les manifestations de la pensée devraient adopter la *typologie* (vocabulaire), la *nomologie* (appareil de prédiction faisant appel à des *lois*) et la *méthodologie* (appareil de justification, expérimental) des sciences naturelles. Voir D. Fissette et P. Poirier, *Philosophie de l'esprit. État des lieux*, Paris, Vrin, 2000, chapitre 3. Pour une exemplification du programme naturaliste en philosophie de l'esprit (et des impasses qu'il rencontre), voir les ouvrages de P. Jacob (*Pourquoi les choses ont-elles un sens ?*, Paris, Odile Jacob, 1997, 347 p.) et J. Proust (*Comment l'esprit vient aux bêtes*, Paris, Gallimard, 1997, 391 p.). Pour une attitude plus nuancée, voir l'ouvrage de P. Engel *Philosophie et psychologie* (voir bibliographie).

9. À dire vrai, les deux termes ne sont pas rigoureusement synonymes, mais on laissera ici ces subtilités de côté.

peut se *réaliser* de manière *multiple* dans mon cerveau, elle se définit de préférence par les relations causales qu'elle peut entretenir avec d'autres états internes ou externes du système (sortir mon parapluie, désirer me mettre au sec, annuler mon rendez-vous...). Et, plus généralement ici – c'est là l'idée cruciale qui permet au fonctionnalisme de justifier le cognitivisme, mais qui peut le faire rentrer en contradiction avec le naturalisme –, un *système* cognitif est donc surtout une entité qui *peut* transformer des états informationnels d'entrée en d'autres états, et ce *quelle que soit sa composition matérielle*. L'architecture du système, assimilable au programme informatique (*software*), est donc indépendante de son matériel (*hardware*). Bref, pour être qualifié de « système de pensée », il n'est pas *nécessaire* d'avoir un cerveau. Comme le disait naguère Putnam, un système cognitif pourrait être fait de gruyère. Voilà une thèse très libérale, qui qualifie de système cognitif bon nombre d'entités sous prétexte qu'elles manipulent et transforment des informations (un thermostat ou une calculatrice sont-ils des systèmes cognitifs ? Qu'en est-il de l'ordinateur ?). Le fonctionnalisme est d'autant plus intéressant pour le cognitiviste en ce qu'il permet d'élaborer une théorie *systématique* du fonctionnement de l'esprit uniquement à partir des propriétés fonctionnelles des états mentaux, relatives à leur syntaxe (voir *infra*). Le fonctionnalisme évite à la fois le chauvinisme et les obstacles des théories de l'identité esprit-cerveau (en étant cependant parfois trop libéral) et les impasses des théories béhavioristes et dualistes de l'esprit.

Un paradoxe peut ici frapper le lecteur : si le fonctionnalisme ne considère pas la *matière cérébrale* comme étant le critère décisif d'une étude et d'une définition de la pensée, comment peut-il être compatible avec la thèse naturaliste ? Une réponse minimale mais décisive consiste à dire que le fonctionnalisme et le naturalisme s'accordent sur leur rejet de tout dualisme ontologique et méthodologique au sujet de l'esprit : n'existe, *in fine*, que la matière (physicalisme). Pour la majorité des philosophes fonctionnalistes aujourd'hui, les états mentaux *dépendent* des états cérébraux (théories de la survenance, de l'identité-occurrence...). La connaissance est un phénomène *naturel* ; l'esprit – ou plutôt ses manifestations – peut être étudié d'une manière scientifique, il n'est pas *supernaturel*. Ces deux positions sont alors très complémentaires : le fonctionnalisme justifie la thèse naturaliste en tentant de montrer comment de la matière cérébrale peut produire systématique-

ment de la cognition au moyen d'une étude des propriétés fonctionnelles d'états informationnels *in fine* réductibles à des états cérébraux. Comme cela a déjà été remarqué, pour la plupart des fonctionnalistes le mental possède une autonomie descriptive par rapport au cérébral. En ce qui concerne l'esprit, on peut donc refuser de souscrire à une forme extrême de naturalisme (en défendant une autonomie descriptive du mental, du cognitif) tout en adhérant à son pendant ontologique (physicalisme). Pour certains partisans du cognitivisme, sceptiques par rapport au naturalisme pur et dur, si l'on souhaite efficacement décrire et expliquer le fonctionnement de la cognition, il est souvent nécessaire (comme en psychologie cognitive) de se situer à un niveau d'explication supérieur au niveau purement neurophysiologique. Mais le cognitiviste qui reconnaît la justesse de la thèse naturaliste tentera de respecter des contraintes *naturelles* dans l'explication du phénomène qu'il étudie (démarche *top-down*, reconnaissant une continuité des sciences entre elles) et trouvera *concevable* l'idée qu'un jour sa théorie puisse se réduire à une théorie physicaliste. Rien ne peut lui interdire de penser que, pour l'instant en tout cas, une explication physicaliste d'un phénomène cognitif *complexe* serait incroyablement *pauvre, triviale, limitée* ou – si l'on rejette tout à fait la thèse naturaliste – *à côté de la plaque*.

Passons maintenant à l'hypothèse cognitiviste proprement dite, et à ses deux sous-thèses :

Selon F. Varela, E. Thompson et E. Rosch, l'hypothèse de base du cognitivisme est que « l'intelligence – humaine comprise – ressemble tellement à la computation dans ses caractéristiques essentielles que la cognition peut en fait se définir par des computations sur des représentations symboliques [...]. Une computation est une opération effectuée ou accomplie sur des symboles, c'est-à-dire sur des éléments qui représentent ce dont ils tiennent lieu¹⁰ ».

Deux notions sont ici centrales : celle de *représentation* (symbolique) et celle de *computation*. Avant de les expliciter, signalons que le cognitivisme classique est aussi un peu hâtivement assimilé, ou du moins apparenté, au *computationnalisme*, à la *théorie représentationnelle de l'esprit*, au *paradigme classique* ou encore au *modèle règles-représentations*.

10. *L'Inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, traduction française de V. Havelange, Paris, Seuil, 1993, p. 73.

Le cognitivisme classique se base sur une théorie particulière du raisonnement (computationnalisme) et sur une théorie particulière des états mentaux, la théorie représentationnelle. Pour cette dernière, être dans un état mental particulier (doué de la propriété d'intentionnalité, c'est-à-dire dirigé *vers* quelque chose), comme croire qu'il pleut, c'est *simplement* entretenir une certaine relation avec une représentation symbolique de l'objet de l'état. Dit autrement, croire, désirer, savoir, ou encore craindre que *p*, c'est avoir une représentation propositionnelle, de format sententiel, de *p* située dans sa boîte à croyance, à désir... Similairement, disposer d'un concept, c'est disposer d'une représentation mentale de ce même concept. J. Fodor ira plus loin, en défendant l'idée que l'objet propositionnel de l'état psychologique est encodé sous la forme d'un « langage » de la pensée, le *mentalais* (cela nous permet entre autres de rendre compte des capacités cognitives des enfants et des animaux, tout comme de la productivité, de la systématité et de la compositionnalité de la pensée)¹¹. Bien sûr, toute représentation *en mentalais* de l'objet de l'attitude psychologique (personnelle ou subpersonnelle) est supposée s'implémenter dans un médium neuronal, mais cette implémentation présente une certaine structure formelle, relative à un langage : ici, le mentalais. Un contenu représentationnel dénotant un état du monde extérieur (le fait qu'il pleuve, par exemple) s'incarne donc dans un ensemble de neurones, et, via ses propriétés physiques, possède des propriétés formelles, ou encore syntaxiques, qui lui permettent d'interagir avec d'autres états représentationnels en étant l'objet d'inférences computationnellement réglées (ma croyance qu'il pleut peut être mise en relation avec mon désir de rester sec et ma vision d'un abribus, ce qui m'amènera à me déplacer).

La théorie du raisonnement – théorie computationnelle – part en effet de cette hypothèse que les représentations propositionnelles ont des propriétés syntaxiques, et, partant, sémantiques. Elle y ajoute alors l'hypothèse que la pensée rationnelle consiste en la manipulation, gouvernée par des règles (des algorithmes), de ces représentations syntaxiquement structurées. Ce qu'il importe d'abord de comprendre

11. Voir le texte de Jerry Fodor « Pourquoi il doit encore y avoir un langage de la pensée ». Traduit en français dans *Philosophie de l'esprit. Psychologie du sens commun et sciences de l'esprit*, textes réunis par D. Fiset et P. Poirier, Paris, Vrin, 2002.

ici, c'est que *la syntaxe a une priorité sur la sémantique*. L'un des slogans cognitivistes est : *Occupez-vous de la syntaxe et la sémantique s'occupera d'elle-même*¹². Pourquoi cette priorité ? La réponse est la suivante : les processus de raisonnement qui prennent place en nous à un niveau subpersonnel ne peuvent avoir comme objets les propriétés sémantiques des représentations, car autrement il faudrait postuler l'existence d'un homoncule qui « comprenne » chaque contenu propositionnel, à l'aide d'autres règles et représentations, et ainsi de suite jusqu'à l'infini. Comment des contenus représentationnels pourraient-ils alors être agissants ? *Par leur forme syntaxique*. Les ordinateurs – et, plus élémentairement les machines de Turing – sont, eux aussi, des machines à manipuler des symboles, en étant pourtant uniquement sensibles à la forme physique (la syntaxe) des symboles. Il suffit alors de s'inspirer de leur fonctionnement afin de rendre compte de la causalité mentale humaine.

Des opérations – gouvernées par des règles logiques valides – qui ne prennent pour objet que les propriétés syntaxiques des symboles ont pour caractéristique de préserver la valeur sémantique de ces symboles ; elles les transforment en suivant des règles, mais à la fin les propriétés sémantiques des symboles sont préservées ; bref, tout se passe *comme si* le processus computationnel comprenait les objets de ses opérations. On peut alors dire que les symboles mentaux ont un pouvoir causal en vertu de leurs propriétés syntaxiques, qui *miment* leurs propriétés sémantiques, et les propriétés syntaxiques sont déterminées par les propriétés physiques du symbole (ici son implémentation neurale)¹³. « Manipuler des symboles », c'est donc raisonner (transformer, produire...) sur des représentations de format sententiel uniquement à partir des propriétés syntaxiques de ces dernières et en suivant des règles formelles (dont nous n'aurions qu'une connaissance « tacite ») de la forme « si... alors... », spécifiées dans un algorithme. Ainsi, par exemple, pour Chomsky, connaître une langue revient à « être dans un certain état mental fait d'une structure de règles

12. Voir John Haugeland, *L'Esprit dans la machine : Fondements de l'intelligence artificielle*, traduction française de J. Henry, Paris, Odile Jacob, 1989, 256 p.

13. Pour ces points assez techniques et les problèmes qu'ils amènent, on renvoie ici le lecteur à l'étude de J.-M. Roy « Causalité intentionnelle et causalité symbolique », dans J.-L. Petit (dir.), *Les Neurosciences et la philosophie de l'action*, Paris, Vrin, 1997, p. 293-325.

et de principes¹⁴ ». Cette connaissance est mise en application lorsque nous traitons des informations auditives, par exemple, qui deviennent alors signifiantes. Une fois de plus, le respect ou le suivi de ces règles n'est pas le produit d'un raisonnement conscient, mais relève plutôt de l'organisation générale de l'esprit/cerveau : le modèle est ici, encore une fois, l'ordinateur; celui-ci effectue des opérations à partir de règles qu'*il* ne considère cependant pas « devant l'esprit ». Dans une perspective plus réaliste, on dira que les cerveaux sont des engins syntaxiques, et non pas sémantiques. La théorie computationnelle du raisonnement résulte de deux grandes entreprises intellectuelles des XIX^e et XX^e siècles : la formalisation des mathématiques (Hilbert, Russell & Whitehead), et le travail pionnier d'Alan Turing (et d'A. Church, sans oublier celui de Gödel), dans les années trente, dont une des intuitions fondamentales était que toute opération (mathématique, par exemple) qui était uniquement sensible aux propriétés syntaxiques pouvait être simulée par une machine (de Turing). On peut alors parler de *mécanismes de la pensée*. L'étude d'une activité cognitive pourrait s'effectuer à un niveau d'analyse indépendant des niveaux biologique ou phénoménologique ; un raisonnement portant sur des choses *signifiantes* – réelles ou imaginaires – peut être mené par un système qui ne connaîtrait rien des significations des objets du raisonnement. La formalisation nous montre comment les propriétés sémantiques d'une représentation peuvent jouer d'un rôle causal, par l'intermédiaire de la *syntaxe*. Émergeant aux XIX^e et XX^e siècles, cette vision du raisonnement remonte au moins jusqu'à Hobbes, et à son idée que « raisonner, c'est calculer ». Dans les années 1960, Von Neumann, McCarthy, Minsky, Newell et Simon ont concrétisé cette intuition fondamentale de Turing, chacun à leur manière (création de programmes et langages, simulation de capacités, théorisation poussée...).

Il semble cependant que notre logique naturelle, étant avant tout une logique (de la) pratique, outrepassa la logique formelle ou la logique informatique, et qu'elle n'obéisse pas toujours à des règles, en tout cas explicitement représentées (illusion régulateur)¹⁵. Comme l'écrit

14. *Règles et représentations*, traduction française de A. Kihm, Paris, Flammarion, 1985, p. 52.

15. Mais le cognitiviste ne défend pas cette position. Si tout mécanisme cognitif est formalisable, la formalisation n'implique pas que tout obéit à des règles présentant des contenus (comme « Fais X lorsque C »). Au contraire, beaucoup de « règles » résultent de ce que Pylyshyn (1984) appelle l'ar-

G. Vignaux, « on peut toujours trouver des règles formelles séduisantes pour modéliser nos activités symboliques, mais cela se fait nécessairement à distance des règles de fonctionnement et de mise en pratique de ces activités¹⁶ ». C'est cependant faire bon marché des possibilités de réduction de l'indétermination de l'attribution des règles et autres algorithmes, au moyen d'expériences et d'études de troubles et de dysfonctionnement en neuropsychologie cognitive par exemple.

Certains diront que, incapables pour l'instant de répondre à des questions comme : « Comment notre système accomplit-il ce type d'activité ? », les cognitivistes se posent alors des questions comme : « Quelles caractéristiques abstraites ce système devrait-il posséder pour réaliser cette action ? » Ce genre de questions constitue notamment l'une des bases de l'intelligence artificielle (étude des conditions de possibilité de la pensée à partir de la *création* de systèmes de pensée, comme les robots ou les ordinateurs).

III. Cognitivisme et niveaux d'analyse

Une exemplification méthodologique du programme cognitiviste peut se trouver dans le travail séminal (posthume) de David Marr sur la vision¹⁷. Toute théorie cognitive, selon Marr, devrait se construire à partir de trois niveaux. Le premier niveau (qu'il appelle, de manière ambiguë, *computationnel*) consiste en une spécification de la tâche (ou du traitement de l'information) qu'un système cognitif effectue (souvent à partir de ses sous-systèmes, il faut alors spécifier les tâches) et des contraintes qu'il rencontre : par exemple, dans le cas de la vision, la tâche est de construire une image tridimensionnelle du monde à partir des simples *inputs* bidimensionnels reçus par la rétine. La deuxième étape, la plus importante (que Marr appelle étape « algorithmique », et qui est authentiquement « computationnelle ») consiste à spécifier un algorithme possible de la fonction spécifiée au premier niveau (un algorithme est une méthode de réalisation – ou de « computation » – d'une fonction ; un ordinateur, par exemple, pour imprimer un document, une

chitecture fonctionnelle du système (biologique) et sont donc « cognitivement impénétrables » (il y aurait sinon, ici aussi, une régression à l'infini, comme celle d'Achille et de la tortue).

16. *Les Sciences cognitives. Une introduction*, Paris, La Découverte/Livre de Poche, 1991, p. 87.

17. *Vision*, San Francisco, Freeman Press, 1982.

de ses fonctions, exécute un algorithme automatique). Il s'agit aussi de déterminer le format des représentations qui seront l'objet de calculs. Établir un algorithme, c'est modéliser un processus avec l'hypothèse que ce processus agit en suivant un ensemble de règles, d'instructions déterminées (autrement dit, en exécutant un algorithme) qui nous permettent de réaliser la fonction. La modélisation du processus nous permet de dégager cet ensemble de règles. Le dernier niveau, enfin, niveau d'implémentation de la fonction, recoupe la réalisation neuro-physiologique (ou neurochimique) de la fonction et/ou des sous-fonctions. Nous comparons la solution du deuxième niveau à ce troisième niveau, afin d'évaluer sa plausibilité. Il est en effet bon de rappeler que nos capacités langagières, perceptives ou motrices dépendent *effectivement* de capacités cérébrales. La neuro-imagerie, au moyen d'outils comme l'EEG (électro-encéphalographie), la MEG (magnéto-encéphalographie), la TEP (tomographie par émission de positons) ou l'IRM (imagerie par résonance magnétique) rendent aujourd'hui davantage possible la localisation cérébrale d'un certain ensemble d'activités cognitives (compréhension de certains types de mots, réalisation d'opérations mathématiques, déclenchement de certaines actions motrices). Il est cependant rare qu'une activité dépende exclusivement d'un « endroit ». D'où parfois la nécessité d'un niveau de description et d'étude plus large que le niveau neurophysiologique, avec les risques d'imprécision et d'abstraction que cela comporte : le niveau fonctionnel, où se situe le paradigme cognitiviste.

On l'aura remarqué, avec ces notions de fonction et d'algorithme, l'idée centrale est qu'un système cognitif réalise, ou compute, une fonction et qu'il est possible de détailler l'algorithme (les règles) qui réalise cette fonction en manipulant des représentations structurées. Comme le dit Fodor, *pas de computation sans représentation* (autrement dit, tout calcul se fait à l'aide de représentations) ou encore *pas de causalité mentale sans représentations explicites de contenus mentaux*.

Dans le même esprit que celui de D. Marr, Z. Pylyshyn a proposé trois niveaux d'explication distincts d'un comportement cognitif : le niveau biologico-physique (la tâche est directement explicable en référence à des lois naturelles (vieillessement cognitif par exemple) ; le niveau syntaxique (ou encore symbolique, fonctionnel) qui est le niveau de la computation (des algorithmes) et des capacités computationnelles (architecture fonctionnelle – *grosso modo* le programme – et opérations rela-

tives à la syntaxe de l'état représentationnel); le niveau sémantique (pour expliquer le comportement, nous devons parler du contenu de l'information qui a provoqué le comportement – croyances, désirs de l'agent; mais ce contenu est seulement causalement actif au second niveau – syntaxe). Une explication simple d'un comportement (psychologie populaire) peut s'en tenir à ce dernier niveau (« Il lui a offert des fleurs, car il croyait qu'elle aimait cela »). Il existe cependant des explications cognitives qui peuvent se dispenser d'invoquer le contenu des représentations, en ce qu'elles tentent de mettre au jour les propriétés de calcul fondamentales (de traitement de l'information) de l'architecture du système, qui sont d'ailleurs *cognitivement impénétrables*: elles ne dépendent pas de ce que nous croyons ou pouvons croire, étant donné qu'elles rendent possibles les interactions computationnelles entre nos états cognitifs. Le niveau du calcul (niveau syntaxique) réalise le travail du troisième niveau, tout en étant rendu possible par le niveau biologique-physique.

IV. Cinq critiques (parmi d'autres) du paradigme cognitiviste

a. Un nouveau paradigme : le paradigme connexionniste

Pendant longtemps l'une des forces du cognitivisme était son caractère unique. À de nombreux moments, Chomsky ou Fodor ont invité de possibles détracteurs à critiquer leurs théories. Comme le disait Fodor¹⁸ dans les années 1970 ,

1. les seuls modèles psychologiques plausibles des processus cognitifs représentent de tels processus comme computationnels;
2. la computation présuppose un médium de computation: un système représentationnel;
3. mieux vaut avoir des théories vaguement plausibles que pas de théories du tout;
4. nous devrions donc attribuer un système représentationnel aux organismes.

Deux fois dans cet argument le principe d'inférence à la meilleure explication est mobilisé: par ce principe, nous reconnaissons l'exis-

18. *The Language of Thought*, New York, Cromwell, 1975, p. 27.

tence d'un processus ou d'une entité (computations et représentations) lorsque la meilleure (et également ici seule) explication d'un phénomène l'exige. Or, depuis une vingtaine d'années¹⁹, le paradigme connexionniste remet en question les deux présupposés centraux du cognitivisme (computation et représentation) en proposant un nouveau modèle de la cognition, plus sensible aux propriétés cérébrales des systèmes cognitifs humains. L'attrait initial du paradigme connexionniste réside dans son désir de concevoir la cognition à partir des propriétés du système cérébral. Le modèle est ici le cerveau, et non plus l'ordinateur. La notion centrale dans le connexionnisme est celle de *réseau de neurones*. Elle est supposée remplacer celle de *représentations symboliques*. Une autre notion est celle de *règles émergentes* (ou *implicites*). Notons que la notion de *réseau de neurones* émergea dès 1943 (McCulloch et Pitts, cybernétique).

Les modèles computationnels de la cognition parviennent à rendre compte de tâches comme jouer aux échecs, calculer ou encore résoudre un problème. Ils éprouvent en revanche beaucoup plus de difficultés pour des tâches basiques comme notre capacité à reconnaître un visage, conduire une voiture ou reconnaître et comprendre une structure linguistique en un temps minimal – activités habituelles dont il paraît moins plausible qu'elles soient gouvernées par des règles. Les modèles connexionnistes expliquent, eux, nos capacités cognitives à partir de systèmes composées d'ensemble d'unités, qui correspondent fonctionnellement – et grossièrement – à des neurones. Les connexions entre chaque unité correspondent aux connexions synaptiques. Par l'intermédiaire des connexions qui la relie aux autres unités, chaque unité peut transmettre un signal (équivalant à une impulsion électrique), en fonction de son niveau d'activation. Ce signal peut être un signal d'excitation (charge positive) ou un signal d'inhibition (charge négative). Ce signal provient d'autres unités activées. En amont, il relève d'un *input* codé holistiquement (un visage, par exemple) auquel le système réagit en transmettant – et en transformant – des signaux, qui, en aval (*output*), doivent aboutir à une configuration particulière (qui signifie – au moyen de valeurs numériques – le visage de X ou de Y). La transmission du signal dans le réseau dépend du « poids » des connexions.

19. La référence du paradigme connexionniste est l'ouvrage de Rumelheart & McClelland (eds.) *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, qui date de 1986.

Pour modifier le comportement du système, on configure ces poids. Modifier le poids d'une liaison, c'est transformer la valeur d'un signal. La force du signal transmis par une unité dépend du niveau d'activation global de l'unité, qui est lui-même relatif, d'une part, au nombre de connexions de cette unité, au poids de chacune de ses connexions, à leur polarité (positive ou négative) et à la force du signal reçu (tout cela constitue la variable), et, d'autre part, au seuil d'activation par défaut (constante) de l'unité (en dessous de ce seuil, l'unité ne transformera pas le signal reçu). Une unité est un peu un carrefour. S'y rejoignent différentes connexions – donc souvent différents signaux – que l'unité va « réorganiser » en fonction de son niveau d'activation, et puis transmettre à nouveau aux autres unités. Une unité ne représente rien ; ce qui peut représenter quelque chose, c'est un ensemble, une configuration d'unités activées. Le connexionnisme ne souhaite pas renoncer au thème de la représentation ; il souhaite plutôt, généralement, considérer celle-ci comme *distribuée* et non pas comme dépendante d'une seule entité physique (au contraire des représentations symboliques, qui sont *locales*) ; elle résulte alors de l'activation d'un grand nombre de neurones ; si l'un d'entre eux devait disparaître, cela n'affecterait pas la « représentation » produite par le système. À la différence des systèmes cognitivo-computationnalistes, les systèmes connexionnistes sont donc *sous-symboliques*²⁰ : les capacités cognitives ne relèvent pas nécessairement d'une manipulation de symboles ou d'états qui ont pour référence le monde extérieur. On ne trouve pas, dans le réseau, d'équivalents permanents de nos représentations symboliques quotidiennes ; il n'y a pas d'unité discrète qui représente le concept de « chien », de « pleuvoir »...

Le connexionnisme pense pouvoir innover de par la théorie du raisonnement et de la compétence qu'il propose : le réseau de neurones apprend à répondre à une tâche cognitive donnée (reconnaître un visage) au moyen d'un *apprentissage statistique*, par essais et erreurs. Le réseau ne consulte aucune règle explicite ; il apprend simplement à se configurer (à configurer le poids de ses connexions) en fonction de l'*input* proposé et de l'*output* souhaité. Il est également capable de *généralisations* (à partir d'une suite de visages présentés,

20. Pour reprendre l'expression de P. Smolensky, l'un des défenseurs de l'approche connexionniste. Voir son article, « Le traitement approprié du connexionnisme », traduit en français dans *Philosophie de l'esprit. Problèmes et perspectives*, textes réunis par D. Fisette et P. Poirier, Paris, Vrin, 2003.

il est capable de dégager un visage *prototype* sans que l'on encode la configuration de ce dernier). Un réseau de neurones peut donc effectuer une tâche cognitive sans utiliser de représentations explicites (si représentations il y a, elles sont l'effet d'une certaine interprétation) et sans appliquer ou suivre des règles ou un programme prédéfini – le système apprend à se configurer, et puis réagit (tout en pouvant se reconfigurer par la suite, en cas de changement de l'*input*). Les représentations sont des vecteurs d'activation ; l'apprentissage consiste en la modification des poids synaptiques ; et la computation (raisonnement) consiste dans le passage de vecteur à vecteur au moyen des poids des connexions.

Pour certains, cependant, le connexionnisme n'est qu'une forme de béhaviorisme, ou reste prisonnier de la psychologie associationniste en étant incapable de rendre compte des capacités de *systematicité* et de *productivité*, inhérentes à la pensée. Fodor et Pylyshyn ont par exemple critiqué les théories connexionnistes en arguant du fait qu'elles sont, dans le meilleur des cas, peu novatrices ou, dans le pire des cas, incohérentes et incomplètes : elles seraient peu novatrices en ce qu'elles ne feraient que montrer comment une représentation qui, d'un certain point de vue, est locale, résulte en fait d'une configuration de neurones. Mais, dans ce cas, elles ne sont pas très pertinentes pour une science de la cognition, qui, rappelons-le, vise moins à décrire l'implémentation de la pensée que son fonctionnement et sa systématicité. La théorie serait incohérente ou incomplète si l'on considère les capacités d'apprentissage et les performances des systèmes connexionnistes : ou bien cette théorie assume ne pas reconnaître l'existence d'un langage de la pensée (qui, rappelons-le, rendrait possible le caractère structuré de la pensée) ou de règles, mais dans ce cas elle doit pouvoir rendre compte autrement de deux capacités fondamentales de la pensée, à savoir la productivité et la systématicité ; ou bien elle parle de règles implicitement représentées, auquel cas elle ne diffère pas de la théorie cognitive qui, soulignons-le, ne postule pas l'existence de règles explicites, mais seulement de représentations explicites (les règles sont bien souvent implicites, mais nous disposons d'une capacité à les rendre explicites, par exemple pour les réviser). Bien entendu, la réponse de Fodor et Pylyshyn a elle-même été critiquée. La question reste cependant la même : peut-on expliquer la cognition sans faire appel aux *représentations* et aux règles (implicites ou explicites) ?

Pour peu que l'on adopte un état d'esprit ouvert, l'opposition entre cognitivisme et connexionnisme paraît moins tranchée qu'elle n'en a l'air. Les deux paradigmes seraient moins concurrents que complémentaires. Le computationnalisme s'occuperait de modéliser et d'expliquer des tâches et capacités cognitives qui, d'un point de vue évolutionniste, sont récentes (jouer aux échecs, résoudre un problème quelconque, entretenir des métareprésentations), tandis que le connexionnisme s'occuperait de phénomènes plus « primitifs » (reconnaissance perceptive, sensori-motricité...). Mais pour certains phénomènes, comme la compréhension langagière, les deux paradigmes restent concurrents. Reste cependant que ces deux paradigmes naturalistes partagent encore au moins trois présupposés : celui de la représentation (la cognition se fait par représentations), celui de l'oubli de la conscience, et celui de l'isolement du système cognitif par rapport à son environnement et à la temporalité.

b. Critiques neuroscientifiques

Certains scientifiques s'inscrivant aussi dans le programme de recherche des sciences cognitives (P. M. et P. S. Churchland, J.-P. Changeux, G. Edelman, A. Damasio, T. Shallice, A. Berthoz, M. Jeannerod) sont parfois hostiles, la plupart du temps indifférents, aux théories cognitivistes ou connexionnistes de la cognition (même si ces dernières souhaitent être plus réalistes que les premières). Ce qui est rejeté ici, c'est l'hypothèse fondamentale (et exposée plus haut) selon laquelle on pourrait étudier le mental, le cognitif, indépendamment de sa réalité cérébrale (même si l'on admet qu'il existe des *contraintes* empiriques), c'est-à-dire à un niveau global, fonctionnel ou algorithmique. Ce niveau serait une mythologie. Ces critiques tournent même parfois en dérision les théories de l'organisation cérébrale élaborées par les cognitivistes classiques²¹. Certains neuroscientifiques auront également tendance à railler la simplicité et l'artificialité des modèles connexionnistes et à critiquer leur tendance à vouloir préserver un sens, fût-il minimal, à la notion de représentation. Pour certains de ces théoriciens, les neurosciences (sciences du cerveau, et plus généralement du système

21. Pour un aperçu, voir J.-N. Missa, *L'Esprit-cerveau. La philosophie de l'esprit à la lumière des neurosciences*, Paris, Vrin, 1993.

nerveux) nous apportent et nous apporteront une nouvelle considération de l'être humain, et doivent de ce fait participer, au même titre que la philosophie ou les sciences humaines par exemple, à une étude des manifestations cognitives culturelles (éthique, création et jugement artistiques, pathologies mentales...)²².

Les approches connexionnistes et neuroscientifiques reprochent au cognitivisme de ne pas être assez radical dans son optique naturaliste, en restant prisonnier du fonctionnalisme et du représentationnalisme symbolique. Ce qui n'est pas le cas des deux noyaux critiques que nous allons maintenant évoquer, qui dénoncent aussi bien les abus naturalistes propres à toute science cognitive que l'approche formaliste privilégiée par le cognitivisme.

c. Approches et intuitions phénoménologiques : corporéité et conscience

Plus conceptuelles et proches du sens commun, d'autres critiques du cognitivisme, principalement d'obédience philosophique, se sont développées. Depuis 1972, un philosophe comme H. Dreyfus a élaboré beaucoup de travaux destinés à montrer, intuitivement, que la pensée humaine ne peut être comprise à partir du paradigme « règles et représentations », et à démontrer qu'un ordinateur ou un robot ne pourront jamais effectuer des tâches cognitives dans des situations réelles (le nombre d'assomptions et de capacités de discrimination – le fameux *frame problem* – étant trop élevé). Dreyfus insiste sur la prévalence du *savoir-faire* (habitudes, routines, prototypes) sur le *savoir propositionnel* (règles et définitions) dans la réalisation de nos tâches cognitives quotidiennes (nouer nos lacets, conduire une voiture, reconnaître une chaise, savoir *quoi faire* lors d'une situation imprévue). Il est impossible, pour Dreyfus, de formaliser tout le noyau de connaissance du sens commun dont nous disposons et qui nous permet d'agir dans le monde. La raison en est que cette connaissance du sens commun ne réside pas dans des règles, mais est plutôt relative à un *Arrière-Plan* de pratiques, de routines, d'institutions et d'interactions corporelles avec

22. Changeux est, en France, le principal promoteur de ces idées. Pour un exemple actuel, voir les propos qu'il tient dans *Le Monde* du 2 février 2005 (p. 23), où, parmi d'autres, il affirme que « les neurosciences inaugurent les lumières du XXI^e siècle ».

le monde. Ce qu'une théorie computationnelle de la cognition manque d'expliquer et d'intégrer dans ses modèles informatiques, c'est alors

1) toute notre connaissance et nos assumptions d'arrière-plan relatives au monde biologique et culturel; cette connaissance relève d'un savoir-faire;

2) notre capacité de discrimination d'éléments pertinents de situations, qui nous permet de répondre intelligemment à des situations inattendues. Cette capacité est analogue à notre « bon sens » de ce qu'il faut faire ou pas dans telle ou telle situation.

Notons que certains chercheurs en intelligence artificielle remettent en question les doutes de Dreyfus concernant (1) en élaborant un inventaire de toutes les propositions propres au sens commun que pourrait « connaître » un ordinateur. Il y en aurait un peu plus de cent millions.

John Searle a également critiqué le computationnalisme, à partir de sa célèbre expérience de pensée de la *chambre chinoise*²³, destinée à montrer que le fait d'opérer sur des symboles uniquement à partir de leurs propriétés syntaxiques comme le fait un ordinateur n'est pas *suffisant* pour être un modèle de la connaissance humaine, en ce qu'un système syntaxique ne peut rendre compte de la dimension de la compréhension, du *sens*, propre à nos activités cognitives (dans d'autres termes, la syntaxe n'est qu'une réduction appauvrissante de la sémantique, cette dernière n'étant d'ailleurs pas uniquement *référentielle*). Imaginons que vous êtes placé dans une chambre close. Ne se trouvent par terre que des morceaux de papier représentant des idéogrammes chinois, ainsi qu'un ensemble d'instructions (du genre « si idéogramme *x* alors idéogramme *y* ») destinés à faire correspondre certains idéogrammes à d'autres idéogrammes. Par une fente située en dessous de la porte, on vous transmet à nouveau un ensemble d'idéogrammes chinois. Vous exécutez les instructions, et renvoyez par l'intermédiaire de la fente les paires d'idéogrammes. Ces paires d'idéogrammes sont en fait composées d'une question et d'une réponse. Un observateur qui parle chinois et qui se situe en dehors de la pièce peut penser que la personne qui se trouve dans la chambre (en l'occurrence vous) sait parler chinois (elle renvoie des paires de questions-réponses sensées). Mais peut-on dire que vous parlez chinois alors que vous ne *compre-*

23. Reprise dans J. Searle, *Du Cerveau au savoir*, traduit en français par C. Chaleysin, Paris, Hermann, 1985, 143 p.

nez rien du tout aux couples d'idéogrammes que vous renvoyez? Le but de cette expérience de pensée est bien sûr de poser la même question au sujet des ordinateurs et des modèles syntaxiques de la compréhension sémantique. Pour Searle, ce qui manque à l'ordinateur pour être qualifié de système de pensée, c'est la capacité de *compréhension*. Le cognitivisme propose donc une image erronée de la pensée humaine, en la réduisant *intégralement* à des schèmes règles-représentations (opérations syntaxiques), et en étant donc incapable de rendre compte de la compréhension. Il n'y a pas de pensée sans compréhension (donc sans conscience, voire diront d'autres sans situation, sans incarnation, sans environnement). Évidemment, cette critique de Searle n'est pas restée sans suites. On a tenté de la contrecarrer de multiples façons²⁴.

Élaborant sa théorie de l'*enaction*, Francisco Varela a tenté de dépasser une assomption commune aussi bien au cognitivisme qu'au connexionnisme : celle de la représentation, conçue comme interface entre l'esprit et le monde. Varela nie l'existence des représentations (antireprésentationnalisme) et tente d'élaborer un modèle de la cognition extrêmement sensible à l'expérience humaine du monde et à la constitution pragmatique du monde de la vie, s'inspirant entre autres de la phénoménologie, de la biologie et des traditions bouddhistes orientales. Son approche semble cependant trop extrême, en ce qu'elle propose de nier l'existence des représentations. Comment dès lors rendre compte du comportement, des erreurs perceptives, des anticipations... ?

La majorité de ces critiques insistent sur l'oubli de la *conscience* dans les paradigmes cognitivistes, connexionnistes mais aussi neuroscientifiques. Or, la conscience semble essentielle à l'esprit (cartésianisme), en ce qu'elle permet notamment aux choses et au monde de nous apparaître *selon une certaine manière*, et avec *un certain effet* (importance des émotions), qui n'est d'ailleurs toujours pas représentatif. Toute tentative d'explication de la cognition la mettant de côté serait donc vouée à l'échec. Il existe pourtant des théories cognitivistes (mais aussi connexionnistes...) de la conscience (Jackendoff, Edelman, Searle...): celle de Dennett en est un exemple, et est justement très critique de cette absolutisation du paradigme cartésien. Pour Dennett,

24. Pour une synthèse, voir J. Moyal, « The Chinese Room Argument », dans Barry Smith (éd.), *John Searle*, Cambridge University Press, 2003, 292 p.

la conscience n'est pas un *lieu* (un *théâtre*, une *télévision*); elle est encore moins un *centre spécial* du cerveau. Elle est plutôt une propriété de certains états mentaux, qui résulte de processus locaux traitant des données sensorielles. Certains abus de langage, certaines expériences que nous estimons « pleines » nous amènent cependant à postuler l'existence de *la* conscience et à la considérer comme essentielle à l'esprit. Comme il l'écrit, « étant donné que vous n'êtes rien de plus que les sous-actions et processus variés qui vous composent, le genre de question suivante est toujours un piège : "Quand exactement ai-je (en opposition aux composantes variées de mon cerveau) été informé (conscient) d'un événement quelconque ?" L'expérience consciente [...] est une succession d'états *constitués par* des processus variés qui prennent place dans le cerveau, et non pas quelque chose d'autre au-dessus de ces processus qui serait *causé par* ces derniers²⁵ ».

d. Confusions conceptuelles et caractère externaliste de l'esprit

Un ensemble d'auteurs, au rang desquels on trouve Meredith Williams, H. Putnam, D. Davidson ou, en France, V. Descombes, se sont entre autres choses inspirés des travaux séminaux de Peirce, Wittgenstein, Anscombe, Geach ou Ryle afin de critiquer le projet d'une *science* mentaliste et internaliste de l'esprit. Les critiques sont essentiellement d'ordre conceptuelles, visant à souligner l'ambiguïté et, souvent, le non-sens, généré par des termes et expressions comme « représentations mentales », « contenu symbolique », « réduction syntaxique de la sémantique » « langage de la pensée » ou « causalité intentionnelle ». L'importance du caractère holiste, social et linguistique de l'esprit est également souligné, et considéré comme non naturalisable (voir les contributions de Philippe Lacour et de Jacques Plouin à ce numéro).

e. L'esprit dans le monde

Depuis une dizaine d'années, certains philosophes et chercheurs ont tenté de ranimer le mot d'ordre suivant utilisé par W. M. Mace pour

25. « Real Consciousness », dans D. Dennett, *Brainchildren. Essays on Designing Mind*, MIT Press, Cambridge (Mass.), 1998, p. 136, souligné par l'auteur (ma traduction).

caractériser la théorie gibsonienne de la perception : « Ne demandez pas ce qu'il y a à l'intérieur de votre tête, mais demandez-vous plutôt à l'intérieur de quoi est votre tête²⁶. » Cette remarque fait signe vers une nouvelle méthodologie pour étudier la cognition dans toute sa complexité. L'un des points de départ se situe dans le constat que les êtres humains manipulent et stockent de l'information, mais d'une manière particulière. L'homme n'est en effet pas – seulement – un classer. Il utilise les (infra)structures de l'environnement et les opérations de ce dernier pour créer, manipuler et stocker des informations, ce qui permet par ailleurs à son système d'exhiber de la productivité et de la systémativité. Il s'agit alors d'étudier l'esprit selon un externalisme actif : il convient de penser *ensemble* la perception, la cognition et l'action, dans un rapport d'*intégration mutuelle* ou encore de *couplage* avec l'environnement, aussi bien biologique que technique et social. Plus radicalement encore, il faudrait, pour certains (van Gelder), considérer l'esprit comme un système dynamique dont toutes les composantes sont interdépendantes. Cette interdépendance assure le fonctionnement de l'esprit, en ce que la pensée prend place *en temps réel* (les états cognitifs *durent* ; ils sont en interaction les uns avec les autres). Il faut donc penser ensemble cognition et temporalité. La cognition doit être conçue de manière dynamique. La représentation conserve une place, mais son rôle et sa nature ne doivent pas être exagérés. Cette manière de penser la cognition s'attache à conserver certaines catégories et intuitions cognitivistes, pour certains domaines ; elle se base également sur certaines propositions connexionnistes tout en prêtant attention aux neurosciences²⁷.

V. Conclusion

L'approche cognitiviste n'est qu'une manière de mener une étude scientifique de l'esprit, à partir de deux intuitions fondamentales : le représentationnalisme et le computationnalisme (modèle règles/représentations). Avec A. Clark, on peut dire que le cognitivisme adopte la

26. « Gibson's strategy for perceiving: Ask not what's inside your head but what your head's inside of », dans R. Shaw & J. Bransford (eds.), *Perceiving, Acting and Knowing*, Hillsdale, NJ: Erlbaum, 1974.

27. Pour un exemple de ce paradigme, voir A. Clark, *Being There: Putting Mind, Body and World Together Again*, Cambridge (Mass.), MIT Press, 1997, 308 p.

méthodologie *MIND* : il se focalise sur des capacités cognitives qui, d'un point de vue évolutionnaire, ont émergé lors de la Maturité de l'être humain ; il tente d'élaborer des théories explicatives *Internalistes* (qui ne font pas appel à l'environnement ; toutes les ressources de l'esprit sont dans la tête) de ces capacités ; ces théories doivent être *Nettes* et globales (formalisation, déni de la temporalité) ; enfin, l'étude prend place selon une perspective de *Design* : on étudie les capacités comme si elles répondaient à une fonction prédéfinie. Peuvent s'ensuivre des critiques d'ordre scientifique (connexionnisme, neurosciences) de cette méthodologie, mais ces critiques se font au nom du même souhait qui anime le cognitivisme : constituer une science naturelle de l'esprit, en étudiant les capacités de ce dernier (perception, langage, raisonnement...) en tant que phénomènes naturels, sans se soucier de son insertion sociale, son caractère public, et surtout signifiant. Puissent les contributions de ce numéro mettre au jour le caractère au mieux incomplet, au pire illusoire, de ce projet*.

BIBLIOGRAPHIE

Ouvrages cités et/ou recommandés
pour une introduction plus complète

ANDLER Daniel (dir.), *Introduction aux sciences cognitives*, Paris, Gallimard/Folio (Essais), 1992, 516 p.

–, *Cognitives (sciences)*, Encyclopaedia Universalis, Paris, 1993.

–, « Processus cognitifs », dans D. ANDLER, A. FAGOT-LARGEAULT et B. SAINT-SERNIN, *Philosophie des sciences I*, Paris, Gallimard/Folio (Essais), 2002 (p. 226-407), 672 p.

BECHTEL William et ABRAHAMSEN Adele, *Le Connexionnisme et l'esprit. Introduction au traitement parallèle par réseaux*, tr. fr. J. Proust, Paris, La Découverte, 1993, 356 p. (épuisé).

CHANGEUX Jean-Pierre et RICŒUR Paul, *Ce qui nous fait penser. La nature et la règle*, Paris, Odile Jacob, 1998, 350 p.

* Je remercie Philippe Lacour de m'avoir confié ce travail introductif et pour ses diverses remarques lors de sa rédaction.

- CHURCHLAND Paul, *Le Cerveau : moteur de la raison et siège de l'âme*, tr. fr. A. Péliissier, Paris, De Boeck, 1999, 368 p.
- CUMMINS Robert et CUMMINS Denise D., *Minds, Brains, and Computers. The Foundations of Cognitive Science. An Anthology*, Blackwell, Oxford/New York, 2000, 552 p.
- DESCOMBES Vincent, *La Dénrée mentale*, Paris, Minuit, 1995, 348 p.
- DUPUY Jean-Pierre, *Aux origines des sciences cognitives*, Paris, La Découverte, 1994, 187 p.
- ENGEL Pascal, *Introduction à la philosophie de l'esprit*, Paris, La Découverte, 1994, 253 p.
- , *Philosophie et psychologie*, Paris, Gallimard/Folio, 1996, 473 p.
- FISSETTE Denis et POIRIER Pierre, *Philosophie de l'esprit. État des lieux*, Paris, Vrin, 2000, 352 p.
- (textes réunis par), *Philosophie de l'esprit, vol. 1 : Psychologie du sens commun et sciences de l'esprit, vol. 2 : Problèmes et perspectives*, Paris, Vrin, 2002-2003, 384 p. chacun.
- FODOR Jerry, *The Language of Thought*, New York, Cromwell, 1975, 224 p.
- , *Representations. Philosophical Essays on the Foundations of Cognitive Science*, MIT Press, Cambridge (Mass.), 1981, 356 p.
- , *La Modularité de l'esprit*, tr. fr. de A. Gerschefeld, Paris, Minuit, 1986, 178 p.
- HAUGELAND John, *L'Esprit dans la machine : Fondements de l'intelligence artificielle*, tr. fr. de J. Henry, Paris, Odile Jacob, 1989, 256 p.
- HOUE O., KAYSER D., KOENIG O., PROUST J. et RASTIER F., *Vocabulaire des sciences cognitives*, Paris, Puf, 1998, 417 p.
- JEANNEROD Marc, *La Nature de l'esprit*, Paris, Odile Jacob, 2002, 256 p.
- JOHNSON-LAIRD Philip, *L'Ordinateur et l'esprit*, Paris, Odile Jacob, 1994, 472 p.
- MISSA Jean-Noël, *L'Esprit-cerveau. La philosophie de l'esprit à la lumière des neurosciences*, Paris, Vrin, 1993, 266 p.
- PÉLISSIER Aline et TÊTE Alain (dir.), *Sciences cognitives, textes fondateurs (1943-1950)*, Paris, Puf, 1995, 313 p.
- PYLYSHYN Zenon, *Computation and Cognition. Toward a Foundation for Cognitive Science*, Cambridge (Mass.)/London, MIT Press, 1984, 320 p.
- SEARLE John R., *La Redécouverte de l'esprit*, tr. fr. Cl. Tiercelin, Paris, Gallimard, 1995, 353 p.

Introduction : cognitivisme et sciences cognitives

TIBERGHIEU Guy (dir.), *Dictionnaire des sciences cognitives*, Paris, Armand Colin, 2002, 336 p.

VARELA F., THOMPSON E. et ROSCH E., *L'Inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, tr. fr. de V. Havelange, Paris, Seuil, 1993, 377 p.

VIGNAUX Georges, *Les Sciences cognitives. Une introduction*, Paris, La Découverte, 1991, 359 p.