

Vincent RIALLE*

IA et sujet humain : entre *physis* et *sémiosis*

« Comme s'il n'y avait pas un moment où la quête de fondement parvient à un terme. Mais ce terme, ce n'est pas la présupposition non fondée, c'est la manière non fondée de procéder » (Wittgenstein, 1969, p. 52).

*« Nous sommes peut-être, quelque part, la pulsion profonde qui produit la *sémiosis*. Mais nous nous reconnaissons uniquement comme *sémiosis* en acte, système de signification et processus de communication. Seule la carte de la *sémiosis*, telle qu'elle se définit à un stade donné de l'aventure historique (avec les scories et les déchets de la *sémiosis* précédente qu'elle traîne derrière elle), nous dit qui nous sommes et ce que (ou comment) nous pensons. » (Eco, 1988, p. 61).*

L'interprétation étant nécessairement celle d'un sujet et l'interaction sémiotique se déployant *entre* des sujets humains, le propos de cet article est de donner quelques éléments de précision sur ce que l'on peut entendre par *sujet*. Plus précisément, entre une cognition structurée comme un système de symboles (cognitivism) et une cognition émergeant d'une organisation distribuée et auto-adaptative d'agents actifs ou réactifs (connexionisme, IA distribuée), la question se pose de savoir s'il y aurait place pour une cognition complexe se construisant par la relation de deux espaces aussi distincts qu'indissociables : celui du sujet, dont le substrat est

* Université Joseph FOURIER, Laboratoire TIMC-IMAG UMR CNRS 5525, Faculté de Médecine de Grenoble, 38706 La Tronche Cedex, Vincent.Rialle@imag.fr

le continuum corps-cerveau-esprit, et celui *des* sujets, dont le substrat est la *sémiosis*, l'espace socio-culturel des systèmes de signes. En outre, la compréhension de la conscience, ramenée dans le champ de la recherche scientifique par des travaux de plus en plus nombreux, ne peut être dissociée de celle du sujet de cette conscience, bien que les deux termes, sujet et conscience, soient à distinguer clairement. Je développerai l'hypothèse selon laquelle le sujet est le produit des langues et des civilisations autant que d'un appareil sensori-moteur très développé. Bien que la culture soit nécessairement *in fine* le résultat d'un processus naturel, la part culturelle de l'entour de l'enfant est spécifique de la construction du sujet humain. Le corps cognitif n'acquiert l'intelligence discursive — propre à l'homme — qu'en vertu d'un entour sémiotique qu'il rencontre et doit apprendre à sa naissance **pour se constituer en sujet**, et qu'il contribuera en retour à construire. Cette hypothèse, pour large qu'elle soit, ne devrait pas freiner le travail du modélisateur : penser un modèle général, heuristique, qualitatif et aux contours nécessairement non définitifs, apparaît aujourd'hui possible et utile.

Cet article reprend dans une première partie la réfutation des hypothèses fondamentales du cognitivisme — développée par l'IA classique — dans son entreprise d'explication de l'intelligence humaine. Elle aborde cette réfutation sous l'angle d'une confrontation des thèmes développés par A. Newell avec deux approches non dualistes de l'esprit issues, pour la première de la neuropsychologie (travaux de A. Damasio), pour la seconde de la philosophie de l'esprit (travaux de D. Dennett). Cette première partie est une manière d'introduire la notion de sujet de la conscience, développée dans la seconde partie de l'article. Celle-ci sera consacrée à un cheminement théorique visant à établir la double détermination de l'intelligence du sujet par le biologique et par le culturel, et à montrer que le sujet est indéfinissable au moyen d'un seul des deux déterminants. Cette démarche serait peut-être inutile, compte tenu du nombre déjà important de travaux sur ce thème, si elle ne s'adressait en priorité à une communauté de l'IA confrontée à l'objectif de modéliser l'intelligence, la rationalité, l'*esprit* humains par le biais d'une robotique avancée. Elle vise alors à rendre à la conceptualisation de l'intelligence humaine les dimensions fondamentales que le cognitivisme conçoit encore difficilement, comme celles de l'échange sémiotique, du temps, des moteurs de diversité et de la métaphore.

Mots-clés : sujet, subjectivité, intelligence, conscience, *sémiosis*, modélisation, intelligence artificielle, culture.

AI and the human subject. On one hand, the activity of interpretation is necessarily performed by an individual subject (for a particular semiotic unit there may be as many nuances of interpretation as there are subjects); on the other hand, semiotic interaction takes place as a relation *between* human subjects. The purpose of this article is to elucidate a meaning of the term "subject" which takes both of these facets into account. Traditionally, cognition is conceived either as a symbol system (cognitivism), or as the result of an adaptive distributed organisation of active or reactive agents (connexionism, distributed-AI, multi-agent systems). This paper seeks an alternative way to model complex cognition, conceived as the product of the relation between two different though inseparable spaces. The first space is that of the *individual subject* ; its substrate is the body-brain-mind continuum, or *physis* . The second space is that of *plural subjects* ; its substrate is the socio-cultural space of systems of signs, or *semiosis*. The notions of subject and consciousness must be clearly distinguished, but the study of consciousness cannot be dissociated from that of the human subject of consciousness. This leads to the hypothesis that the human subject is a product of language and civilisation as well as of biological functioning. Although in the last resort culture is a product of nature, the cultural aspect of the environment plays a specific role in the construction of the human subject. The cognitive body acquires discursive intelligence, specific to the human kind, only by means of a semiotic environment encountered at birth. In order to attain the status of a subject, the child must appropriate this environment and at the same time contribute to its construction. Although very broad, this hypothesis need not hinder modelling : it appears both possible and necessary today to design a general, heuristic, qualitative model.

The first part of this article reviews the refutation of the fundamental hypotheses of cognitivism and «classical» AI in their attempt to explain human intelligence. This refutation is undertaken by confronting the topics developed by Newell with two non-dualist approaches to mind : one from neuropsychology (Damasio), the other from the philosophy of mind (Dennett). This first part introduces the notion of the subject of consciousness. The second part consists of a theoretical development which aims at establishing the dual determination of the subject's intelligence by both biological and cultural dimensions : the subject cannot be defined by either of these two dimensions alone. These themes are already well worked ; however, this article addresses that part of the AI community which aims at modelling human intelligence, rationality and *mind* by means of advanced robotics. In this context, it seems important to stress certain fundamental dimensions of human intelligence which the cognitivist paradigm

has difficulty taking into account, such as semiotic exchange, time, the generation of diversity, and metaphor.

Key Words : Subject, Subjectivity, Intelligence, Consciousness, Semiosis, Modelling, Artificial Intelligence, Culture.

INTRODUCTION

Quelle direction de recherche l'intelligence artificielle devrait-elle ajouter à ses — déjà nombreux — domaines d'investigation pour répondre à cette question en forme de koan zen : «l'intelligence artificielle est-elle possible ?»¹ «La direction du sémiotique» répondrons-nous sans plus de détours. Cependant, la réponse n'a que peu d'intérêt en regard de la réflexion qui y conduit. C'est à cette réflexion que nous nous attacherons dans cet article.

L'IA a connu quelque succès à rendre compte d'un côté de l'intelligence de l'expert en situation de résolution de problèmes ou de planification, de l'autre de processus associatifs, de reconnaissance d'objets et de catégorisation plus propres au connexionnisme. Parviendra-t-elle à franchir le pas du langage, et rendre compte d'une intelligence *discursive*² qui se développe par l'usage d'une langue et de nombreux systèmes de symboles dans des réseaux complexes d'échanges entre les personnes d'une société humaine³ ?

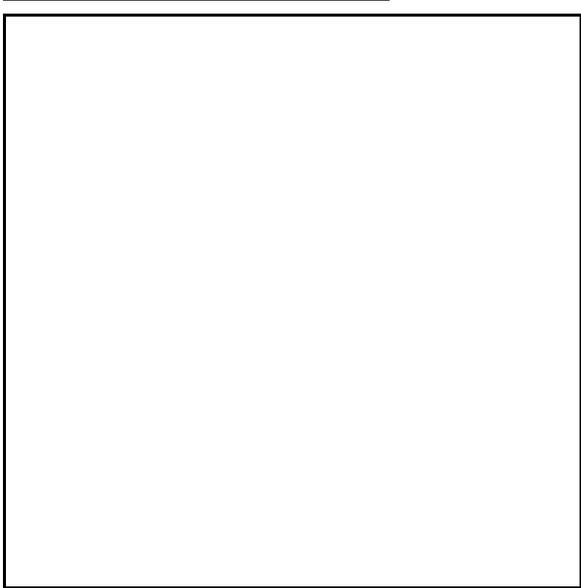
¹ D. Hofstadter (éminent amateur de koans) : «Historiquement, les gens ont toujours eu des idées naïves au sujet des qualités qui, une fois mécanisées, constitueraient indéniablement l'intelligence. Il semble parfois qu'au lieu de produire un résultat qui, de l'avis de tous, est l'intelligence réelle, chaque nouvelle étape vers l'IA dévoile uniquement ce que l'intelligence *n'est pas*. » (Hofstadter et Dennett, 1987, p. 642).

² Sur l'intelligence spécifiquement *discursive* et les moyens de l'aborder par l'IA, on lira particulièrement Jorion (1990).

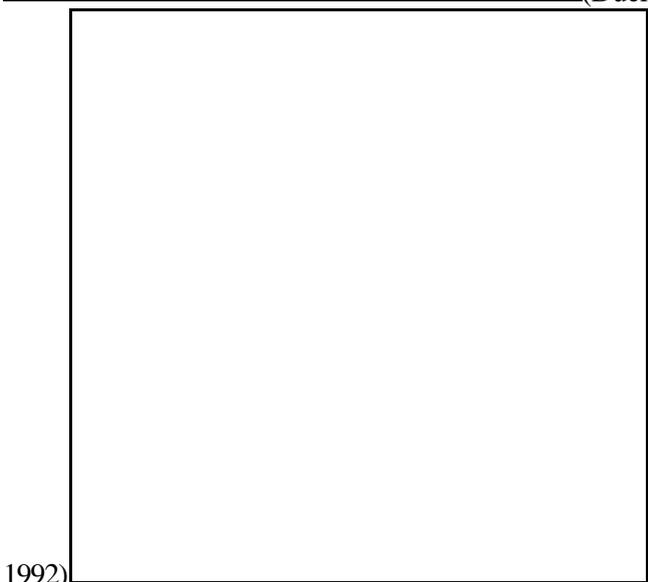
³ Il s'agit de l'explication sociogénétique du sujet de la connaissance, que Piaget reconnaissait au même titre que l'explication phylogénétique et l'explication psychogénétique, tout en ayant consacré son œuvre à cette dernière

En retour, de même que le cognitivisme et le connexionnisme ne sauraient, pour rendre compte du cognitif, faire l'économie d'une science des signes, une sémiotique saurait-elle, pour intervenir dans les recherches sur la cognition, faire l'économie d'une étude des liaisons entre son domaine et celui de la neuropsychologie dans lequel s'inscrit l'interprète ?

Cet article tente d'apporter quelques éléments pour éclairer cette dimension de la cognition qui fait du sujet humain à la fois la brique de base et l'aboutissant d'une intelligence discursive. Le propre de l'intelligence discursive est de ne pouvoir se concevoir que dans une

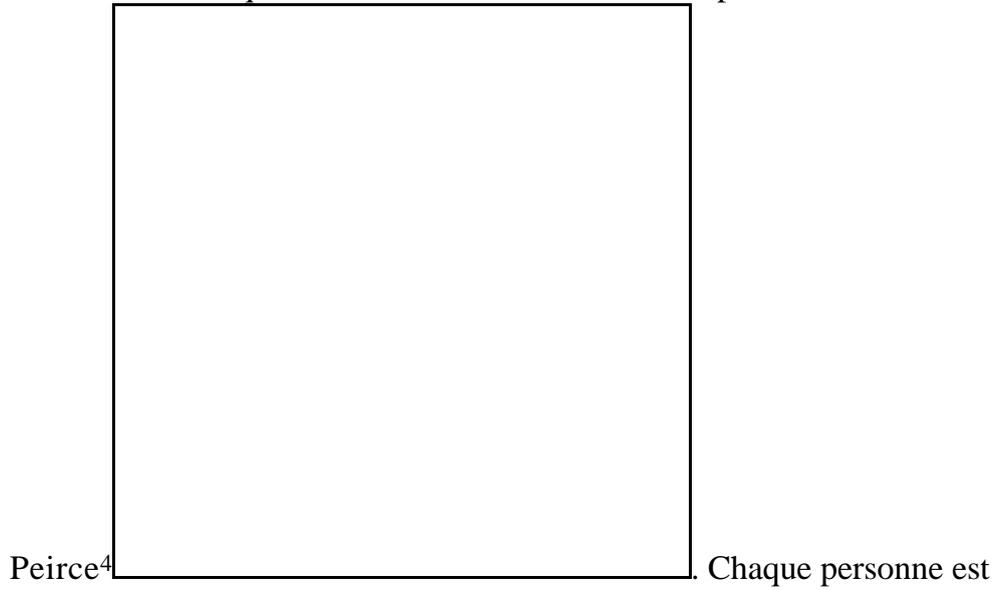


(Ducret, 1984 ; Piaget,



1992)

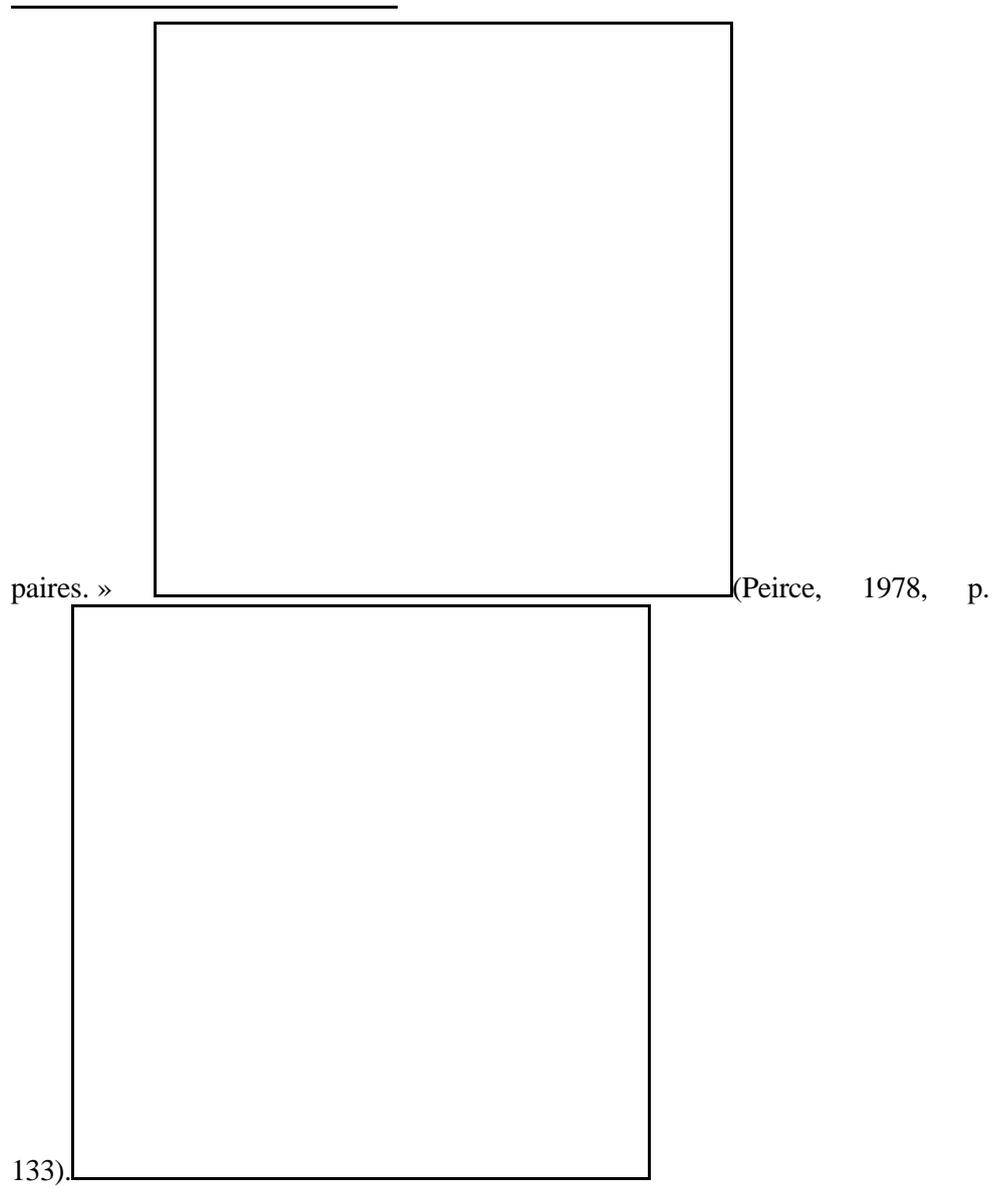
relation : elle se déploie chez le sujet dans une relation entre la nature (grec : *physis*) et l'espace de tous les systèmes sémiotiques propres à une société, qui entoure l'enfant d'homme à sa naissance et que nous nommons *espace sémiotique* ou *sémiosis*, en référence à la *sémiosis* telle que définie par C.S.



⁴ Le concept de *sémiosis* est introduit par Locke (cf. Rastier, cet ouvrage) et est repris par C.S. Peirce qui l'utilise pour qualifier le procès de signification : « (...) par « semiosis », j'entends (...) une action ou influence qui est ou implique la coopération de *trois* sujets, tel qu'un signe, son objet et son interprétant, cette influence tri-relative n'étant en aucune façon réductible à des actions entre

simultanément une instance partielle et une source de production de cette intelligence. Cependant, alors qu'une intelligence artificielle non focalisée sur le sujet et libre des *a priori* transcendants du cognitivisme se développe dans la voie de l'herméneutique matérielle (e.g., Bachimont, 1993, 1996, Rastier, 1995), nous aborderons ici la question du sujet en tant que *nœud* entre le biologique et le social.

Nous tenterons d'étayer l'idée selon laquelle la dimension sémiotique de l'IA passe par la modélisation de cette intelligence discursive, avec un point particulier à souligner : celui de la



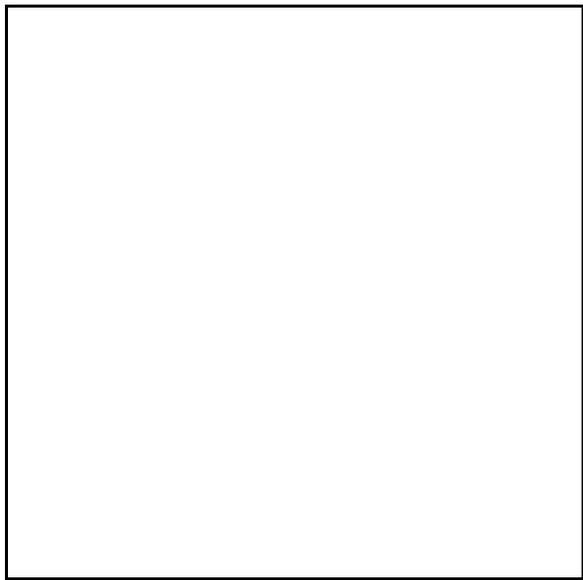
conscience, de la subjectivité du sujet, de son intentionnalité⁵ et de leur collusion étroite avec la langue⁶ en tant que système sémiotique de base⁷. Point n'est besoin cependant de tenter de produire un modèle complet, dont il serait ensuite aisé de démontrer les insuffisances. Nous nous limiterons à étudier les conditions d'une approche théorique de l'intelligence qui n'exclut ni le développement neurobiologique et psychocognitif du sujet, ni l'environnement socio-culturel et sémiotique. Cet environnement *conditionne* ce développement au niveau *du sujet* tandis qu'il *émerge* de l'activité *des sujets* (nous n'avons pas de verbe pour exprimer cette action complexe).

Quant à la théorisation proprement dite du *sujet* de la conscience humaine (indistinguable *a priori* du *sujet* de l'intelligence discursive), nous nous limiterons à exprimer la conviction qu'elle ne saurait être donnée par une axiomatique directe des lois d'essence (celles notamment de la phénoménologie depuis Husserl), mais plutôt par une médiation herméneutique intrinsèque à l'élaboration théorique (un exemple de cette attitude épistémique est donné par la mathématisation des lois d'essence de J. Petitot

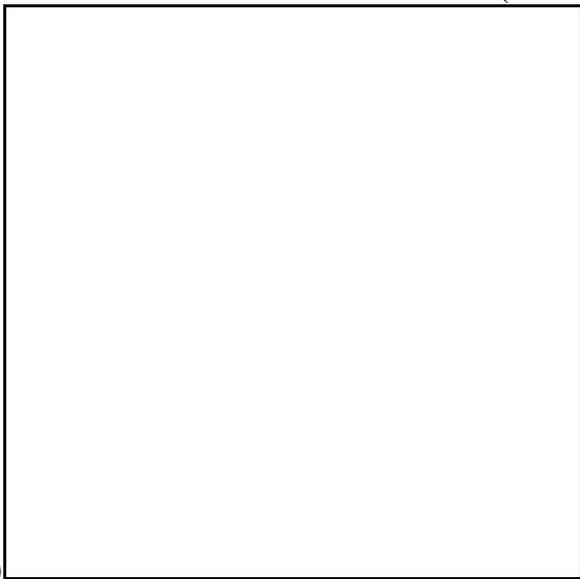
⁵ Convenons avec Searle que «(...) la conscience et l'intentionnalité sont liées entre elles de manière essentielle au sens où nous ne comprenons la notion d'état intentionnel inconscient que dans les termes de son accessibilité à la conscience. » (Searle, 1995, p. 14).

⁶ Cf. Rastier (1996) : «les langues (et les autres systèmes de signes) jouent un rôle de médiateur dans la genèse de la conscience humaine».

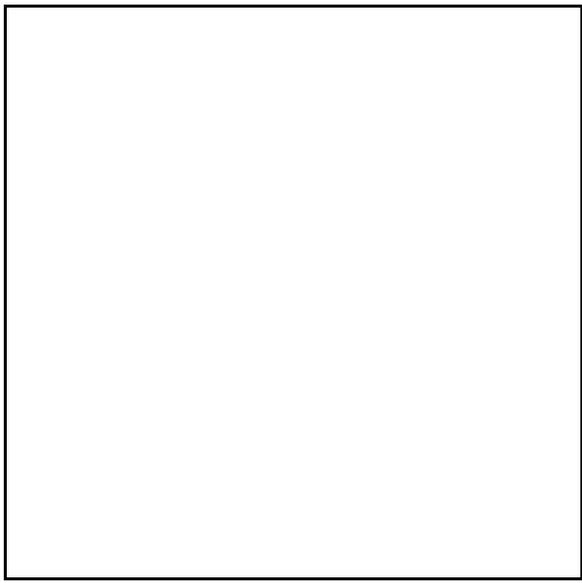
⁷ « Puisque l'on considère de plus en plus (...) que la langue est non seulement le système sémiotique le plus ou le mieux analysable, mais aussi celui qui peut modéliser tous les autres, en transformant n'importe quelle autre sémiotique au niveau de son contenu, alors le modèle du signe linguistique s'impose graduellement comme le modèle sémiotique par excellence. » (Eco, 1988, p. 44).



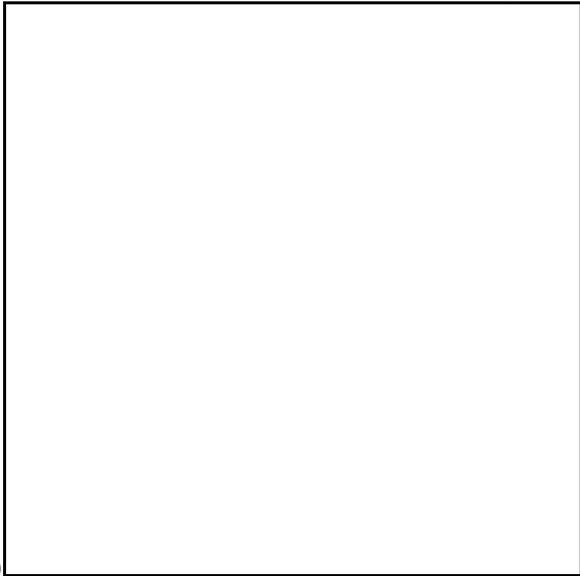
(Petitot,



1993). En particulier, si l'élaboration philosophique du sujet proposée par l'*être-au-monde* heideggerien (Dreyfus, 1993) a pu s'avérer utile — face notamment au positivisme logique et jusque dans la conception de logiciels



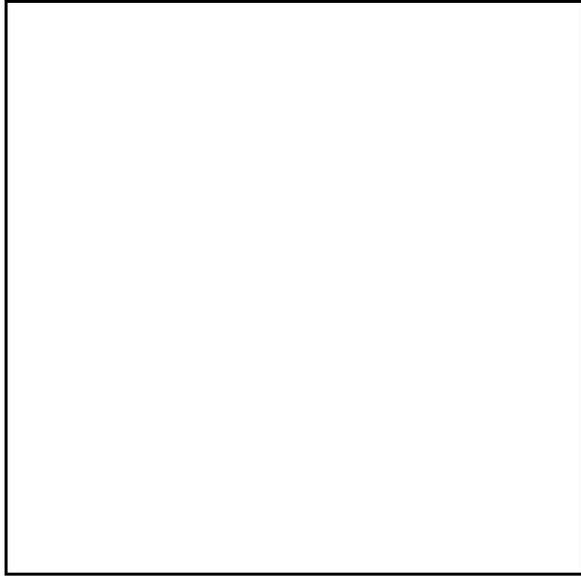
(Winograd,



1993) —, une telle élaboration n'est pas affranchie du risque de constituer une nouvelle axiomatique — incompatible en particulier avec un soucis de dépassement des barrières interprétatives —, et ne dispense pas du travail d'analyse des tenants et aboutissants de cet être-au-monde, lesquels n'apparaissent en rien transcendants.

Une dualité semble s'imposer entre le niveau neuronal physico-chimique, lieu d'ancrage du sujet, et le niveau sémio-culturel. Mais dualité analytique n'est pas synonyme de dualisme cartésien, et la raison, utilisée dans cette démarche, n'a pas d'autre forme de

transcendance que celle définie par Putnam⁸ ou vue par M. Serres⁹. Cette hypothèse, pour large qu'elle soit, ne devrait pas freiner le travail du modélisateur : penser un modèle général, heuristique, qualitatif et aux contours nécessairement non définitifs, apparaît aujourd'hui possible et utile, à la lueur notamment des approches récentes de la conscience. Celles-ci seront abordées, dans une première partie de l'article, au moyen d'une confrontation entre l'approche cognitive d'Allen Newell (Newell, 1980 ; Newell, 1982 ; Newell *et al.*, 1989 ; Newell, 1990 ; Newell, 1993) et les approches philosophiques de Daniel

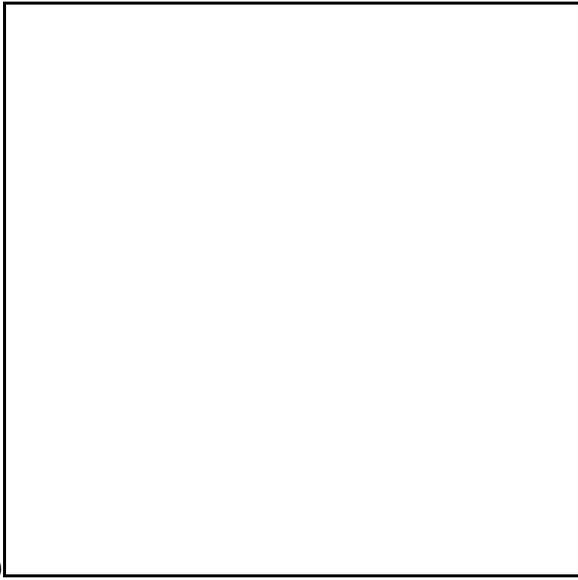


Dennett

(Dennett, 1993 ;

⁸ « Mon idée est que les « règles en vigueur » (*standard*) que retiennent, explicitement ou implicitement, une culture ou une sous-culture, ne peuvent pas *définir* ce qu'est la raison, pas même une raison en contexte, dans la mesure où, pour leur interprétation, elles présupposent cette raison (elles présupposent une « disposition raisonnable » [*reasonableness*]). (...) En ce sens, la raison est à la fois immanente (puisqu'on ne saurait la trouver hors d'institutions et de jeux de langage concrets) et transcendante (puisqu'elle est une idée régulatrice que nous utilisons afin de critiquer la conduite de *toute* activité et de *toute* institution. » (Putnam, 1992, p. 23).

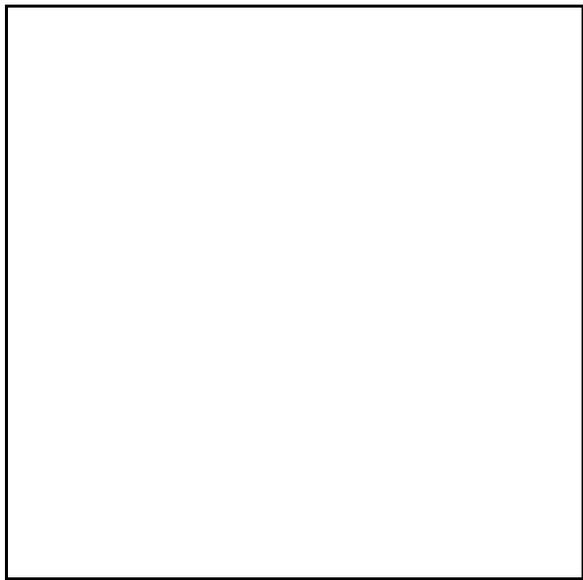
⁹ « Qu'appelle-t-on donc penser ? Compenser ce qui n'est pas au moyen de la raison, porter la tare rationnelle entre l'existence et le néant ou le possible, comme si la raison faisait le rapport de l'être au non être, ou qu'elle justifiât ce qui est à partir de ce qui n'est pas. Elle touche donc à la création quasi divine, et suppose une familiarité mortelle avec le néant ou le possible. Cette pensée rationnelle, cette pesée ou proportion compensatoire comble exactement le manque ontologique. » (Serres, 1991, p. 210).



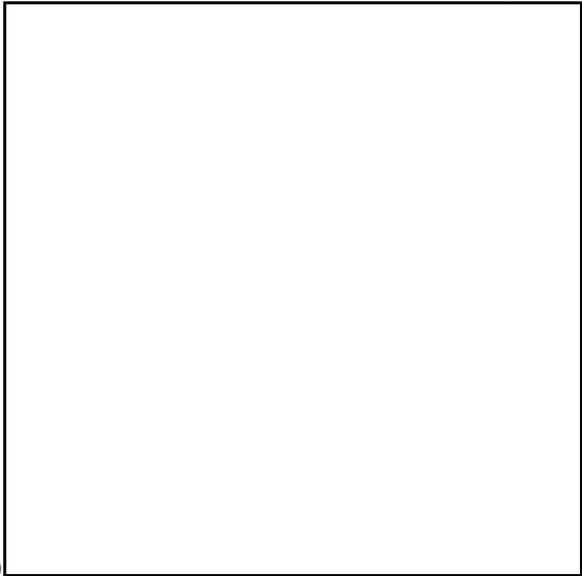
Dennett, 1993)
neuropsychologique

d'Antonio

et
Damasio

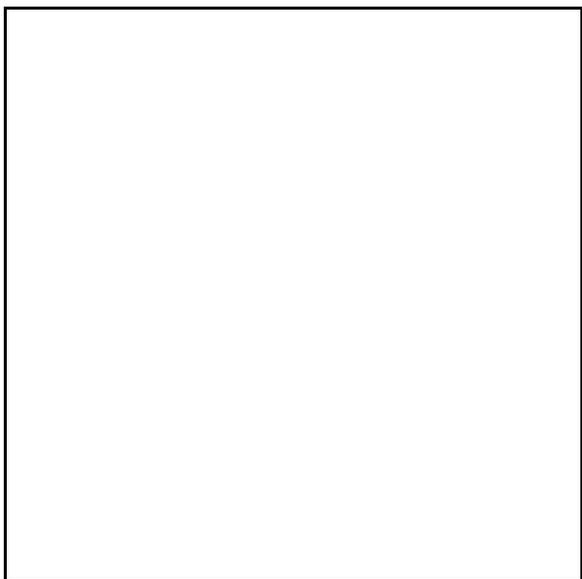


(Damasio,

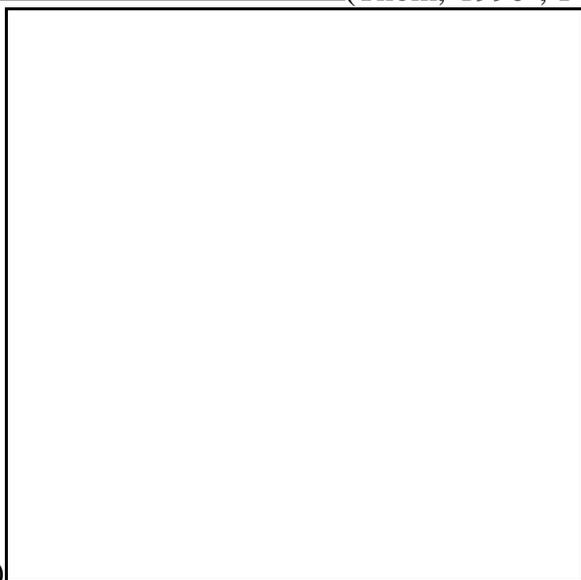


1995)

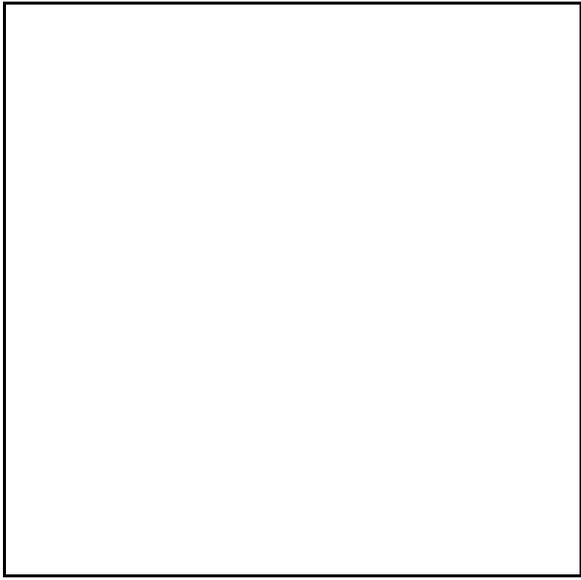
Ce choix de travaux n'est pas représentatif de l'ensemble des recherches touchant à la cognition et au langage. En particulier, d'importants travaux sur l'ancrage perceptif du langage et sa dynamique formatrice du sujet ont vu le jour dans la perspective morphodynamique de la théorie des catastrophes, introduite par Thom et Zeeman



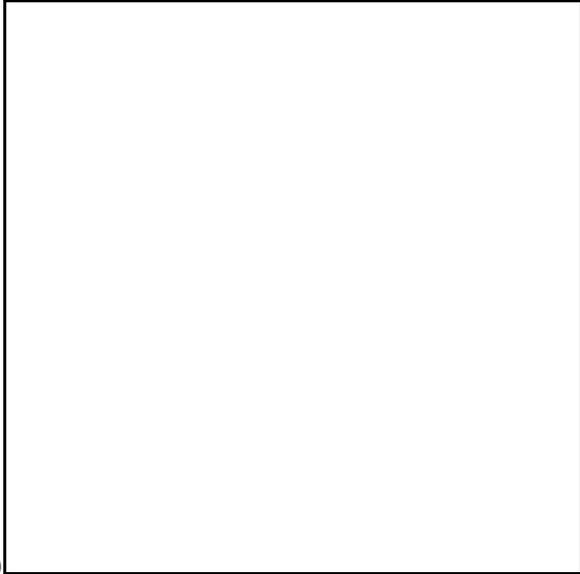
(Thom, 1990 ; Petitot, 1992



; Virole, 1995). De plus, l'espace sémiotique dont il sera question en particulier dans la seconde partie apparaît aujourd'hui chez plusieurs auteurs comme partie intégrante de l'espace complet de déploiement de la cognition humaine. Miermont, par exemple, définit un *espace sémiotique implicite* dans lequel se déploie le sujet à partir de sa naissance

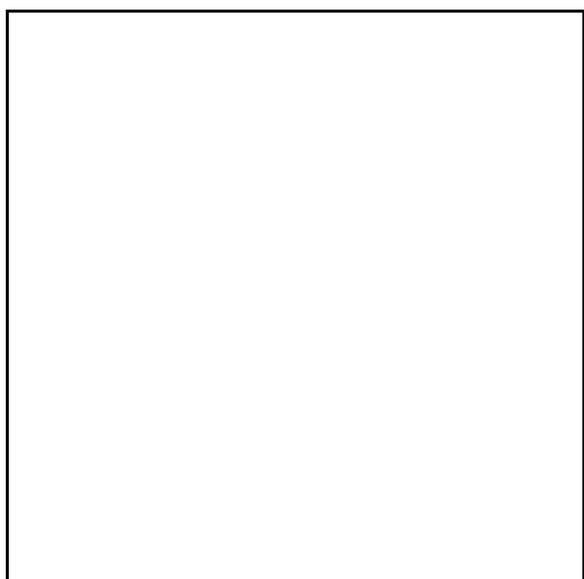


(Miermont,

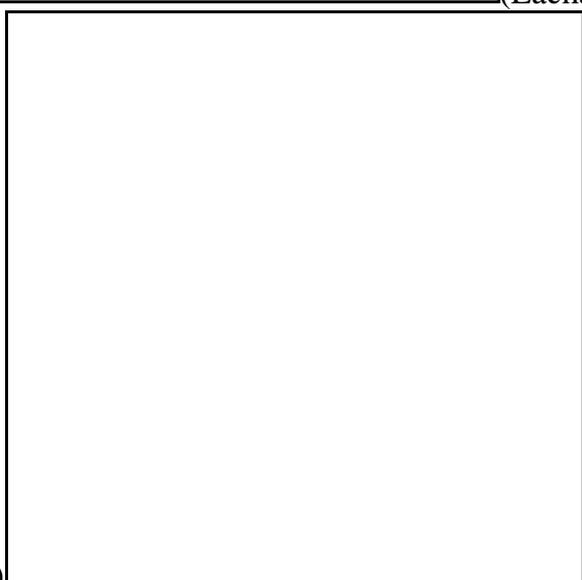


1996)

. D'autre part, dans le contexte sémiotique actuel, l'ordinateur tient, selon Miermont, une place à la fois comme outil de compréhension et comme créateur de *sémiosis*. N. Lacharité de son côté clarifie, au sein d'un modèle complet de *système interprétant*, ce qu'il nomme les trois niveaux de complexité : celui des *synthèses perceptives*, celui des *processus sémiotiques* et celui des *synthèses logiques*



(Lacharité,



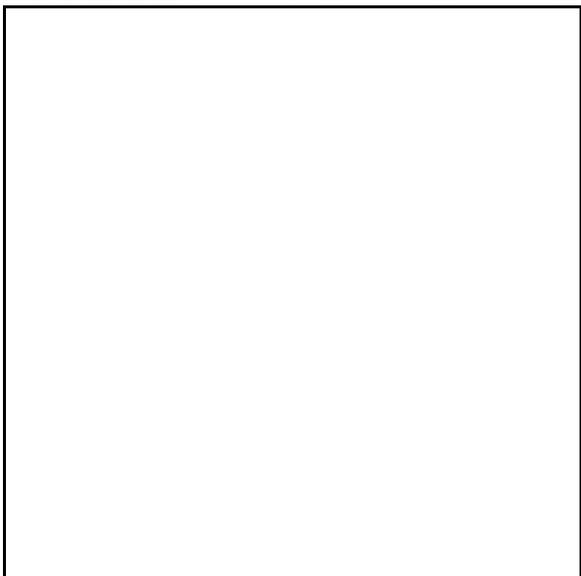
1996). La particularité de ce modèle est une grande précision dans la définition des constituants du système interprétant et des relations entre ces constituants, sans faire pour autant de concessions devant la complexité de ces relations. Enfin, la psychanalyse, enfant terrible de la science du XXe siècle, s'intéresse de très près à la logique des symboles dans la «machine désirante» du sujet humain, qu'elle distingue de la «machine connaissante» propre aux sciences cognitives actuelles¹⁰.

¹⁰ «(...) les machines connaissantes ne sont pas forcément des machines désirantes, les théories de la connaissance, ou d'ailleurs de la signification, ne sont pas les mêmes que celles du désir, en d'autres termes, la psychose ne saurait se réduire à un simple défaut de logiciel, pas plus que le désir ne saurait se réduire à

Le choix de nos trois auteurs «classiques» n'est donc qu'un parcours possible, certes dans un paysage de figures emblématiques, parmi ceux que l'on peut faire aujourd'hui dans l'immense champ de la littérature consacrée à ce sujet ces dernières années. Ce parcours n'est pas neutre et semble céder en particulier au champ de forces déterminé aujourd'hui par les fonctionnalismes et la philosophie analytique nord-américaine, mais dans une optique de dépassement des apories d'une approche purement fonctionnaliste¹¹.

une combinatoire mathématique (...). Ce constat invite alors à interroger à nouveaux frais ce que doit être une « machine désirante » ou au moins une machine douée d'intentionnalité. C'est là une question vers laquelle la psychanalyse toute entière est tournée » (Vincent, 1996, p. 197-198).

¹¹ Voir en particulier le n°21 d'*Intellectica*, consacré à une étude approfondie des fonctionnalismes sous la direction d'E. Pacherie



(Pacherie,

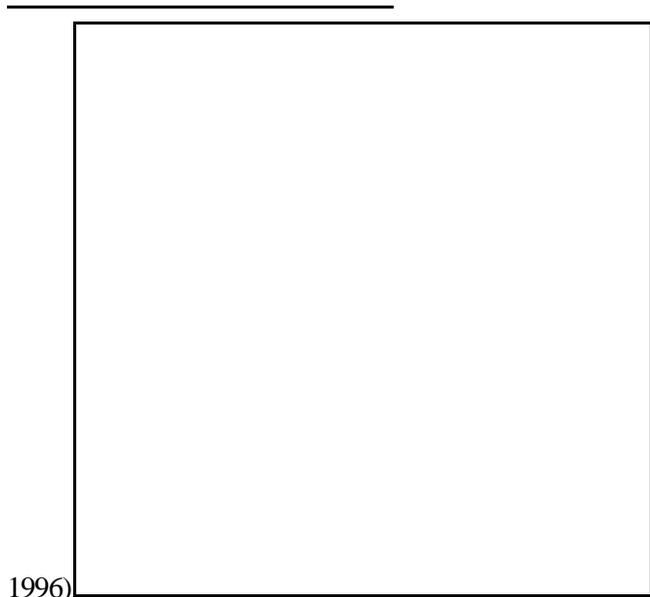
1996) et en particulier la critique

d'Y.-M. Visetti (Visetti,

La seconde partie sera consacrée à un cheminement théorique visant à établir la double détermination de l'intelligence du sujet par le biologique *et* par le culturel. Cette démarche serait peut-être inutile, compte tenu du nombre déjà important de travaux sur ce thème, si elle ne s'adressait en priorité à une communauté de l'IA confrontée à l'objectif de modéliser l'intelligence, la rationalité, l'*esprit* humains par le biais d'une robotique avancée. Elle vise alors à rendre à la conceptualisation de l'intelligence humaine des dimensions fondamentales qui restent encore peu considérées par le cognitivisme, telles que celles de l'échange sémiotique, du temps irréversible, des moteurs de diversité, du processus métaphorique...

1. LE COGNITIVISTE, LE PHILOSOPHE ET LE NEUROLOGUE

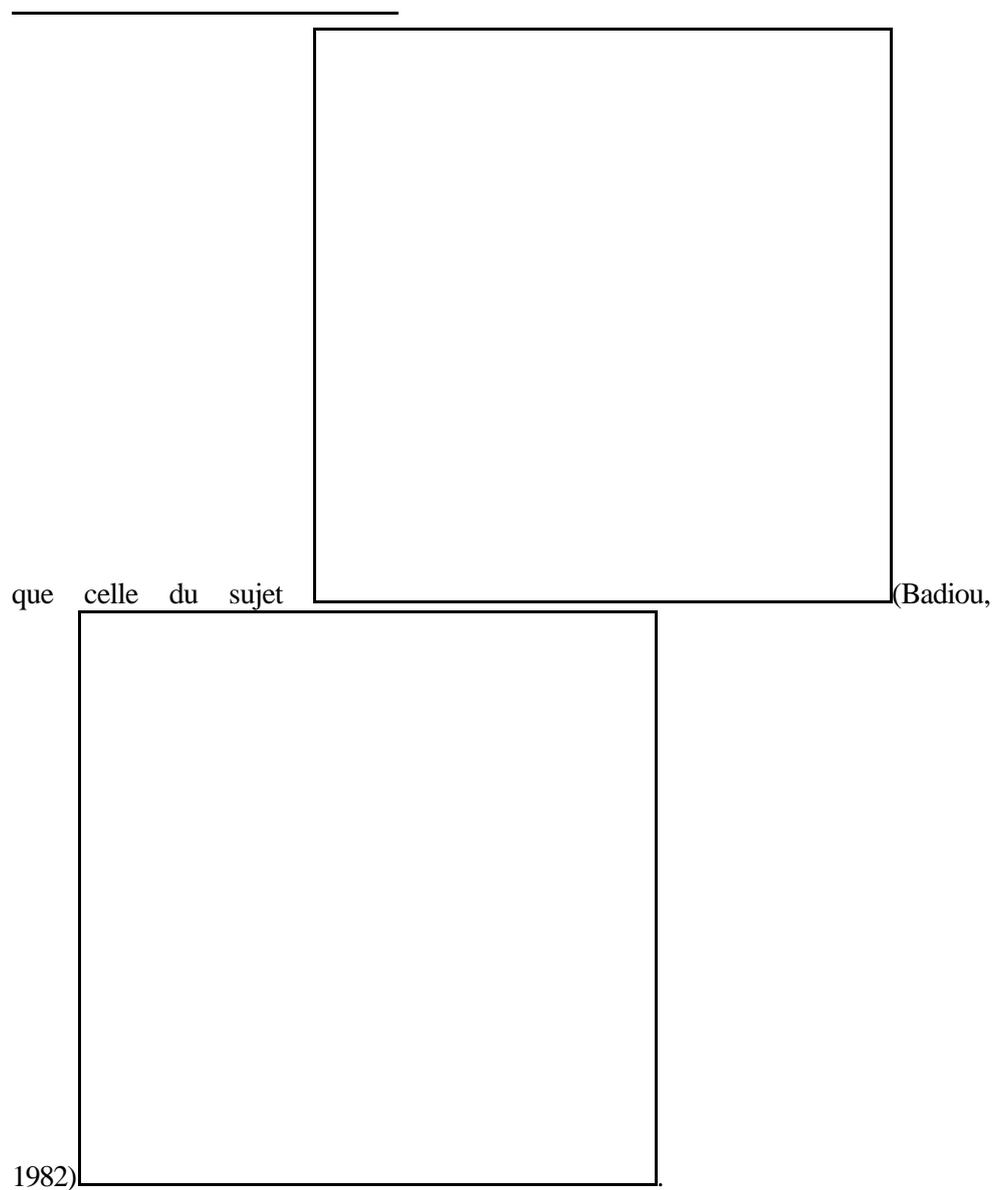
La question du sujet de l'interprétation engage directement celle de la conscience qui, après avoir perdu le devant de la scène en psychologie¹², depuis le déclin de l'introspectionnisme du XIXe siècle,

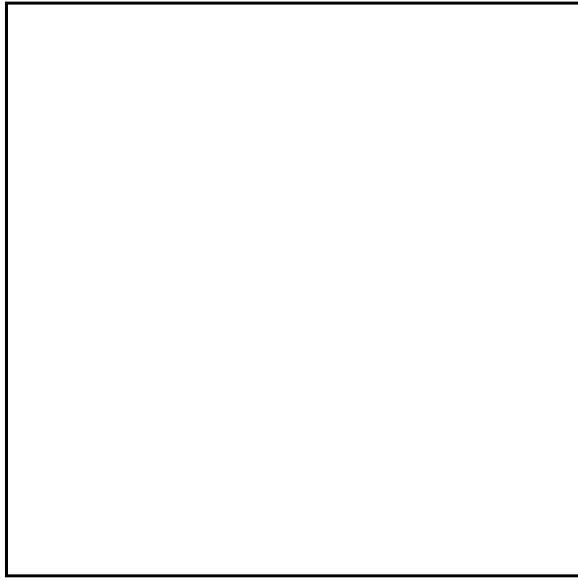


1996)

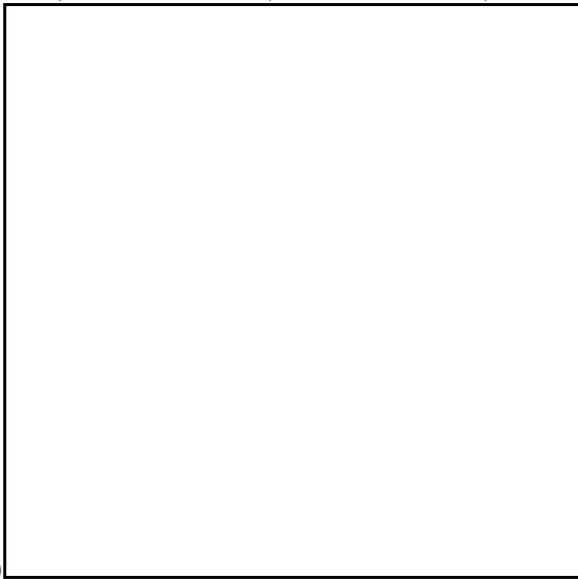
¹² Il n'en va pas de même pour la philosophie qui a particulièrement étudié la question de la conscience, dans la courant de la phénoménologie notamment, ainsi

s'est vue posée avec une insistance presque surprenante à partir des

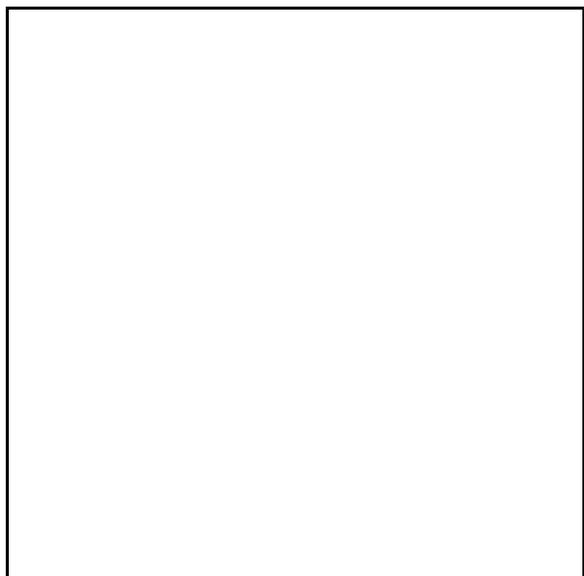




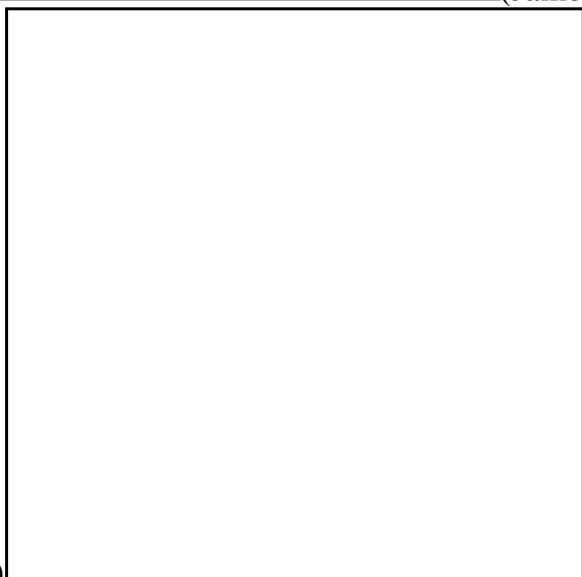
années 80 (e.g., (Jackendoff, 1987 ; Searle, 1987 ; Edelman, 1992 ; Dennett, 1993 ; Varela *et al.*, 1993 ; Verschure, 1993 ; Crick, 1994 ; Searle, 1995 ; Block *et al.*, 1996 ; Chalmers, 1996 ; Hameroff *et al.*,



1996), etc.). Nombre de chercheurs en sciences de la cognition ont donc subitement redoublé d'intérêt pour une question séculaire et notamment explicitement posée un siècle auparavant par William James



(James, 1890,

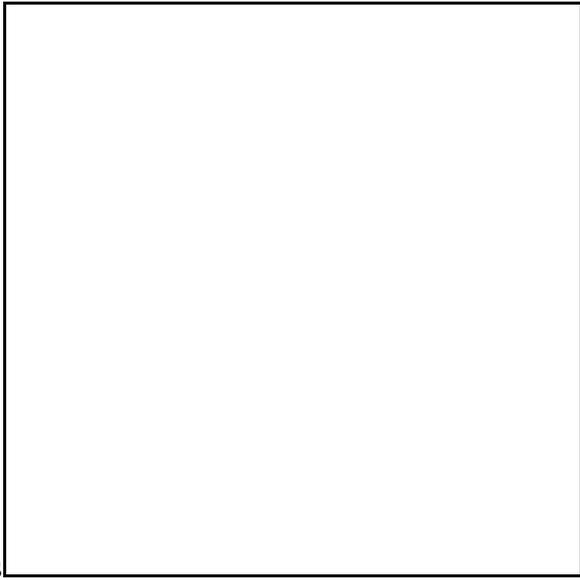


1981). Dans ce mouvement, plusieurs théories sont publiées aujourd'hui, qui toutes rencontrent et tentent de répondre au problème du *dualisme cartésien*.

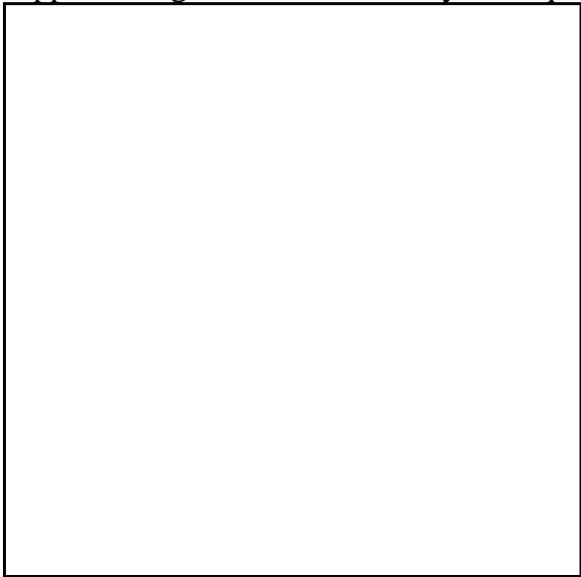
Parcourons rapidement, à travers les travaux de Newell, Dennett et A. Damasio, le paysage partiel des idées qui ont cours dans ce domaine.

I.1. IA «classique» et niveau intentionnel chez Newell

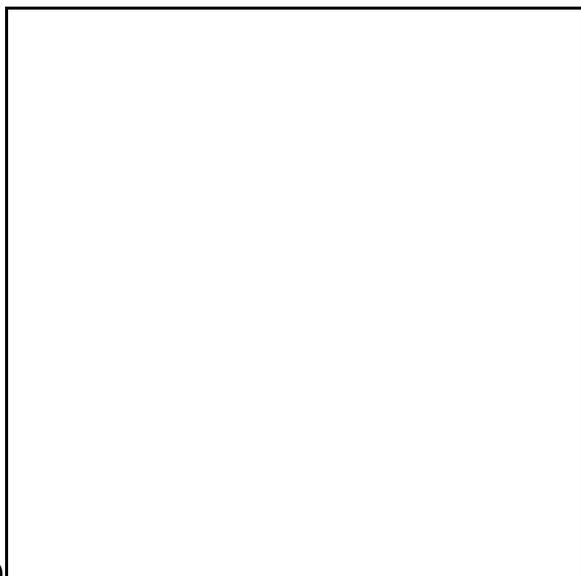
Il existe aujourd'hui une floraison de «communautés» en intelligence artificielle depuis son individualisation en tant que branche scientifique dans les années cinquante. Allen Newell, co-fondateur et parmi les plus illustres représentants de l'IA, en retient



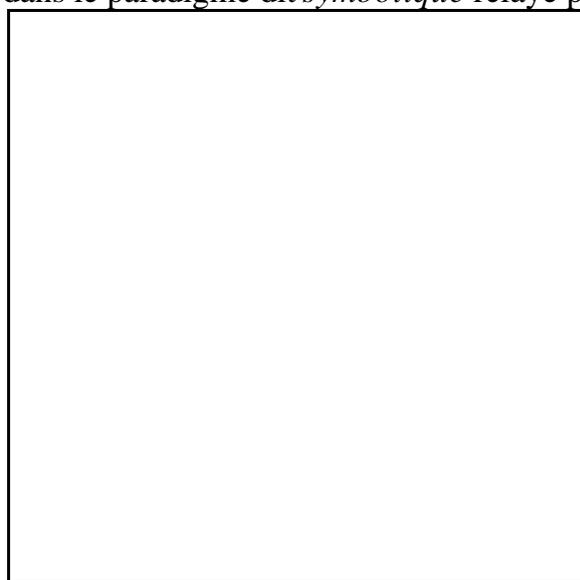
trois : la communauté des systèmes à bases de connaissances, celle des logiciens et celle de l'apprentissage symbolique automatique



(Newell,

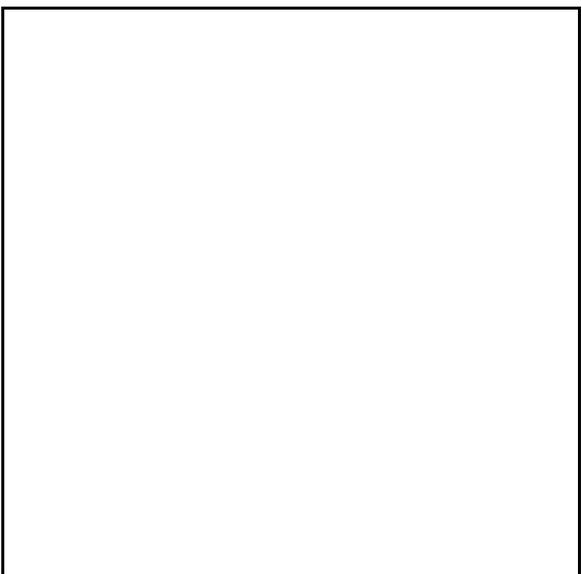


1993) Il s'avère vain de rechercher, parmi elles, une communauté consacrée à part entière à l'étude de l'intelligence discursive et du rôle du langage et des systèmes sémiotiques dans la construction de la cognition. Il existe néanmoins une branche de l'IA classique¹³ dévolue à la linguistique, mais celle-ci est largement incluse dans le paradigme dit *symbolique* relayé par celui

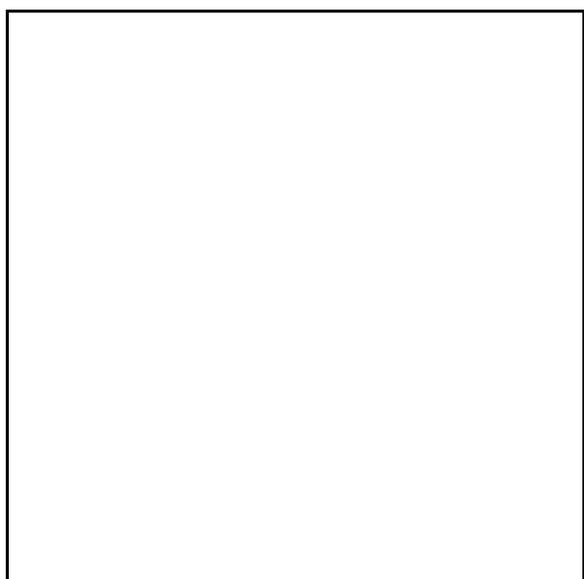


de l'IA distribuée (Sabah,

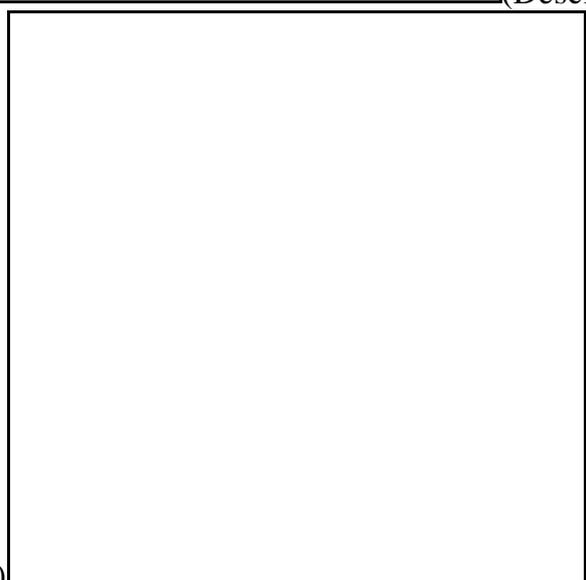
¹³ Le terme IA classique réfère à l'Intelligence Artificielle fondée sur le paradigme du Système de Symboles Physiques (Newell, 1980) et en opposition avec l'ensembles des approches constructivistes de l'intelligence au sens large.



1993) poussée des systèmes cognitifs, avec une étude formels



(Desclés,

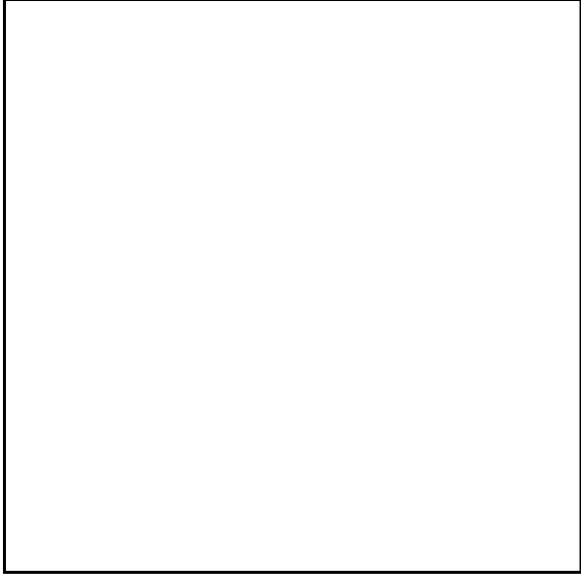


1991)

Nous retiendrons, pour l'analyse des conceptions cognitivistes des mécanismes de pensée, les travaux de Newell, dont l'objectif déclaré était l'étude de l'esprit humain¹⁴, en ce qu'ils rassemblent la plupart des arguments du paradigme cognitiviste, y compris ceux qui le protègent de tout procès expéditif. Conformément au programme originel de l'IA dont il est l'un des initiateurs, Newell s'est consacré à modéliser la complexité organisationnelle du traitement de l'information *chez l'homme* pour la simuler sur ordinateur. Il s'est attaché à développer un

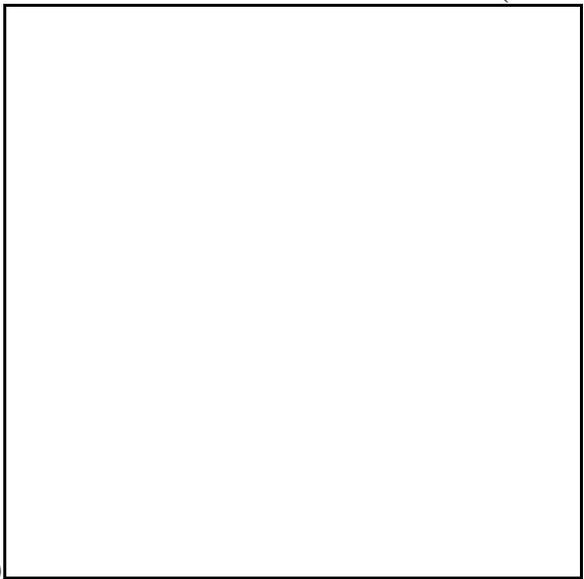
¹⁴ Son collaborateur de toujours, H. Simon, confirme : « (...) (Newell) described his career as aimed at understanding the human mind » (Simon, 1993, p. 251).

modèle d'organisation de la connaissance, interne au sujet humain, pour rendre compte d'un type particulier de complexité: celle, déductive et heuristique, des raisonnements de savants et d'experts, c'est-à-dire des raisonnements dépendant d'une connaissance initiale et poursuivant un but, tous deux clairement définis. Pour cela, il a cherché à montrer la nature et l'ampleur de cette complexité pour produire ensuite des moyens de la *penser*, d'où sa théorie du *knowledge level*



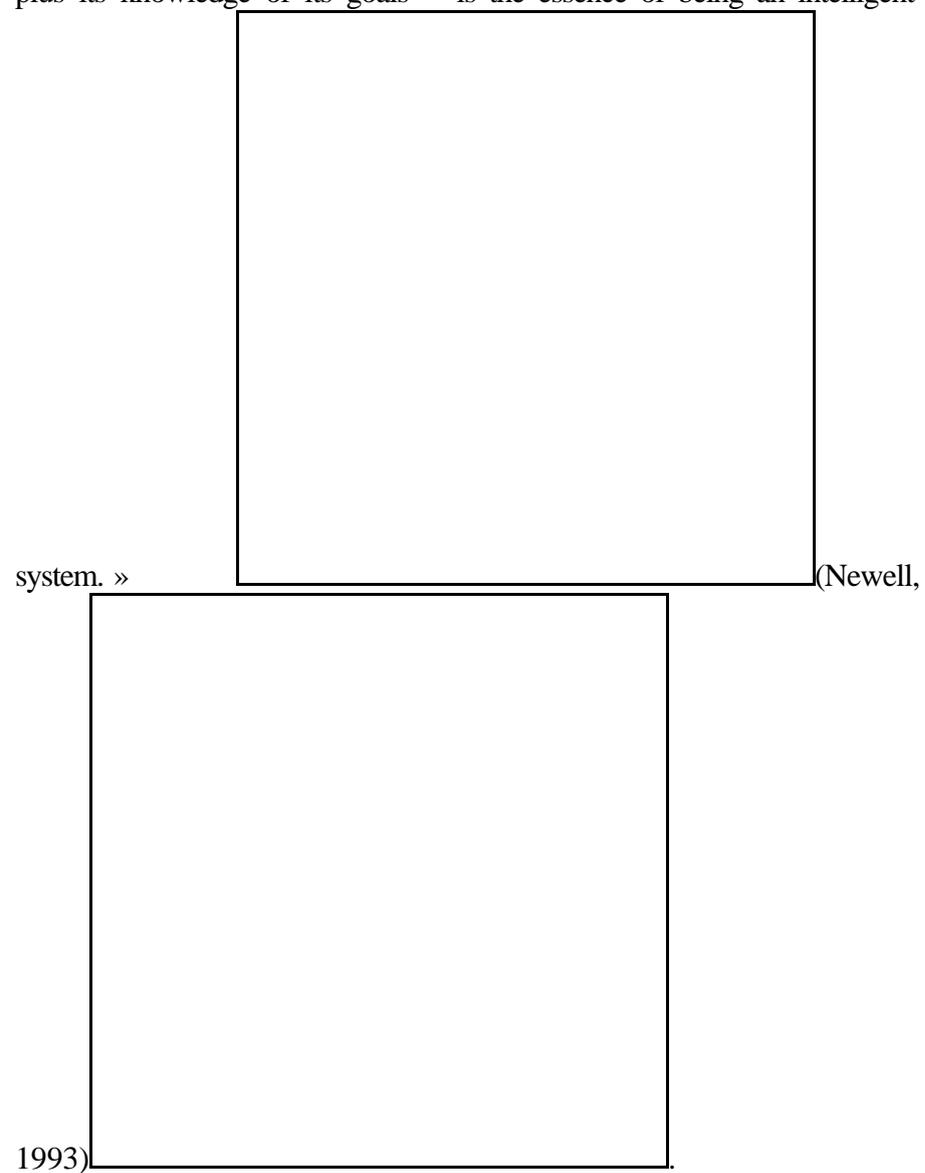
(Newell, 1982 ; Newell,

1993)

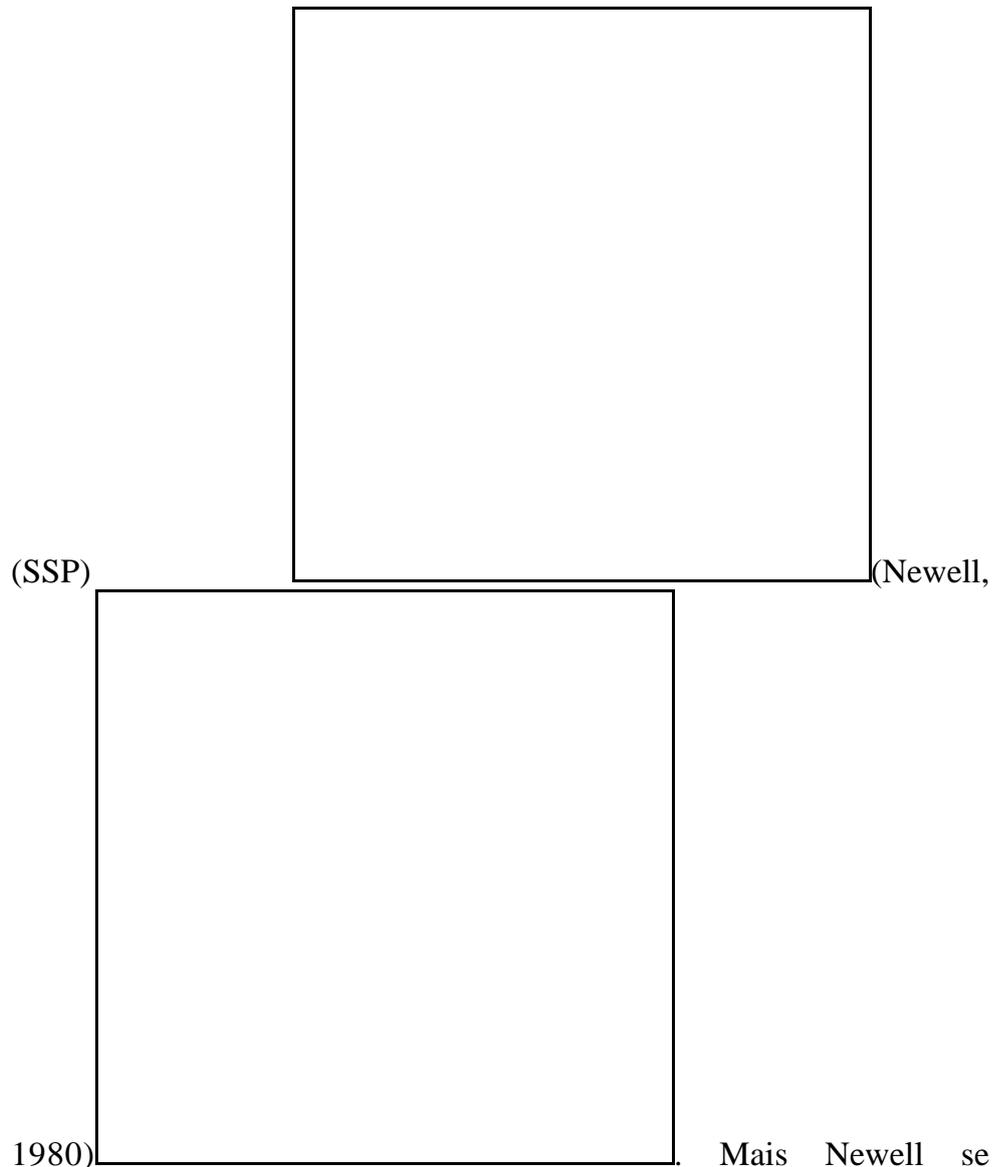


« “The knowledge level” (...) describes the nature of *knowledge* as the medium of a system level that lies above the *symbol* or *program level* in the hierarchy of levels of computational systems. The unique character of the knowledge level is that its central law of behavior, the principle that the

system will take whatever action attains its *goals* given its knowledge, can be computed without positing any internal structure of the system. Put another way, the knowledge level abstracts completely from representation, structure, and process. That this is possible-that a system's behavior can be predicted based only on the *content* of its representations plus its knowledge of its goals - is the essence of being an intelligent



Les travaux de Newell, comme l'ensemble des travaux de l'IA classique, reposent sur le paradigme du système de symboles physiques



Mais Newell se distingue néanmoins de toute acception simplificatrice du cognitivisme en ce qu'il a cherché à explorer le raisonnement humain dans ce qu'il a de plus spécifique, de plus existentiel ou de plus authentiquement humain si l'on peut dire, mais sans jamais se départir du paradigme de la représentation et de l'espace du problème¹⁵.

Il rejette notamment la critique visant à dire qu'un système de symboles est un système clos (limité par la *fermeture transitive* de ses

¹⁵ « (...) the key issue in the problem space is this creative active decision on the part of the agent. It's a really existential act. That's how it's going to represent its world and its problems. » (propos de Newell rapportés par Agre (1993, p. 447).

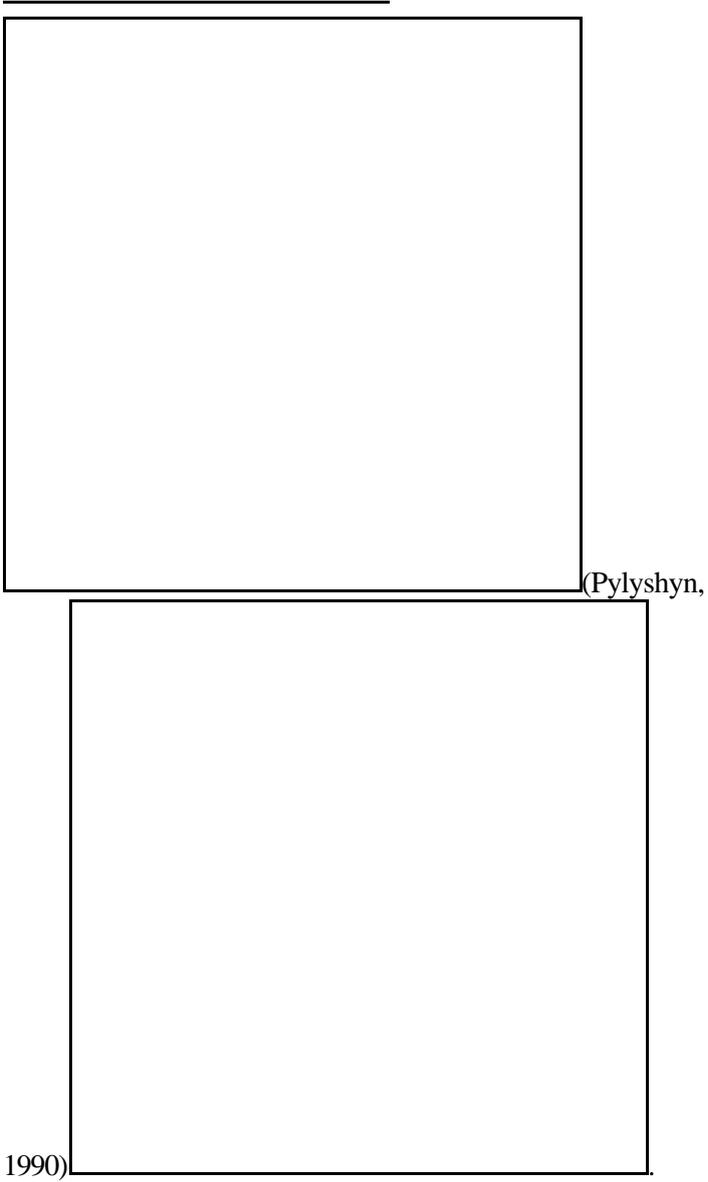
opérateurs) qui ne saurait reproduire un apprentissage humain, en montrant la faille qui selon lui s'introduit dans cet argument :

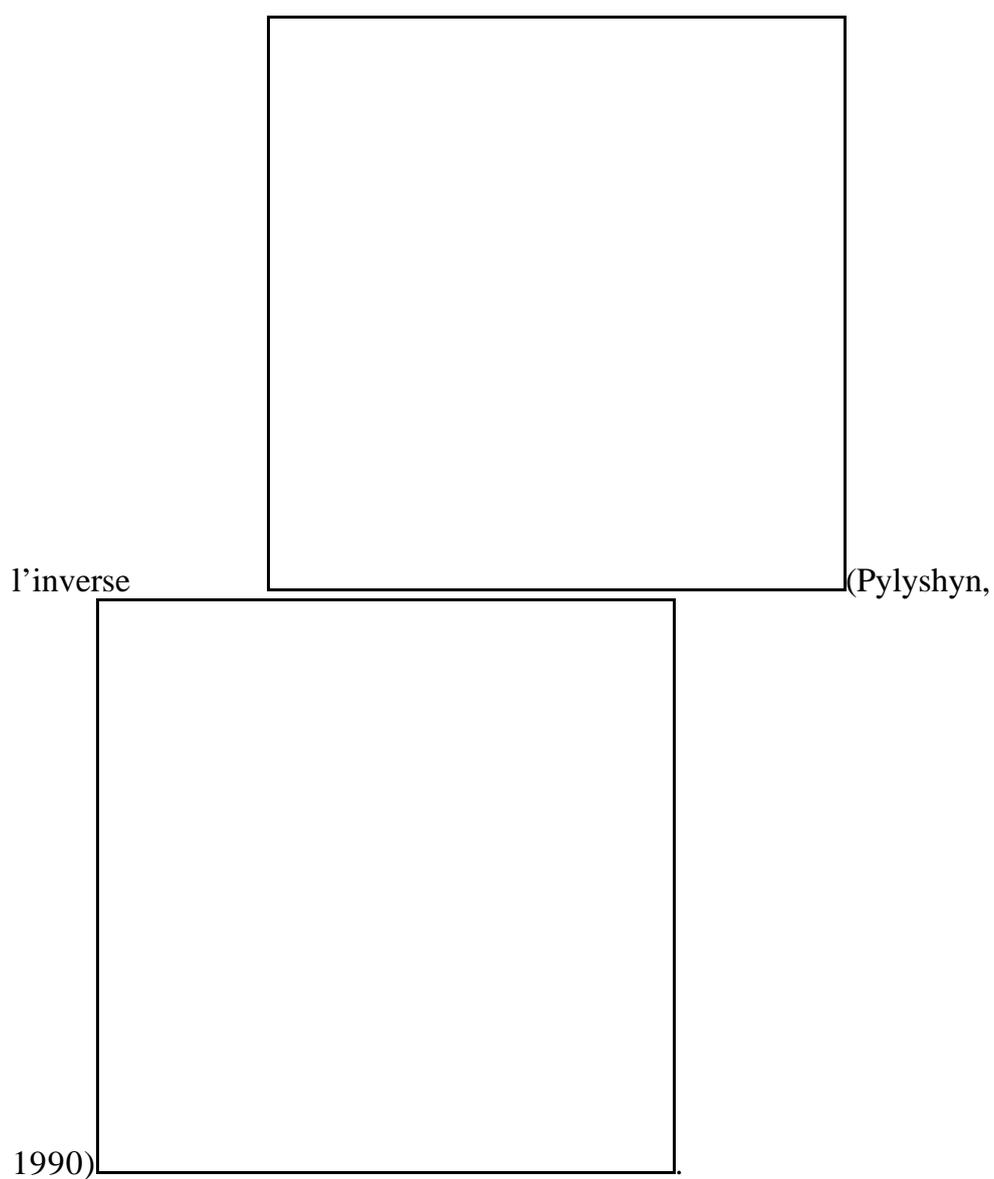
« The flaw is that it is simply not the case that the knowledge to be attributed to a generative system is a transitive closure of the operators over the system. That is, the flaw is to believe that if I give you a set of axioms then the theory of the knowledge level says that you must attribute as knowledge the transitive closure of this thing. The right statement is that something can be described at the knowledge level if you can somehow state the knowledge of the system such that you can predict the behavior of the system from the goal and the knowledge. And if you try and use the transitive closure as the predictor, you will do terribly in predicting it because the system itself never gets close to its transitive closure. » (propos de Newell rapportés par (Agré, 1993, p. 437)).

Newell n'était assurément pas dupe des limitations du SSP et avait compris, avec de nombreux chercheurs, épistémologues, philosophes, qu'il existait nécessairement une zone de la cognition antérieure à toute formulation explicite. Mais l'hypothèse cognitiviste avisée de Newell amène néanmoins à conclure *nécessairement* qu'un immense système formel épuise non l'explication proprement dite mais l'*implémentation* de l'esprit humain. Dans cette optique, le substrat biologique est celui de la machinerie physique sur laquelle est « installé » le programme de l'esprit humain¹⁶, dans un mouvement allant fondamentalement du

¹⁶ « (...) cognition is literally a species of computing, carried out in a particular type of biological mechanism. »

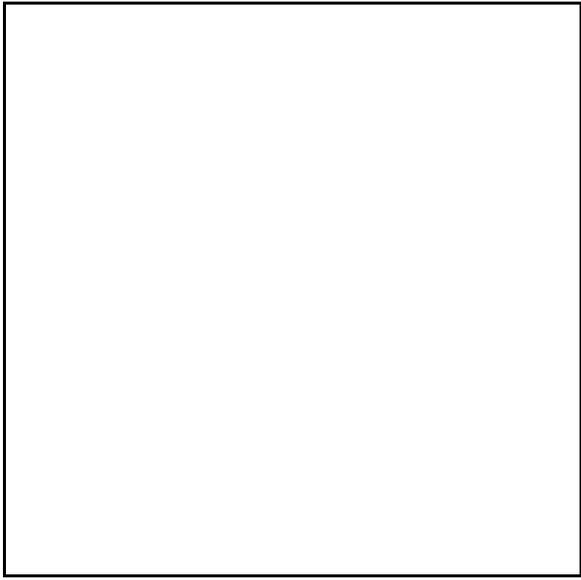
niveau de la connaissance au niveau physique ou biologique, et non



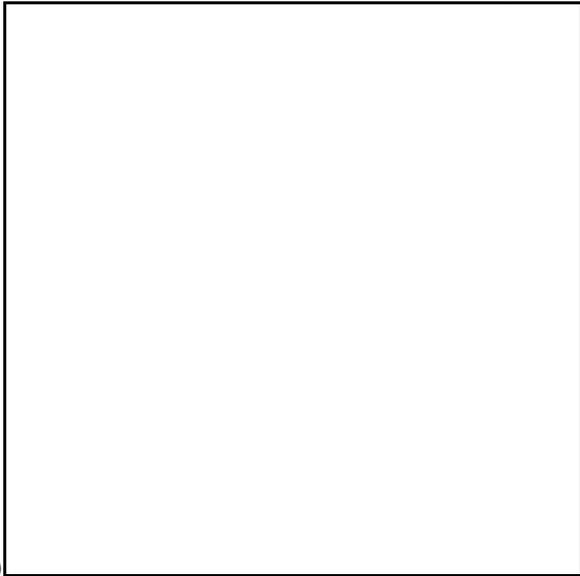


Cette approche de la cognition ne va pas sans soulever quelques problèmes de modélisation cognitive ; en particulier celui de la construction de cette complexité, surtout en l'absence d'une considération de l'entour sémiotique et social.

Le point d'ancrage de cette IA dans les sciences de la cognition est la résolution de problèmes *par l'homme* : ce qui intéresse au premier chef le chercheur en IA qui se recommande des sciences cognitives est le «human problem solving»



(Newell et Simon,

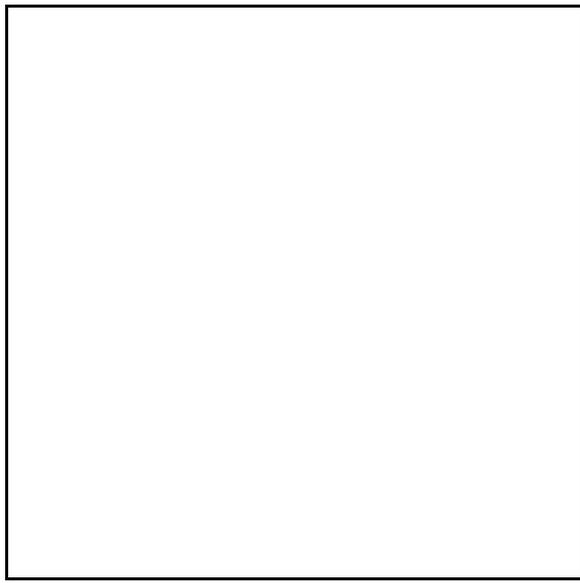


1972) dont les caractéristiques fondamentales et difficiles à conceptualiser pour l'IA sont l'*approfondissement progressif* (*progressive deepening*) et l'hypothèse selon laquelle la structure et la configuration générale de la mémoire humaine dictent au raisonnement ses cheminements¹⁷. Le point névralgique se trouve inscrit dans cet *a priori* inaugural de l'IA : l'esprit qui résout le problème est «pur esprit», même si pour résoudre

¹⁷ « What you see is that memory shapes the method. And the big thing that comes out is that the search strategy that humans use is essentially progressive deepening. And it's funny that the whole development of AI doesn't recognize progressive deepening as a strategy» (propos de Newell rapportés par Agre, 1993, p. 419).

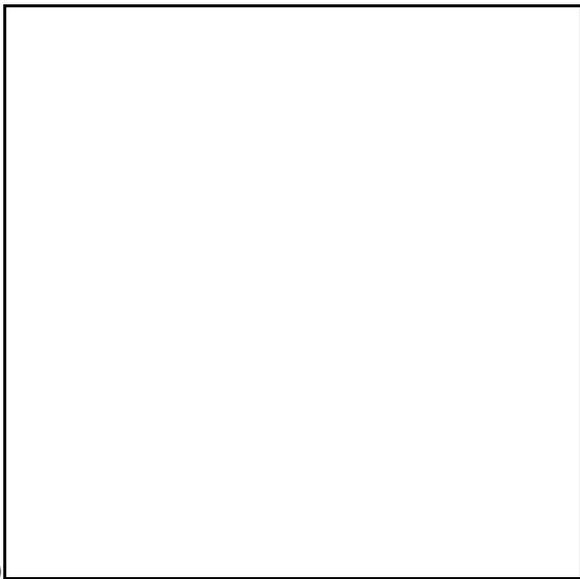
le problème il passe par des phases d'apprentissage. Selon ce point de vue, il serait saugrenu de prétendre que c'est *le corps* qui résout le problème : les opérations sur des symboles (e.g., $2+2=4$) sont entièrement indépendantes des états corporels et mentaux du sujet qui les «implémente». Elles existent dans un monde intemporel qui n'est pas celui de la logique formelle mais celui à partir duquel peuvent être formulées des connaissances et des buts. Quel est ce monde, sinon un monde transcendant ? *A contrario* A. Damasio rapporte cette observation fondamentale de *l'esprit humain* (cf. I.3) : un raisonnement digne du qualificatif de *complexe* conduit *par un sujet humain* ne saurait s'envisager sans le contrôle heuristique permanent du système sensoriel et de l'affectivité (par le biais d'un grand nombre de *marqueurs somatiques*). Dit plus radicalement, le raisonnement le plus complexe n'est que l'expression la plus raffinée d'un système sensori-moteur très évolué dont l'entour est à la fois naturel et culturel. Contrôle qui, s'il vient à disparaître, fait du sujet, fût-il le plus performant dans l'exercice de la logique, un malade que la société doit assumer.

Il existe un niveau antérieur à celui, nommé *symbolique*, de la conception du système computationnel en termes de représentations, de structures de données et d'états du système. C'est le niveau de la connaissance, qui n'est autre qu'un niveau intentionnel : celui où se forme originellement ce que l'on attend du système et qui sera *ensuite* conçu précisément, en termes symboliques, puis concrètement en termes physiques. Le niveau de la connaissance chez Newell est celui dans lequel prend naissance un *projet* de réalisation de système computationnel : le système, avant d'être spécifié en termes de structures de données et d'états internes, existe en termes de fonctions à remplir et de connaissances disponibles ou nécessaires pour remplir



ces fonctions

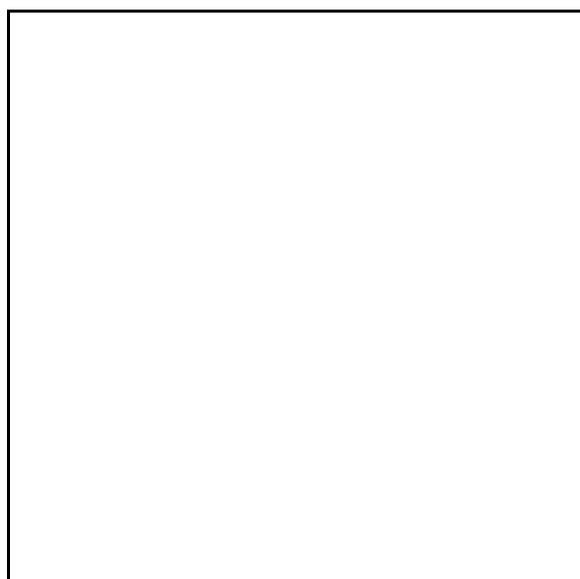
(Newell,



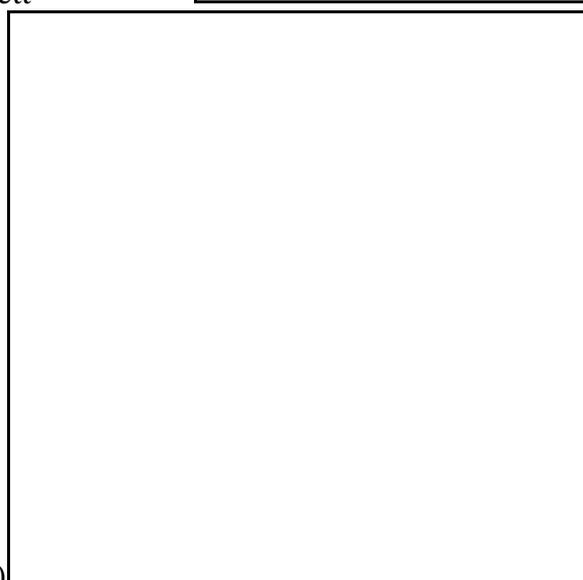
1993)

L'expression

Knowledge level pourrait alors être traduite par *niveau intentionnel de la connaissance*, terme que n'aurait pas désavoué Newell lui-même en raison de son affinité avec la *posture intentionnelle* définie par

Dennett¹⁸

(Dennett,



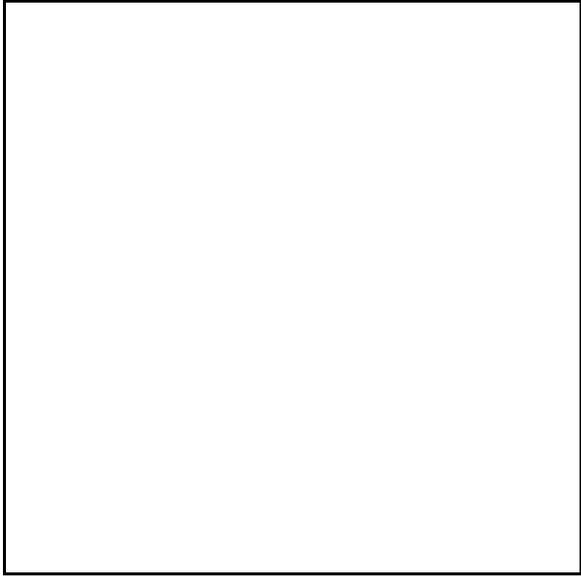
1988)

Ce niveau intentionnel ne peut en aucun cas être confondu avec le niveau symbolique ou computationnel, ce qui fait toute la difficulté de sa définition précise : aussi loin que l'on pousse l'expression formelle d'un système de symboles physiques, il subsistera toujours une partie non réductible, imputable au niveau intentionnel, sans quoi ce dernier serait sans objet¹⁹. Le niveau intentionnel est défini par Newell comme

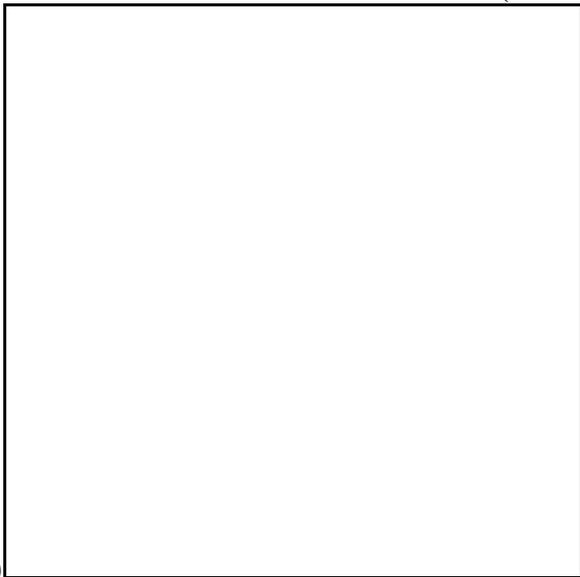
¹⁸ « (...) a concept of *intentional stance* or *intentional system* was produced by Dan Dennet (...), which is in most ways identical to the knowledge level. » (Newell, 1993, p. 36).

¹⁹ « The knowledge level is, in a certain sense, formless and it is not clear what to do with it. » (Newell, 1993, p. 35).

le niveau de la connaissance à l'état brut, telle qu'elle se présente à l'esprit d'un concepteur de systèmes intelligents. Dans cette optique, ce niveau est entièrement *a priori* : il n'existe aucun niveau causal antérieur susceptible de l'expliquer. C'est la connaissance à *l'état pur*, le point ultime duquel naît le système intelligent dans la pensée du concepteur. Sa forme computationnelle détaillée est ensuite affaire de formalisation dans un système de symboles rigoureux, puis sa forme physique est affaire de technologie. Sur un plan pratique, le *niveau intentionnel de la connaissance* correspond à un mode d'énonciation destiné à faciliter la définition d'un système particulier, en faisant l'économie des précisions en termes de *représentations*, de *structures de données* et *d'états du système*



(Newell,



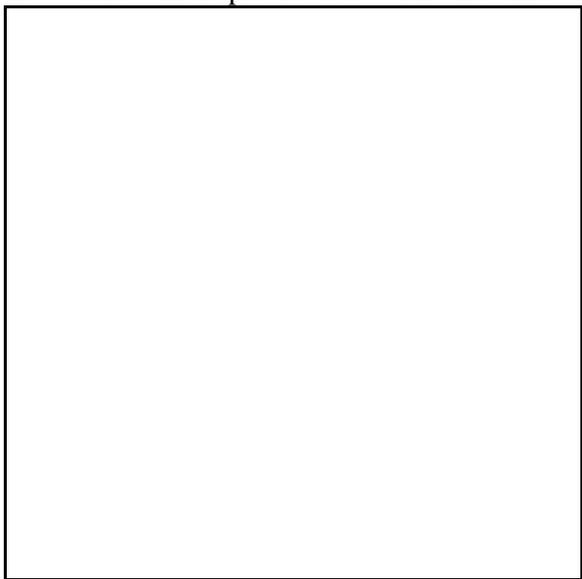
1993)

. Ce concept permet de

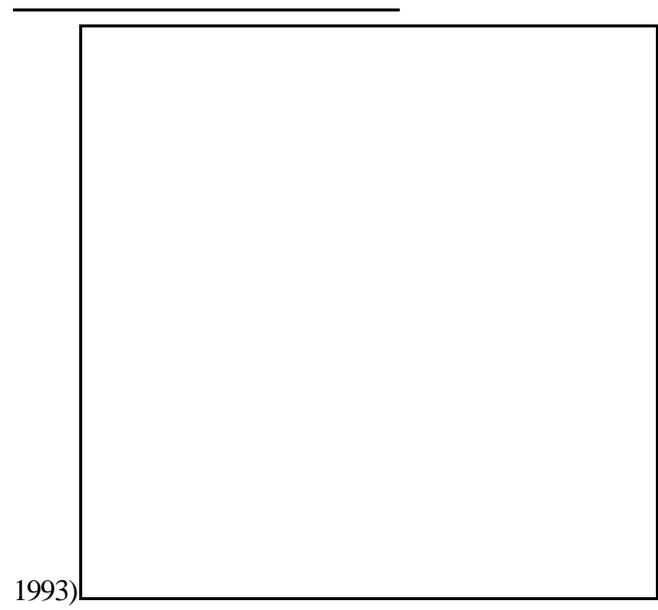
ramener dans le discours scientifique les propositions non formelles jaillissant du pur esprit du concepteur en IA. Que ces propositions antérieures à toute formulation aient partie liée avec le propre *système interprétatif* du concepteur, hérité de son entour culturel, cela n'a pas été envisagé par Newell.

Il reste que la complexité de la cognition humaine réside principalement dans la dynamique de sa construction, qui explique du même coup le pourquoi de ses fonctions. C'est de cette dynamique même — dans l'espace-temps physique et sémiotique du sujet —, que naît la complexité de l'organisation. Mais passionné par le *knowledge level* tel qu'il l'entendait, Newell ne pouvait accorder à l'ontogenèse qu'un intérêt secondaire²⁰.

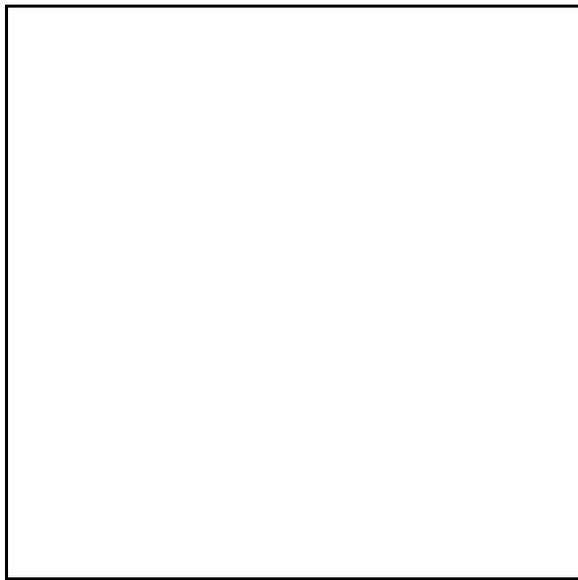
²⁰ cf. Newell : « Ultimately *it may turn out* to depend upon history, it may depend on some point in the history of the system when it came in contact with something in the world which provided it with that knowledge. » (Cité par Dennett dans



(Dennett,

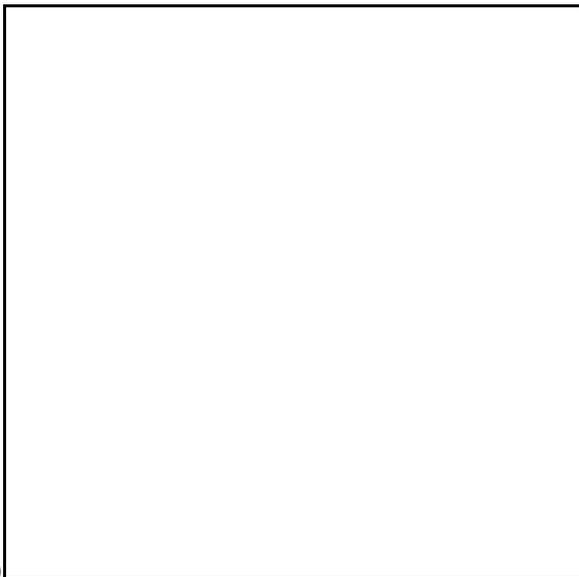


1993) p. 290), (c'est nous qui soulignons).



Soar

(Newell,



1990), le système le plus représentatif des expérimentations des idées de Newell, ne saurait satisfaire un philosophe des sciences cognitives tel que D. Dennett. Bien que s'appuyant largement sur les modèles computationnels, Dennett s'amuse adroitement de ce système :

« Soar is all business. Soar is either working or sound asleep, always learning-by-chunking, always solving problems, never idling. There are no profligate expenditure or dubious degressions, no along-for-the-ride production clustering up the problem space, and Soar is never too tired and cranky to take on yet another impasse. Or so it seems. Perhaps if we put just the right menagerie of operators on stage, or the right items of supplementary knowledge in memory, a sprinkling of sub-optimal goals,

etc., a lazy, mathophobic, lust-obsessed Soar could stand forth for all to see. That is what I mean about how easy it is to misplace the level of Soar; perhaps all this brisk, efficient problem solving should be viewed as the biological (rather than psychological) activities of elements too small to the naked eye of the folk-psychological observer. » (Dennett, 1993, p. 287).

Ainsi, Dennett voit moins dans Soar l'implémentation d'une *théorie unifiée de la cognition* (avec la connotation éminemment totalisante d'une telle expression), qu'une simulation, au demeurant incomplète, d'automatismes biologiques pré-conscients, sous-jacents aux mécanismes psychologiques observables.

Le cœur du problème de la cognition spécifiquement humaine, à savoir la question du langage et de l'entour sémiotique, reste à éclaircir :

« Early on (Newell) gives an eloquent survey on what he calls the "efflorescence of adaptation" by human (and only the human) species (...), but does this paean to productive versatility proclaim that the symbols of an *internalized natural language* are necessary, or is it rather that one needs a pre-linguistic language of thought -- in which case we may wonder why the human language of thought gives us such an edge over the other species, if it does not get most of its power from the external language we learn to speak. » (Dennett, 1993, p. 292-293).

I.2. La conscience expliquée par D. Dennett en tant que processus narratif à versions multiples

L'itinéraire proposé Dennett pour conduire le lecteur à son modèle de la conscience débute, comme celui de A. Damasio (cf. *infra*), au niveau des perceptions — première divergence notoire par rapport à l'approche de Newell — pour s'achever dans une conscience conçue comme un processus narratif :

« Les stimuli visuels provoquent des chaînes d'événements dans le cortex qui produisent graduellement des discriminations de plus en plus spécifiques. (...) [des] parties du cerveau viennent à se trouver dans de nouveaux états qui discriminent différents traits, par exemple d'abord le simple point de départ d'un stimulus, puis la localisation, puis la forme, puis plus tard la couleur (dans un chemin différent), plus tard encore le mouvement (apparent), et finalement la reconnaissance des objets. (...) ce processus distribué se déroule pendant des centaines de millisecondes, pendant lesquelles différentes additions, incorporations, amendements et réécritures de contenu peuvent survenir, dans des ordres divers. Ceux-ci finissent par produire, au bout du compte, quelque chose qui ressemble plutôt à un flux narratif ou à un récit suivi, qu'on peut considérer comme sujet à un

processus d'édition constant par des nombreux processus distribués alentour dans le cerveau et se produisent dans le futur. Les contenus surgissent, sont révisés, contribuent à l'interprétation d'autres contenus ou à la modulation du comportement (verbal ou autre) et, dans ce processus, laissent leur trace dans la mémoire, qui ensuite peut finalement s'estomper ou être incorporée dans ou être réécrites par d'autres contenus, en totalité ou en partie. Cet amas de contenus est plutôt comme un récit en raison de sa multiplicité ; à tout point du temps, il existe des versions multiples des fragments narratifs à différentes étapes du travail d'édition et dans différents endroits du cerveau. » (Dennett, 1993, p. 172-174).

La conception selon laquelle sujet et objet de l'interprétation sont deux entités différentes suppose d'emblée ce que Dennett nomme un *Théâtre Cartésien*. Dans le Théâtre Cartésien existe une irréductibilité fondamentale entre l'interprète et l'interprété, renvoyant directement au dualisme entre le corps et l'esprit. Dennett montre principalement, au cours de ses 600 pages d'explication, que le moi psychologique, celui que nous appelons ici le sujet et qui se définit essentiellement par sa *conscience*, peut être rendu par la métaphore du centre de gravité : le moi est le *centre de gravité de la narration* ; narration dont les objets sont les présentations mentales directement inscrites par l'entour culturel et dont le moteur est biologique :

« Un moi, selon ma théorie, n'est pas un vieux point mathématique, c'est une abstraction définie par les myriades d'attributions et d'interprétations (y compris les attributions de soi et les interprétations de soi) qui ont composé la biographie du corps vivant dont il est le Centre de Gravité narrative. Comme tel, il joue un rôle singulièrement important dans l'économie cognitive incessante de ce corps vivant, parce que, de toutes les choses qui se trouvent dans l'environnement et dont un corps actif doit faire des modèles mentaux, aucune n'est plus cruciale que le modèle que l'agent a de lui-même » (Dennett, 1993, p. 529).

« (...) les moi ne sont pas des perles de l'âme qui existent de façon indépendante, mais les artefacts des processus sociaux qui nous créent, et comme d'autres artefacts de ce genre, ils sont soumis à des changements brusques de statuts. La seule «force» qui revienne à la trajectoire d'un moi (...), c'est la stabilité qui lui est impartie par la toile de croyances qui le constitue, et quand ces croyances se rompent, elle se rompt, soit pour toujours, soit temporairement. » (Dennett, 1993, p. 525).

Dennett se demande s'il doit appeler « matérialisme sémiotique » sa théorie selon laquelle nous produisons notre moi par le langage au cours d'un processus narratif (Dennett, 1993, p. 508). Pour Dennett en effet, le moi psychologique est une abstraction (tout comme le moi

biologique) qui se construit, se «tisse», par un processus narratif : nous nous racontons en permanence ce que nous sommes, faisons, aimons, détestons... et cette narration devient ce que nous avons la conviction d'être.

I.3. Le processus narratif de la conscience biologique chez A. Damasio

Toutefois, le matérialisme sémiotique de Dennett, qui se présente comme un non-dualisme radical, présente pourtant une zone d'ombre que la théorie de A. Damasio éclaircit. Cette zone concerne la non-séparabilité du centre de gravité narrative et du corps. Dans un dialogue simulé avec un sceptique vis-à-vis de sa théorie, qui craint en adhérant à la thèse de Dennett de perdre l'immortalité que lui garantit la métaphysique cartésienne, Dennett prête au défenseur de sa théorie (c'est-à-dire lui-même) ces mots :

«Et si vous étiez une perle de substance matérielle, un groupe spectaculairement spécifique d'atomes cérébraux, votre mortalité dépendrait des forces physiques qui le font tenir ensemble (...). Si vous pensez à vous comme un centre de gravité narrative, en revanche, votre existence dépend de la persistance de cette narration (...), qui pourrait *théoriquement* survivre indéfiniment à de nombreuses inversions de *médium*, être téléportée aussi aisément (en principe) que le journal du soir, et indéfiniment stockée comme de la pure et simple information. » (Dennett, 1993, p. 533-534)²¹.

C'est peut-être là un talon d'Achille de la théorie de Dennett. Ce qui devrait être «stocké», c'est l'information sur la *configuration* dendritico-axonale et synaptique, sur le système hormonal, sur la composition sanguine, etc., qui ne constituent pas le support de la pensée, mais la pensée elle-même à ses racines. C'est l'une des thèses majeures de A. Damasio (Damasio, 1995, p. 146 et suiv.). Nous avons un million de milliards de synapses (10^{15}) dont la configuration interne d'un nombre important d'entre elles se modifie à chaque instant. Outre ce caractère changeant de la connectivité du système neuronal, l'outil qui constaterait son état le modifierait du même coup. Enfin, le programme qui a amené le système cérébral dans un état donné n'existe nulle part.

A. Damasio reprend l'hypothèse d'une narration constituant le «je», en montrant que cette narration est une forme de second ordre, établie

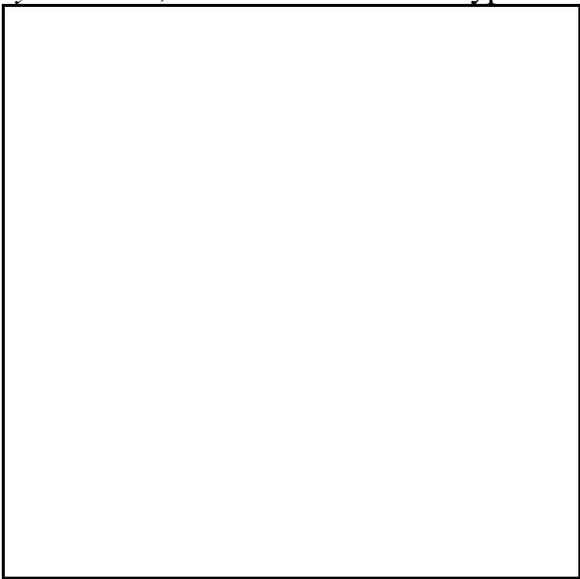
²¹ Cette thèse avait été admirablement mise en scène par Arthur C. Clarke dans son roman de science fiction *La cité des étoiles*, Denoël, 1985.

dans la continuité même d'une narration de premier ordre directement corporelle : une narration *biologique*.

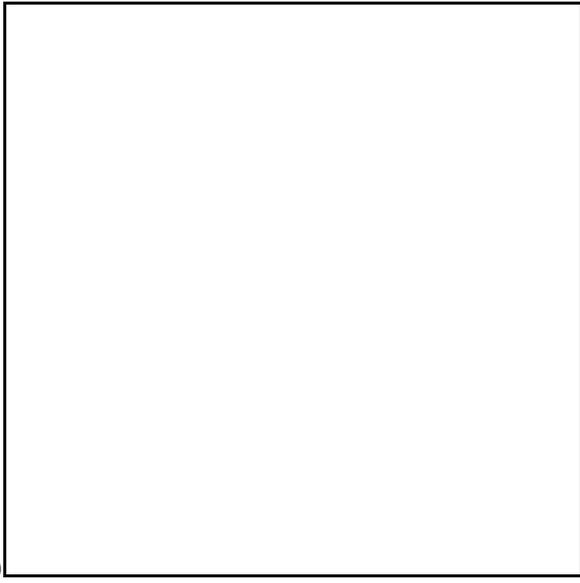
«Les êtres humains possèdent des capacités narratives de second ordre, fournies par le langage, grâce auxquelles peuvent être construites des narrations verbales à partir des non verbales. La forme raffinée de subjectivité qui est la nôtre émerge sans doute de ce dernier processus. Le langage n'est peut-être pas à l'origine du moi, mais il est très certainement à l'origine du « je » » (Damasio, 1995, p. 305).

Si l'on adopte la position selon laquelle il n'existe nulle part dans le cerveau, ou à l'extérieur du cerveau, un lieu précis où siègerait la conscience observatrice du monde depuis son observatoire ou théâtre cartésien, position commune à Dennett et A. Damasio, on est amené à concevoir avec eux une conscience provenant de *parcours* opérés par des mécanismes biologiques nécessaire à la vigilance : le moi biologique (différent de la conscience de soi et entièrement non conscient) (Damasio, 1995, p. 301); parcours situés à l'intérieur d'un réseau de contenus mentaux : les *représentations potentielles* (Damasio, 1995, p. 142).

Ces contenus proviennent de diverses zones du cerveau et sont *synthétisés*, selon l'hypothèse de A. Damasio



(Damasio,

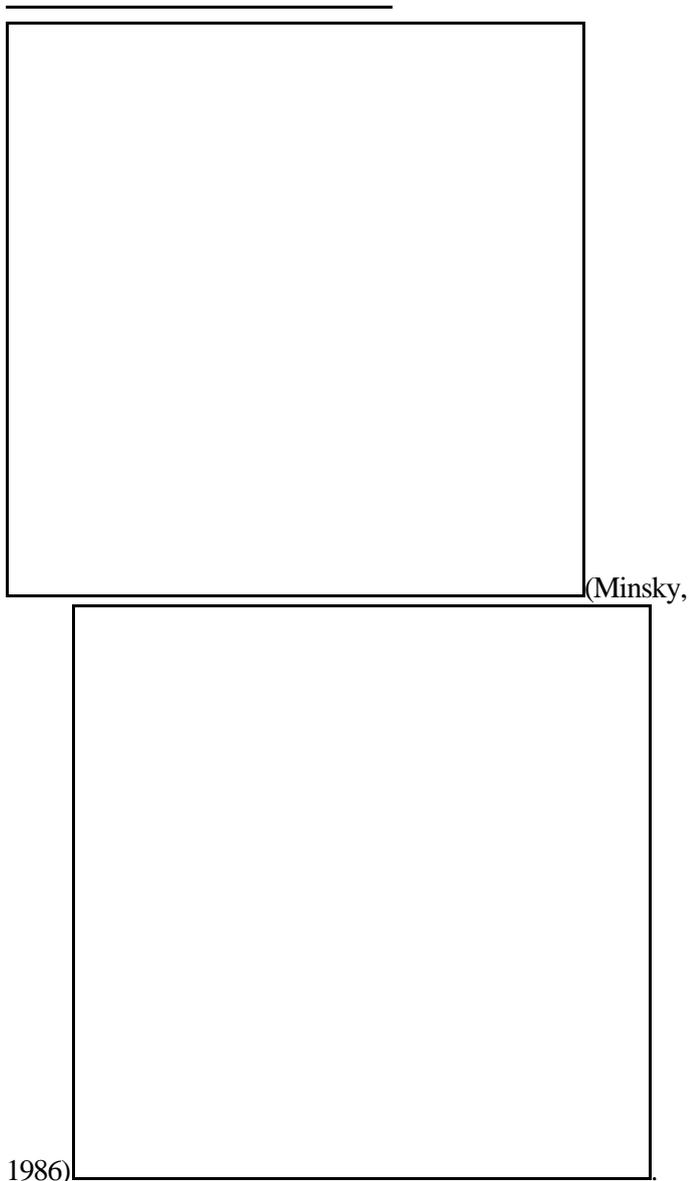


1995) dans des zones spécifiques appelées *zones de convergence* (Damasio, 1995, p. 139), dans lesquelles s'effectuent notamment les opérations de déduction (Damasio, 1995, p. 210). Celles-ci permettent de «construire» des entités complexes à partir d'éléments plus simples²². Les neurones, ou

²² Hypothèse neuropsychologique qui fait écho à l'hypothèse computationnelle d'une *société d'agents et d'agences* telle que formulée par M. Minsky

amas neuronaux, de discrimination de percepts, transmettent leur résultat de détection à d'autres amas neuronaux chargés de répondre à des conjonctions de percepts éventuellement très complexes, comme dans les mécanismes de reconnaissance de visages.

Le moi biologique en tant que système de vigilance sous-jacent aux états corporels, sensations et actions, serait incapable par lui-même d'une connaissance à proprement parler. Il doit être complété par ce que A. Damasio nomme une *représentation tierce-partie* ou un *méta-moi*, non sans quelque maladresse dans la dénomination (qui n'enlève



rien à l'idée) puisque rien n'est *re-présenté* dans ce mécanisme neuronal.

« Le moi tel qu'il est décrit ci-dessus ne peut pas savoir. Cependant un processus que nous pourrions appeler le «méta moi» est peut-être en mesure de savoir, à condition que : 1. le cerveau puisse élaborer une certaine description de la perturbation de l'état de l'organisme, perturbation qui a été engendrée par les réponses du cerveau à la présence d'une image ; 2. la description puisse engendrer une image du processus de la perturbation ; 3. l'image du moi perturbé puisse être affichée en même temps que l'image qui a déclenché la perturbation. En résumé, la description dont je parle concerne la perturbation de l'état de l'organisme, par suite des réponses du cerveau à l'image de l'objet X. Cette description n'emploie pas de langage, bien qu'elle puisse être traduite sous forme verbale. » (Damasio, 1995, p. 303).

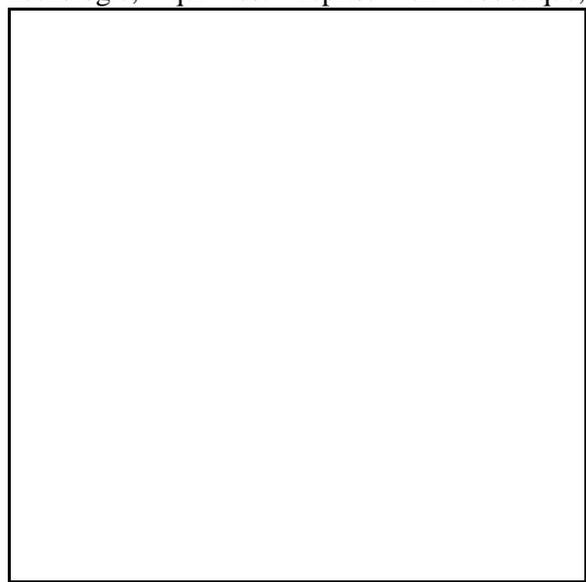
« Le «méta-moi» que j'envisage est une construction purement non verbale, une représentation schématique des rapports entre les protagonistes principaux établie depuis un point de vue extérieur à chacun d'eux. En fait, cette représentation tierce partie constitue une narration non verbale, moment après moment, des événements affectant ces protagonistes. Cette narration peut être faite sans utiliser le langage, mais en recourant aux mécanismes élémentaires de représentation dans l'espace et dans le temps des systèmes sensoriels et moteurs. Je ne vois pas de raison pour laquelle les animaux qui ne possèdent pas de langage n'élaboreraient pas de telles représentations » (Damasio, 1995, p. 305).

Bien que les deux approches, de Dennett et A. Damasio, soient différentes dans leur méthode, et que A. Damasio déclarât explicitement qu'il ne se retrouvait pas dans la machine virtuelle joycienne de Dennett (Damasio, 1995, p. 306), la théorie du philosophe et celle du neurologue se retrouvent sur ce point : la conscience, même la plus élémentaire, est un processus de parcours de type narratif, dans un ensemble d'*images* selon la terminologie de A. Damasio (Damasio, 1995, p. 132 et 134).

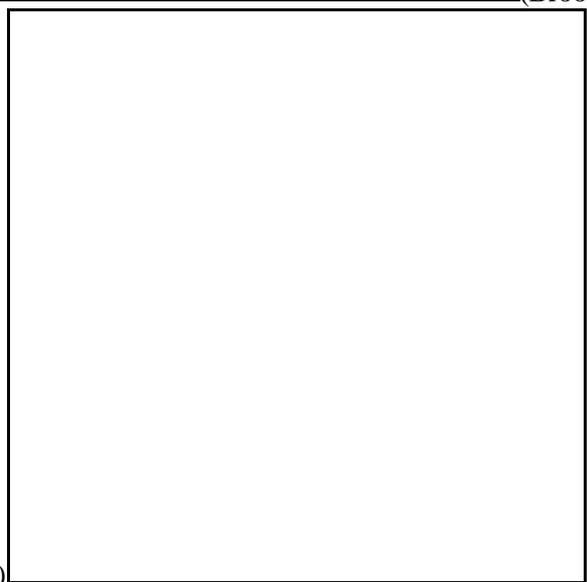
Selon l'hypothèse de A. Damasio, partagée par un certain nombre de neuropsychologues, il *n'existerait pas* de représentation d'information en tant que telle dans le cerveau, mais seulement des processus de réactivation de parcours perceptifs (Damasio, 1995, p. 138), chaque perception donnant lieu à, ou modifiant, une représentation potentielle. Dans les systèmes biologiques, *a fortiori* dans le cerveau humain, la mémoire est directement construite sur les processus perceptifs, de sorte qu'il est impossible de séparer ces processus des processus de

mémorisation, c'est-à-dire du problème de la représentation²³. Mais ce que vise essentiellement l'approche neuropsychologique de A. Damasio est l'explicitation des mécanismes de raisonnements complexes. En bref, l'efficacité de ce raisonnement repose en majeure partie sur l'information venue du corps pour assurer l'instinct cognitif de conservation, au moyen de *marqueurs somatiques*. Ces marqueurs informent le système cognitif global sur la *viabilité biologique* du raisonnement en cours. Le contrôle opéré par les marqueurs somatiques est particulièrement actif dans les raisonnements qui

²³ C'est cette même hypothèse, apparaissant comme la plus vraisemblable en neurologie, qui est reprise en robotique, par R. Brooks notamment

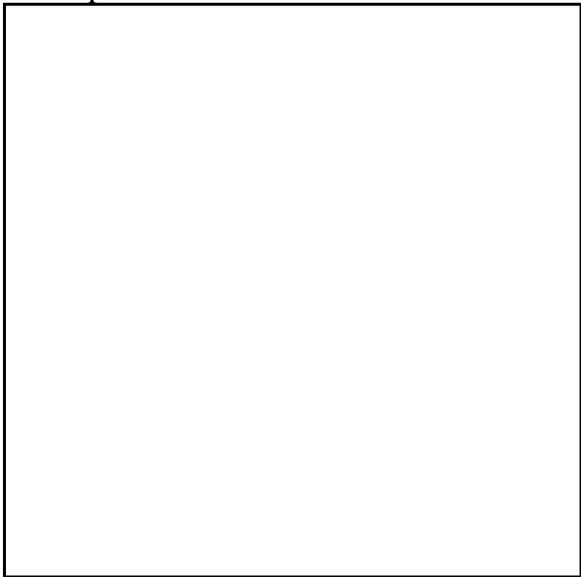


(Brooks,

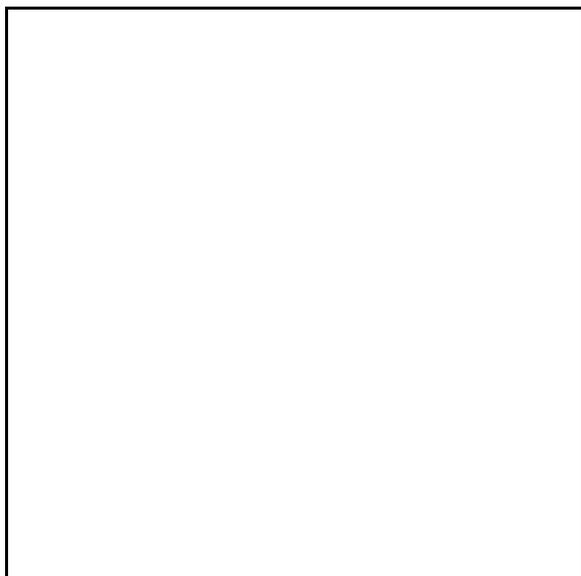


1991)

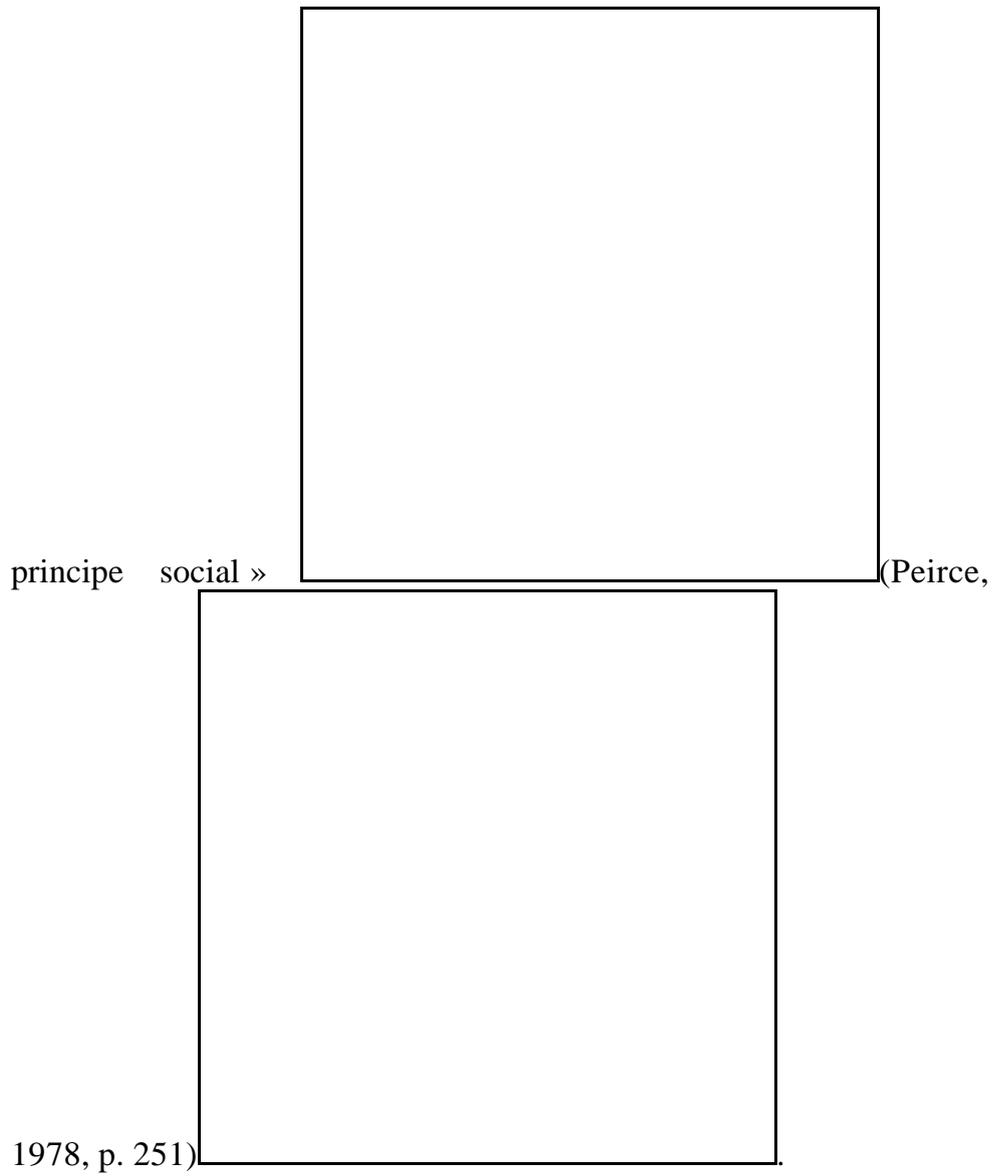
mettent en jeu le devenir de la personne et reposent sur un grand nombre de données perceptives ou issues elles-mêmes de raisonnements. Les stratégies de choix d'action ou de comportement dans des contextes sociaux complexes offrent les exemples les plus marquants de ce type de raisonnement. Dans ce type de résolution de problème, la conscience ne peut examiner tous les parcours déductifs possibles dans un horizon temporel imposé par le contexte. La conscience recourt alors aux heuristiques constituées par les marqueurs somatiques. En résumé très schématique, on ne prend conscience *in fine* que des déductions que le corps, à l'insu de la conscience, a déjà examinées puis autorisées par ses marqueurs somatiques en termes de bon ou de mauvais pour l'homéostasie, la survie, le développement. Les émotions constituent la synthèse d'un grand nombre de marqueurs somatiques, et ont pour fonction d'infléchir le comportement général vers un retour à l'état d'homéostasie (ressenti comme le bien-être) duquel il s'est écarté. Il faut souligner ici que l'hypothèse neuropsychologique du raisonnement complexe que propose A. Damasio corrobore la définition Peircéenne de l'interprétant logique, doublement déterminé par le biologique et par le social: «(...) en ce qui concerne le désir, il serait facile de montrer (si cela en valait la peine) que l'interprétant logique est un effet de l'interprétant énergétique dans le sens où ce dernier est un effet de l'interprétant affectif.»



(Peirce, 1978, p.



135). En d'autres termes, il faut *désirer* pour être logique, et «La logique à sa racine dans le



Dès lors, il est à craindre que le puissant résolveur automatique de problèmes, tel qu'il est envisagé par l'IA cognitive, ne réponde au tableau symptomatologique du malade Elliot, cérébro-lésé frontal, de A. Damasio. La caractéristique majeure de ce type de malade est de pouvoir conduire à la perfection des «raisonnements» très complexes sur le plan combinatoire, représentationnel et logique, et de se trouver dans le même temps incapable de *décider* dans les contextes sociaux courants (ils tentent en effet en permanence de parcourir tous les schèmes déductifs possibles à partir d'une situation donnée, effectuent des comparaisons entre les parcours selon tous les critères possibles

dans le but de trouver le meilleur... qu'ils ne trouvent bien sûr jamais)²⁴, d'où perte de l'autonomie les conduisant à une déroute personnelle : perte d'emploi, dégradation de leur entourage affectif et familial, opprobre et rejet social (étant *vus* comme intelligent mais dépourvus de scrupules, de sens des responsabilités...). La prise en charge thérapeutique leur est souvent refusée (leur intelligence purement combinatoire et symbolique y compris sur les situations empruntées au domaine de la vie courante sociale et relationnelle leur permettant de réussir haut la main les tests de performance psychotechniques habituels (Damasio, 1995, p. 57 et suiv.).

Pour simuler algorithmiquement des raisonnements complexes tels qu'étudiés par A. Damasio, l'algorithme devrait disposer du système informationnel constitué par la peau, l'appareil musculo-squelettique, les organes internes, les systèmes limbiques et les cortex, et prévoir les étapes de synthèse de ces flux — presque partout bidirectionnels — d'information que sont les émotions.

2. LES DETERMINANTS DE L'INTELLIGENCE HUMAINE : *PHYSIS*, *SEMIOSIS* ET SUJET

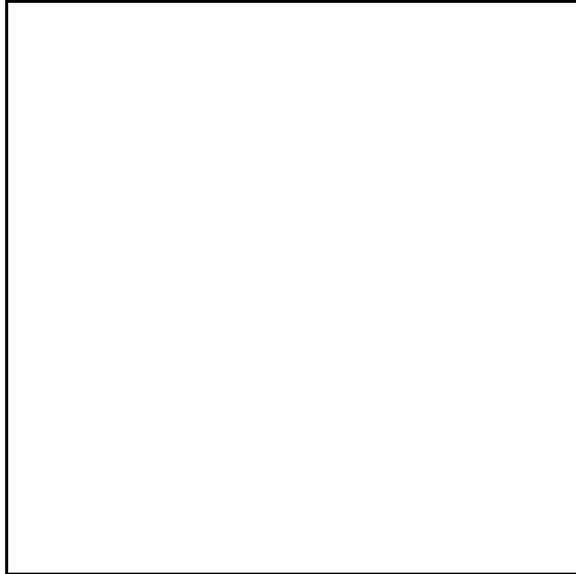
Comment s'effectue le passage du méta-moi biologique — proprement animal — au moi discursif, celui qui sait en nommant ce qu'il sait ? Ce sera l'objet de cette deuxième partie, qui tentera d'apporter quelques éléments de réflexion sur le *sujet de la conscience*.

²⁴ « L'observation de patients tels qu'Elliot suggère que la froide stratégie invoquée par Kant et d'autres auteurs, ressemble plus à la façon dont les personnes atteintes de lésions pré-frontales procèdent pour prendre une décision qu'à celle des individus normaux. » (Damasio, 1995, p. 223).

II.1. Deux espaces fondateurs du sujet

II.1.1 Au-delà de l'approche binaire

Reprenons l'opposition entre le global et l'entour telle que présentée par F. Rastier, avant de nous risquer à quelques développements théoriques. Dans cette formulation, l'*entour* (*Umwelt*) désigne «le monde perçu, et vécu en tant qu'il est interprété», alors que « [le] substrat, notamment physique, de l'entour, le *Welt* venant en arrière plan, nous l'appellerons *arrière-monde*, en empruntant ce terme



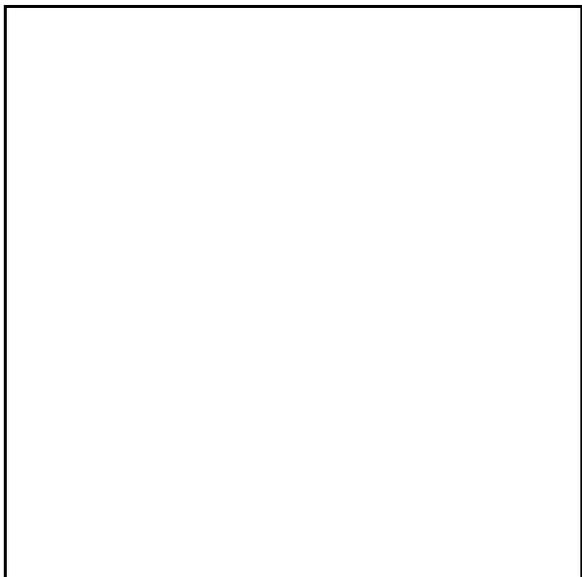
à Clément *et al.* »

(Rastier,

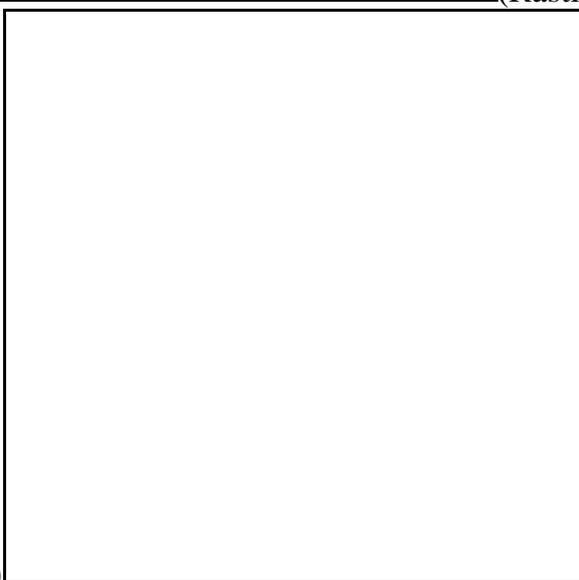
1996)

L'entour englobe la sphère des présentations ainsi que la sphère sémiotique, alors que l'arrière-monde comprend la sphère physique.

Pour Rastier, l'entour inclut le sujet en tant qu'interprète des présentations (les *images* chez A. Damasio) qu'il forme dans son contact avec l'arrière-monde. Cette distinction entre entour et arrière-monde partage plusieurs traits communs avec la classique distinction binaire entre *sémiosis* et monde physique, chez Eco notamment. Mais par son caractère purement binaire, ce système d'oppositions ou d'articulations entre *physis* et *sémiosis* ne permet pas de sortir du vieux « face-à-face de la Pensée et du Réel »



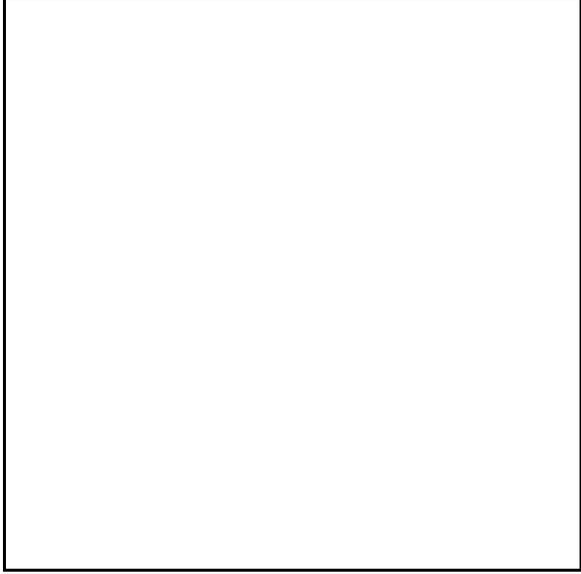
(Rastier,



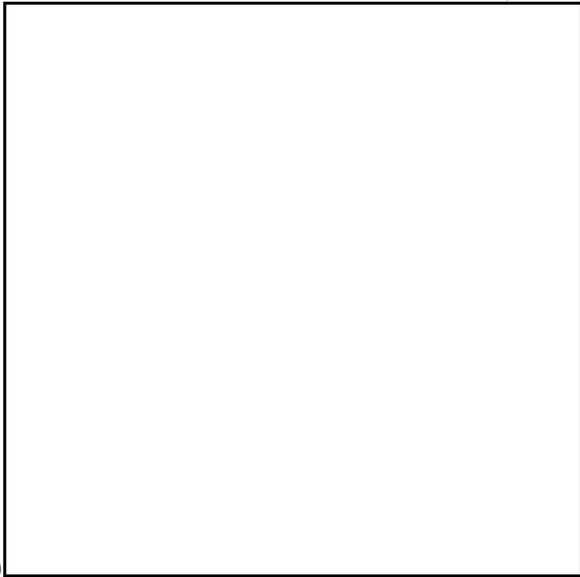
1996) caractéristique de la métaphysique occidentale. Il existe bien un *réel* et des *représentations*, malgré l'ambiguïté de ces termes, mais celles-ci font partie de celui-là : si l'on en reste à cette formulation, la circularité est immédiate. Le problème posé par cette formulation est similaire à

celui de la projection sur un plan d'un objet de dimensionnalité supérieure à deux : l'impossibilité d'observer sur ce plan ce qui lui est orthogonal. Ce plan est celui de la *physis*. Ce qui lui est irréductible, la *sémiosis*, n'y apparaît plus que par les interconnexions neurales entre les images cérébrales des objets sémiotiques et, en conséquence directe, «du sujet dans cette affaire, on n'en voit pas de trace» comme le déclarait naguère le cognitivisme (Sperber, 1987, p. 47).

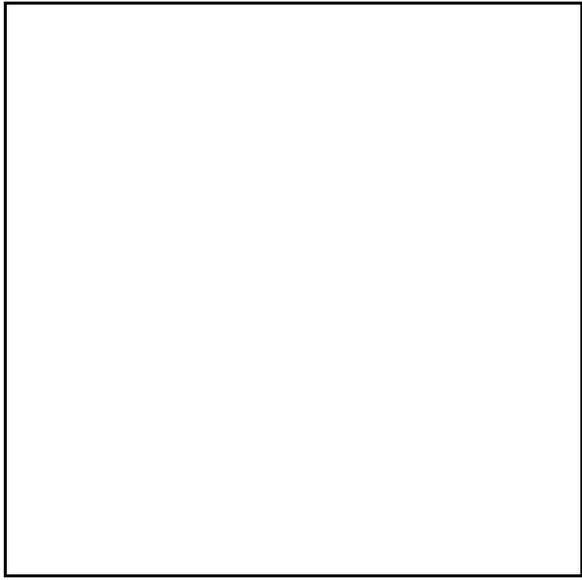
La dimension manquante a trait à la question de l'interprétation



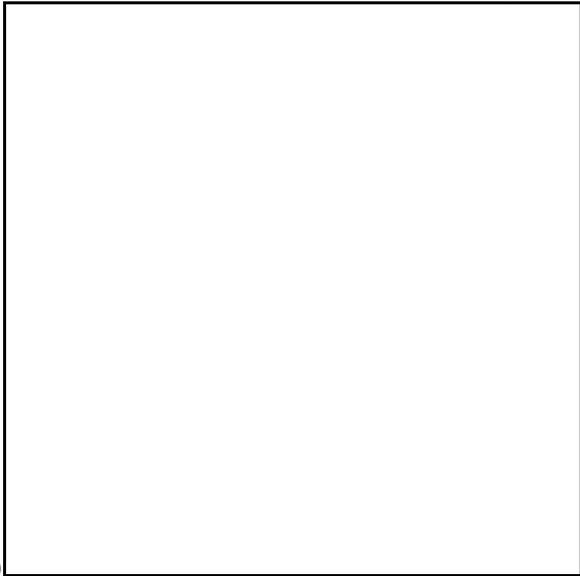
(Rastier,



1996) de la langue dans l'interaction entre sujets (e.g., c'est-à-dire à l'usage



(Jorion, 1990 ; Vernant,



1994), pour laquelle une réduction au biologique est non pas impossible mais *dissolvante*.

L'interprétation étant nécessairement celle d'un sujet et l'interaction sémiotique se déployant *entre* des sujets humains, notre propos est de préciser ce que l'on peut entendre par *sujet*. En outre, la compréhension de la conscience subjective ramenée dans le champ de la recherche scientifique²⁵, n'est pas distinguable de celle du sujet de la

²⁵ voir en particulier l'*hétérophénoménologie* de D. Dennett (Dennett, 1993, chap. 2 et 3), dont le programme est exactement celui de la *phanéoscopie* de C.S. Peirce, c'est-à-dire « la description de ce qui est devant l'esprit ou la conscience, tel qu'il apparaît » (Peirce, 1978, p.10)..

conscience, bien que les deux termes, sujet et conscience, ne soient pas synonymes.

II.1.2 Relation des espaces

Considérons l'hypothèse selon laquelle *le sujet est une relation* : on ne le trouve ni dans la nature ni dans la *sémiosis*, mais en tant que médium et médiateur de leur interaction.

Selon cette hypothèse, le sujet est producteur et produit de la *sémiosis* comme de la *physis*. Il appartient au monde naturel par sa base biologique et à la *sémiosis* par le langage (Fig. 1). Si l'interaction vient à disparaître, le sujet disparaît aussi.

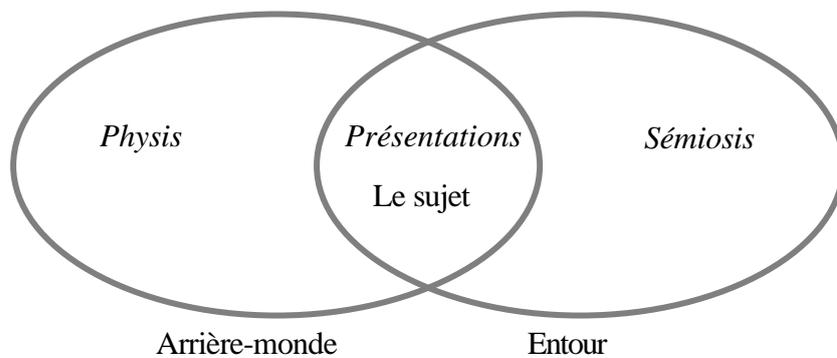
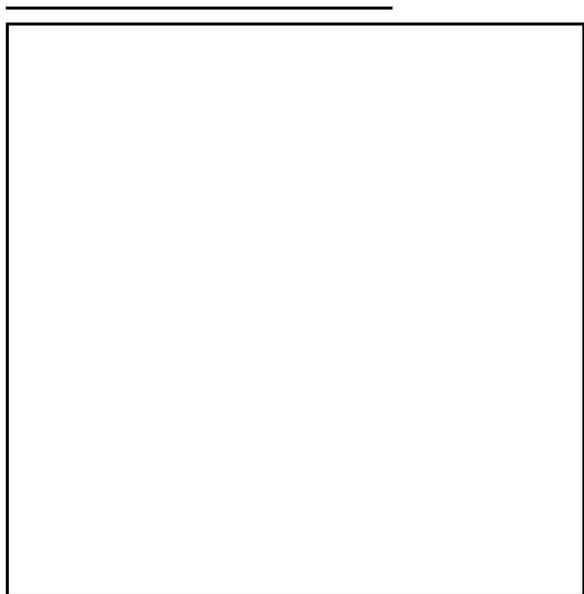


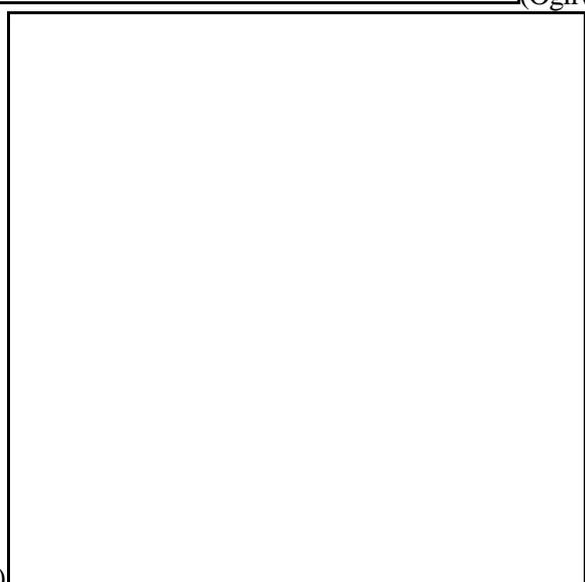
Figure 1. Métaphore de l'intersection

Mieux qu'une simple intersection, qui suggère une bi-appartenance du sujet aux deux espaces qui le déterminent, la métaphore géométrique la plus à même de rendre compte du triptique *physis-sujet-sémiosis* semble être le nœud borroméen (Fig. 2). Celui-ci se présente comme un entrelacs indissociable de trois cercles, dont aucun ne coupe l'espace intérieur de l'autre et dont la suppression de l'un sépare les deux autres²⁶.

²⁶ Nous empruntons la métaphore du nœud borroméen à Lacan qui l'a lui-même empruntée aux mathématiciens (cours de topologie de M. Guilbaud notamment) pour montrer les relations structurantes entre le *réel*, le *symbolique* et l'*imaginaire* ; le sujet *de l'inconscient* résulte du *nouage* même de ces trois domaines. Si nous voulions nous risquer à un rapprochement, nous dirions que l'anneau représentant le sujet (Fig. 2), correspondrait plutôt chez Lacan à la fonction *imaginaire* qui détermine le « moi », alors que l'anneau de la *sémiosis* correspondrait à la fonction *symbolique* lacannienne. Mais la comparaison ne saurait être poussée plus loin sans entrer dans la théorie du sujet selon Lacan, au demeurant très complexe. De plus, nous nous attachons ici non pas au sujet de l'inconscient mais à celui de la conscience. Signalons parmi d'autres une introduction à la théorie du sujet selon Lacan par B. Ogilvie



(Ogilvie,



1987) contributions de l'ouvrage collectif de P. Kaufmann ainsi que diverses

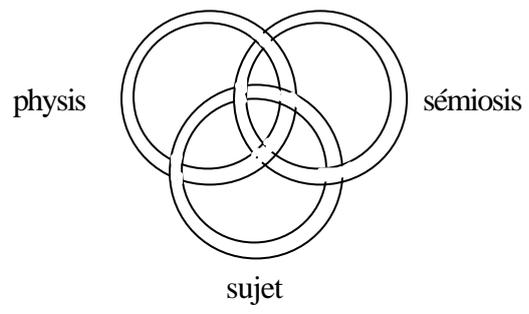
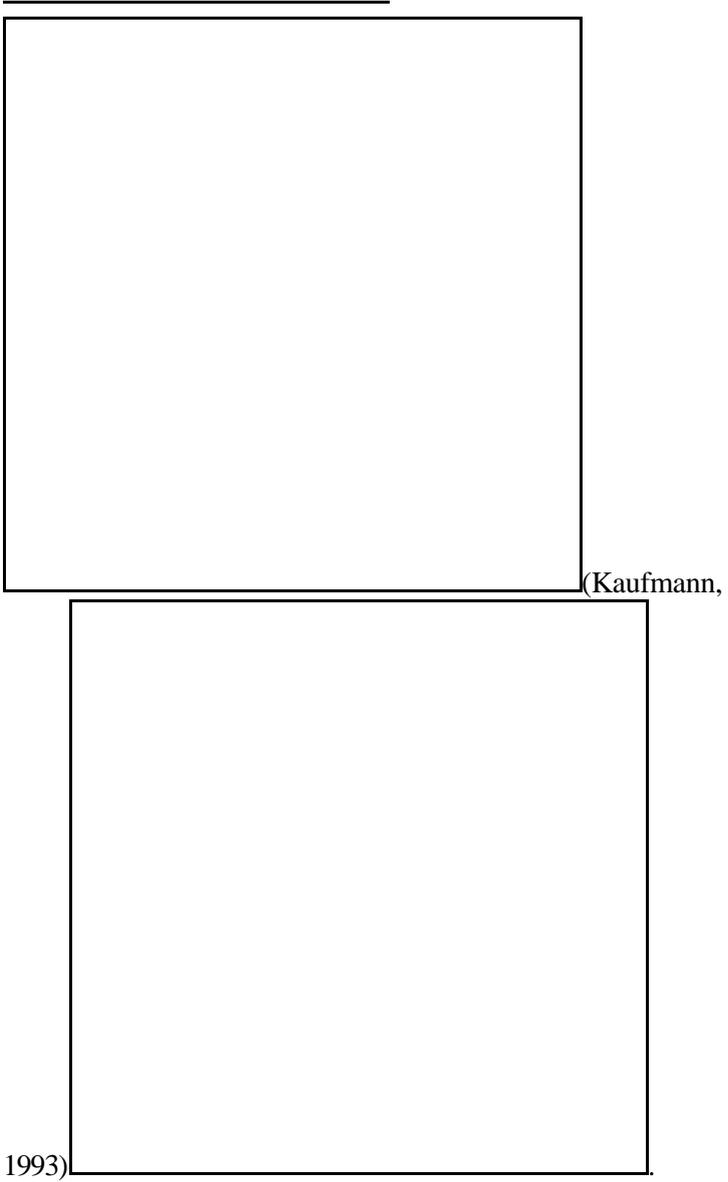


Figure 2. Métaphore du nœud borroméen



On peut énoncer, pour faire simple, que le sujet conscient est la résultante d'un processus sémiotique (discursif, narratif...) dans un moi constitué par un processus biologique. La conscience est un parcours narratif. Le sujet n'est ni uniquement physique, ni uniquement sémiotique, mais le point étrange de jonction de ces deux mondes ; étrange en ce qu'il est encore à l'heure actuelle a-scientifique²⁷. Si les sciences du vivant, ancrées dans la *physis*, rencontrent aujourd'hui le problème de la conscience du sujet humain, les sciences du signe le rencontre aussi. Peirce en particulier, au moment où il définit la pensée en tant qu'ensemble d'interprétants, indique aussi que la limite du champs de la *sémiosis* est atteinte lorsqu'on aborde la connaissance des contenus de la conscience : « A l'exception de la connaissance, dans le cas présent, du contenu de la conscience dans ce cas (connaissance dont l'existence peut être mise en doute) toute notre connaissance et notre pensée se font par signe » (Peirce, 1978, p. 30).

Remarquons également que ce que nous appelons *le sujet* est, dans les grandes classes qui fondent la définition de l'humain (chez Locke et Peirce notamment, et repris par Eco²⁸), le lieu par excellence de *l'éthique*, reliée mais non irréductible à la *physis* et à la *sémiosis*.

Dans l'espace complet des conditions de développement du sujet, le langage verbal figure comme partie de l'espace sémiotique, complémentaire de l'espace physico-chimique. L'ontogenèse du sujet se déroule alors dans un milieu à la fois naturel et culturel. Le corps cognitif n'acquiert l'intelligence discursive — propre à l'homme — qu'en vertu d'un entour sémiotique qu'il rencontre et doit apprendre à sa naissance, et qu'il contribue en retour à construire.

II.1.3 Le sujet dans l'espace biologique

S'il est lieu de l'interprétation, le sujet contient aussi la cause biologique de l'interprétation, sans être pour autant entièrement déterminé par le biologique ni lui être aisément réductible²⁹.

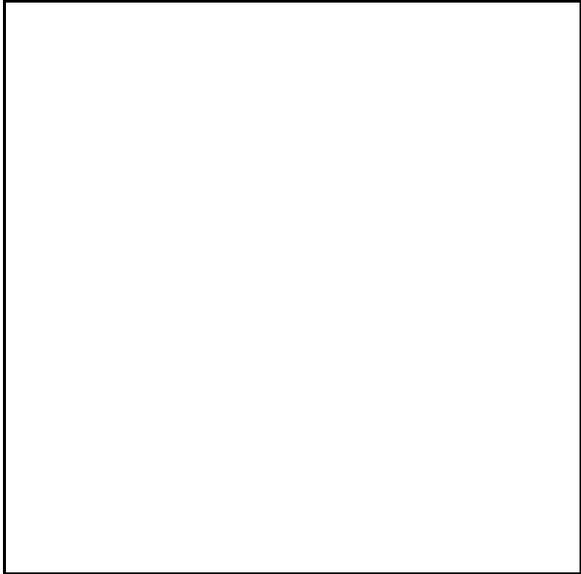
²⁷ Les approches psychanalytiques de la construction du sujet, réputées hors du champ de la science, devraient connaître un regain d'intérêt dans les sciences cognitives en raison ne serait-ce que de leur richesse heuristique, mais nous resterons ici à un niveau très simple dans lequel le sujet en tant que *sujet du désir* ou *sujet de l'inconscient* ne sont pas abordés.

²⁸ « (...) la sphère de l'entière connaissance humaine se réduit à la physique, l'éthique et la sémiotique. » (Eco, 1988, p. 10),

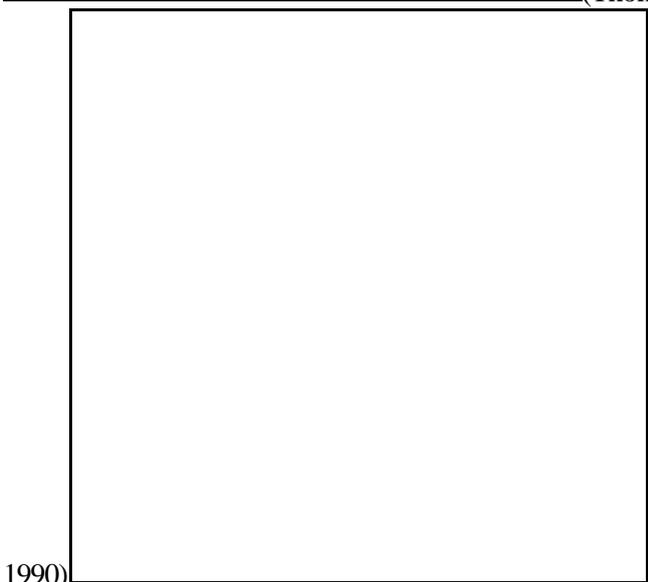
²⁹ A propos de la question de la réduction. Il est toujours tentant de chercher la réduction méthodologique, comme par exemple la réduction de la conscience au réseau neuronal-hormonal, puis aux étages cellulaire, corpusculaire et ainsi de suite

Le détour par A. Damasio visait à insister sur le fait qu'il existe un mécanisme complexe que l'on résume par le terme de «moi biologique» (bien qu'il soit non localisé et sans rapport avec l'idée d'un *homoncule*), et dont l'un des effets est d'opérer en permanence un contrôle sur les enchaînements associatifs ou déductifs de la pensée.

jusqu'à l'antimatière. Mais à chaque étage de réduction, il y a *disparition* de ce qu'on cherche à expliquer et l'intérêt de la démarche rencontre ici ses limites. Ce qui n'enlève rien à l'intuition que tout phénomène est le carrefour ou l'aboutissement d'un nombre (souvent non calculable) de causes, ou que des phénomènes locaux dans des domaines distincts peuvent être rapportés à des lois générales communes (e.g., la théorie des catastrophes de Thom et Zeeman



(Thom,

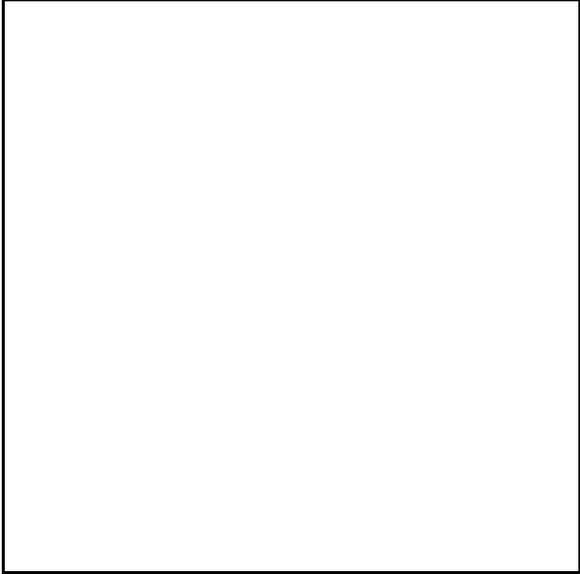


1990)

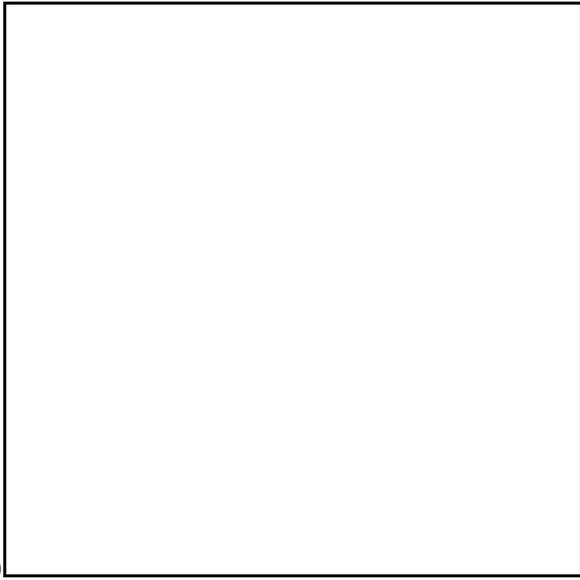
).

La narration biologique dont parle A. Damasio est compulsive et s'actualise par la modification corporelle *hic et nunc*, alors que la narration biographique, dont parle notamment Dennett (cf. cit. *supra*) est historique et concerne l'histoire du sujet, *i.e.* l'interprétation de son passé et de son avenir. Sans cette narration, le sujet n'est pas.

C'est le même moi biologique qui «lit» les émotions sur la périphérie du corps avant de les livrer au processus de sémiotisation. Sans l'accession au symbolique, c'est-à-dire sans sa formulation verbale, l'émotion — signe ou message corporel — reste indifférenciée, c'est-à-dire inconsciente : l'être organisé est agi par elle, en tant qu'il *est* elle. Lorsque l'émotion est symbolisée, en d'autres termes liée à des signifiants verbaux qui visent à en rendre compte, elle accède à un autre statut et devient constitutive du sujet, passant ainsi, par le langage, de la *physis* à la *sémiosis*. Accédant ainsi au réseau mnésique, constitutif du sujet



(Jorion, 1990 ; Vincent,



1996), elle active, au sens de l'influx nerveux, d'autres signifiants qui, par cet effet même détermineront des modifications au niveau corporel. Il ne saurait en effet y avoir de pensées, sauf à reprendre la thèse dualiste, qui ne modifient peu ou prou les connexions neurales qu'elles mettent en jeu et produisent de nouveaux états corporels. D'où l'actualisation par le sujet d'émotions déclenchées par les signes qu'il produit et les mots qu'il prononce, oralement à des interlocuteurs ou non oralement dans son discours intérieur.

Le mécanisme qui préside à la constitution du sujet, dont les essais de modélisation notamment de Dennett tentent de rendre compte, opère le passage de l'instantané du corps (l'inconscient physiologique de la vie animale) à la narration biographique, l'histoire du sujet (donnant lieu à l'inconscient psychologique). Le sujet qui a rompu avec sa propre histoire, par amnésie accidentelle ou catastrophe psychotique, «flotte» dans son corps et doit retrouver sa biographie pour recouvrer son autonomie³⁰.

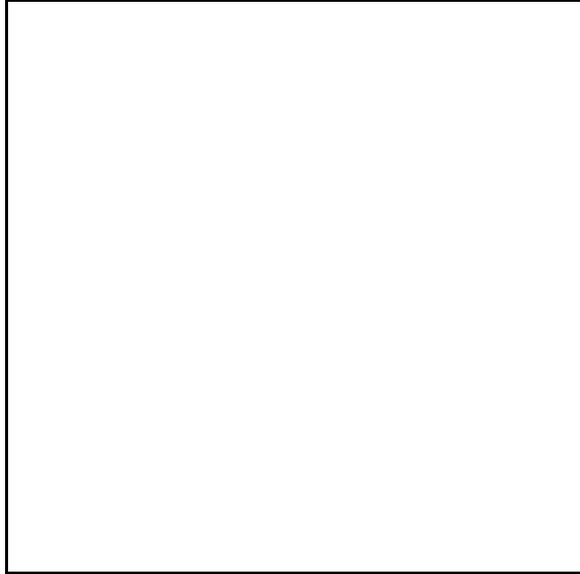
II.1.4 Le sujet dans l'espace sémiotique et le problème de la transcendance

L'espace sémiotique correspond à l'étage d'organisation qui vient immédiatement après le sujet — en tant qu'être biologique doté de la capacité linguistique — dans les étages de complexité de l'organisation naturelle (corpuscules élémentaires, atomes, cellules, corps humain, société). C'est l'espace de déploiement des systèmes de signes qui

³⁰ Le signe distinctif du psychotique est la rupture semble-t-il quasi-définitive avec son autobiographie et l'impossibilité quasi complète de la reconstruire.

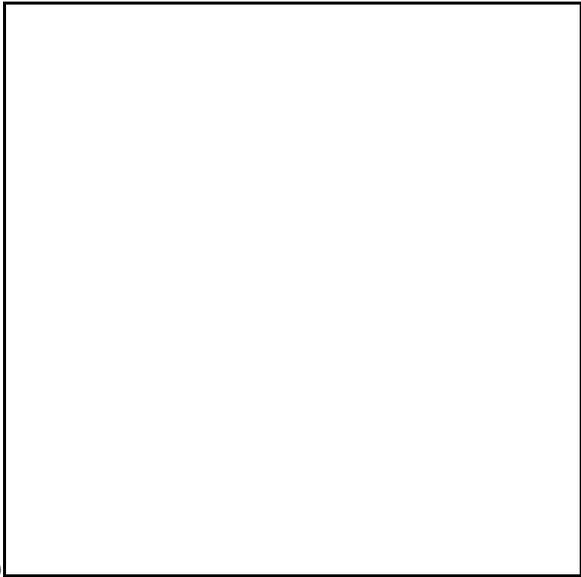
préexistent et s'imposent au sujet, comme s'imposent à lui les conditions matérielles (*Welt*) entourant sa conception, sa naissance³¹,

³¹ « L'humain naît dans une langue - que l'enfant à naître entend déjà, réverbérée, à travers le liquide amniotique - au milieu de signes, entouré d'artefacts. C'est dans la sémiotisation et par la sémiotisation que l'individu biologique se transforme en



sujet humain situé»

(Rastier,



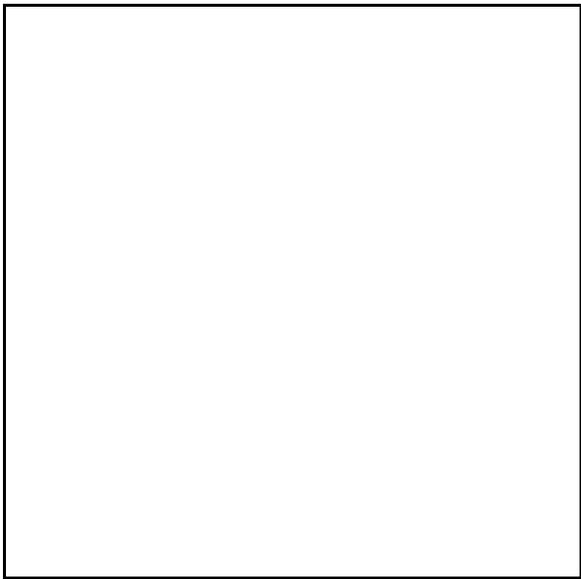
1996)

« Le langage est simplement un lieu privilégié de l'activité interprétative que nous déployons pour constituer et modifier notre entour. En d'autres termes le langage est une partie du monde où nous vivons »

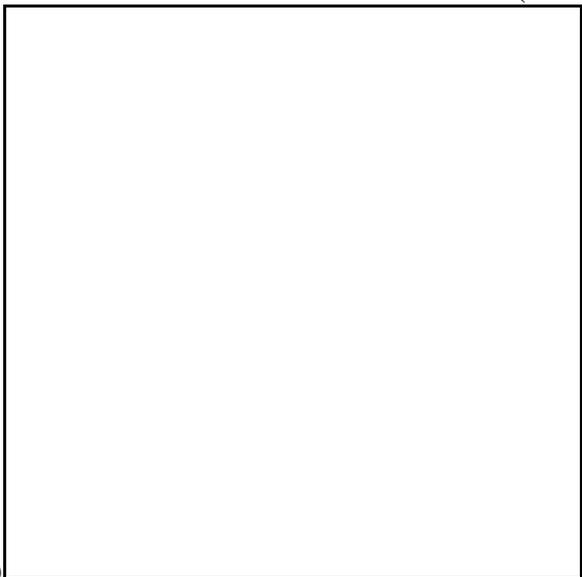
et son existence. Les mots sont situés *entre* les agents, avant d'être inscrits *dans* le système cognitif du sujet. Cet entour sémiotique fait du sujet :

a) un interprète : le lieu du processus d'interprétation qui est à l'origine de l'ensemble des pratiques sociales ;

b) un générateur de diversité et d'invention³² : le lieu de l'imprévisible rendant toute prospective sociale, politique ou scientifique hautement incertaine ;



(Rastier,



1996)

³² « (...) on a un signe seulement quand une expression s'inscrit immédiatement dans un rapport triadique, où le troisième terme, l'interprétant, génère

c) un point de relation du social et du biologique (cf. supra), rendant possible une large gamme des devenir, des meilleurs aux pires, aussi bien du côté individuel-biologique que collectif-social.

Les points *b* et *c* renvoient directement au problème de l'ontologie du sujet, que rencontrent aujourd'hui avec acuité (sous ses divers modes : conscience, intentionnalité, subjectivité...) les sciences cognitives aussi bien cognitivistes que constructivistes. Cette rencontre n'est pas fortuite, elle est le *pot* autour duquel ont tourné, faute de pouvoir ébranler les fondements du positivisme, la psychologie cognitive en particulier, et la plupart des sciences humaines depuis le siècle dernier.

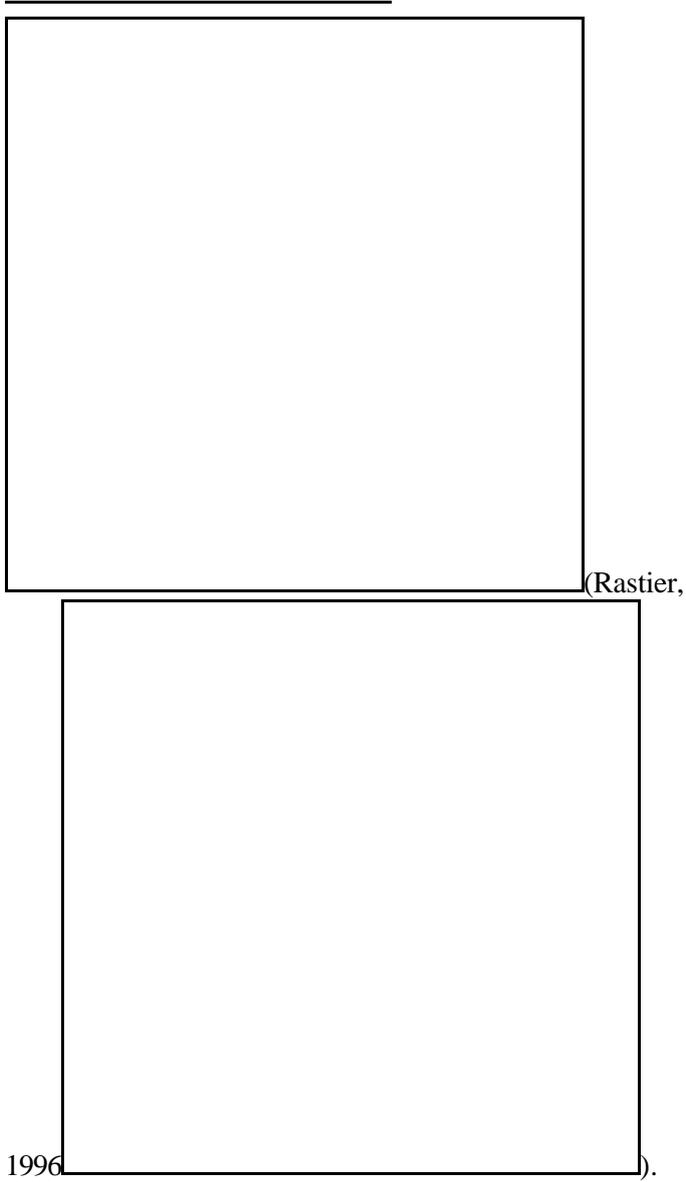
Toutefois, l'ontologie du sujet ne devrait pas, sous peine de rester l'apanage des spiritualistes, être synonyme de recours à la transcendance. Une ontologie non transcendantale du sujet apparaît plutôt comme un point essentiel mais non unique vers lequel convergent aujourd'hui avec prudence les sciences de la cognition, et dont témoigne l'effervescence actuelle du débat. Mais cette ontologie rencontre depuis le siècle dernier l'écueil du réductionnisme : l'étage physico-chimique est impuissant à expliquer l'esprit ou le sujet (ces mots recouvrant les mêmes indéterminations) comme le voudrait le positivisme. En d'autres termes, il est impossible de réifier ce qui n'est pas de l'ordre des objets physiques³³.

automatiquement une nouvelle interprétation, et ce à l'infini. » (Peirce, cité par Eco (1988, p. 14)).

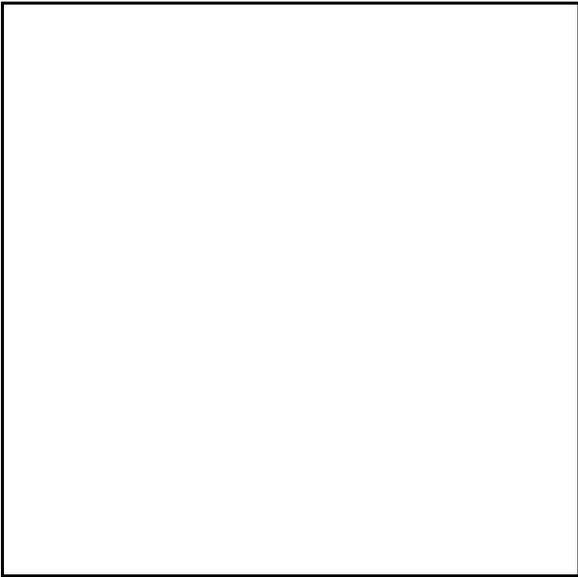
³³ «L'ontologie non transcendantale» préconisée ici n'est pas très éloignée de la «stratégie non ontologique» proposée par Rastier : « au lieu, pour ainsi dire de creuser sur place les représentations pour parvenir à leur substrat physique, il nous semble qu'une stratégie non-ontologique est possible. Elle engage, non plus sur l'axe "vertical" de la représentation, mais sur l'axe "horizontal" de l'interprétation (...), à partir en quête des trois zones de l'entour pour restituer leurs frontières, leurs transitions et leur unité. »

II.2.L'espace symbolique de l'IA classique *versus* l'espace sémiotique

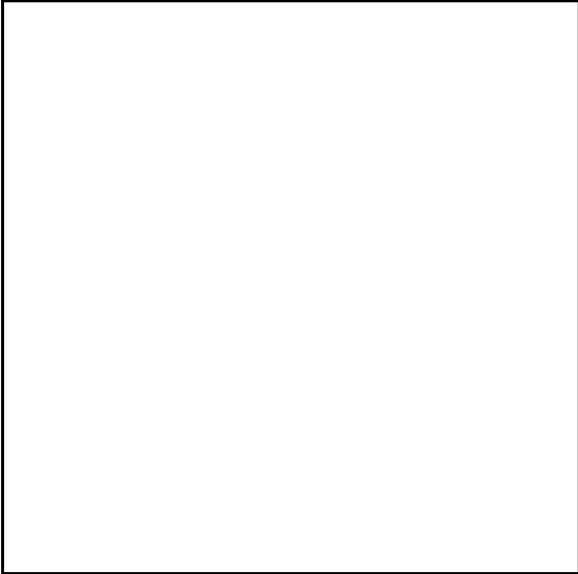
Une connaissance en tant que telle n'advient *que* dans l'espace sémiotique, mais doit se détacher, pour se singulariser, comme une forme sur un fond d'automatismes. Sinon elle se ramène soit à un *habitus* social ou linguistique, dont P. Bourdieu analyse précisément



les mécanismes

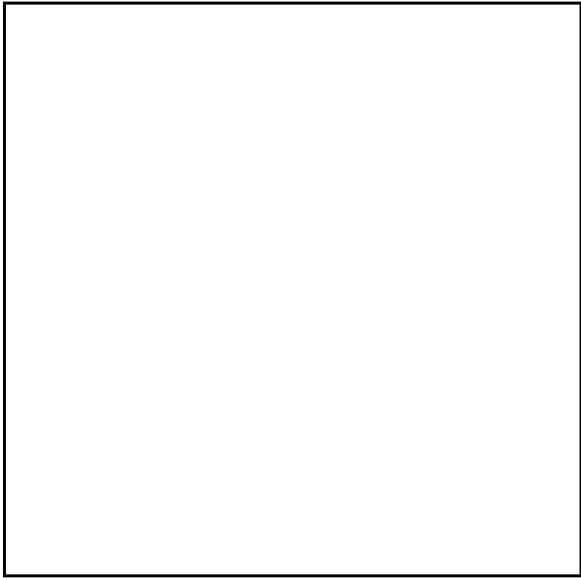


(Bourdieu,

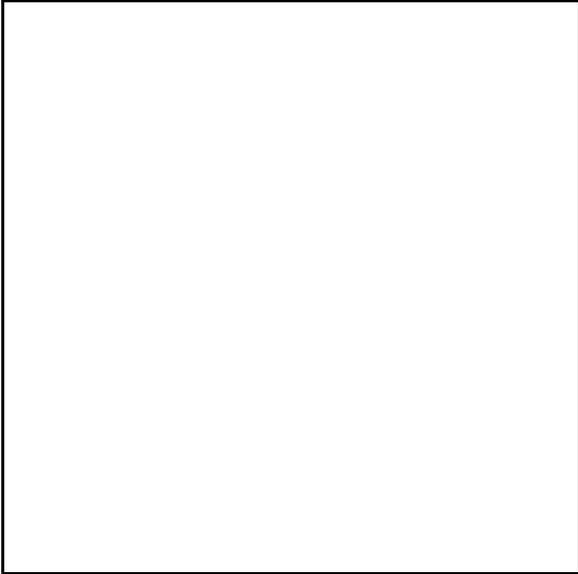


1982), soit à un *réflexe*,
(e.g., *savoir lire*, écrire, conduire une voiture...). Mais cet espace
sémiotique est un espace d'interprétation, et *n'est pas*, comme nous
allons le voir, l'espace symbolique de l'IA classique.

Pour le cognitivisme, cet espace symbolique est le lieu d'où sont
énoncés les principes premiers, tels que le *principe de raison
suffisante* ou les axiomes des théories formelles, qui conduiront à ce
que Newell appelle des *théories unifiées de la cognition*



(Newell,



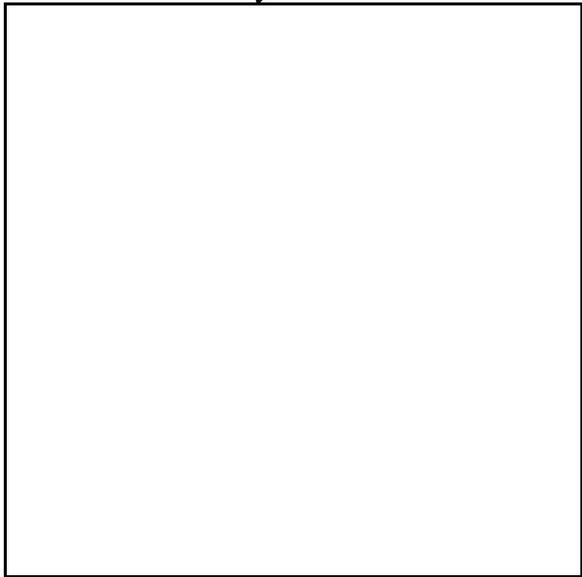
1990)

C'est donc un espace transcendant, lieu d'un absolu théorique (les conditions de complétude et de consistance y sont remplies) et premier à toute formulation ultérieure d'un système intelligent. Ces théories unifiées, bien qu'irréductibles aux deux théories de la physique que sont la mécanique quantique et la relativité généralisée, semblent partager avec celles-ci un *a priori* théorique commun : le principe de raison suffisante et l'impossibilité corrélatrice, selon Prigogine et Stengers, d'expliquer l'unidirectionnalité du temps :

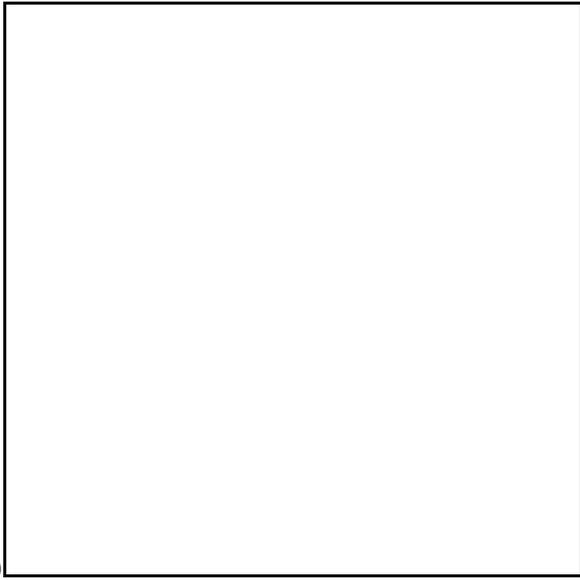
« L'impossibilité de définir une différence intrinsèque entre l'avant et l'après, à laquelle est vouée la dynamique, nous est aujourd'hui évidente, mais dès l'origine, elle était à la fois affirmée et dissimulée par un principe qui, avec

les travaux de Galilée, Huyghens, Leibniz, Euler et Lagrange, est devenu le principe même de conceptualisation de la dynamique. Leibniz a baptisé ce principe *principe de la raison suffisante*. En termes leibniziens ce principe énonce l'équivalence entre la cause « pleine » et l'effet « entier » » (Prigogine et Stengers, 1992, p. 26).

Si les mêmes conditions produiront toujours, *in illo tempore*, les mêmes conséquences, des conditions initiales ne se *reproduiront* jamais à l'identique dans le monde physique, tant au niveau quantique qu'aux au niveaux biologique-neuronal et social. La théorie du chaos déterministe enfonce définitivement ce coin, qui indique que deux états initiaux aussi proches que l'on puisse concevoir donneront lieu à des trajectoires divergentes dans l'espace de phases d'un système chaotique. Des systèmes chaotiques se rencontrent en particulier dans le système nerveux central



(Demongeot et al.,



1996) et donc dans le *substratum* de la pensée.

Le symbole cognitiviste (la cause *pleine*, l'effet *entier*) ne rencontre jamais complètement (c'est-à-dire dans une *équivalence forte*) l'existence physique, et l'aporie des théories cognitives computationnelles fortes (la *strong equivalence* de Pylyshyn) dans l'explication de l'esprit humain n'est résolue qu'au prix d'un *décret* de coïncidence (*matching*), c'est-à-dire d'un trou théorique entre la «représentation» théorique et la «présentation» neuronale et sémiotique³⁴. A partir de cette rupture, la théorie formelle se déploie dans son espace de symboles et d'opérateurs théoriques purs et dessine un esprit intemporel, dont la réalité physique et humaine est une *instance imparfaite*³⁵. Le dualisme cartésien est inhérent à l'*a priori* du système de symboles physiques, et la sortie du dualisme, qui reste une sorte de devoir pour le scientifique moderne, se fait par le biais du robot. L'équivalence forte devra alors être installée au cœur de

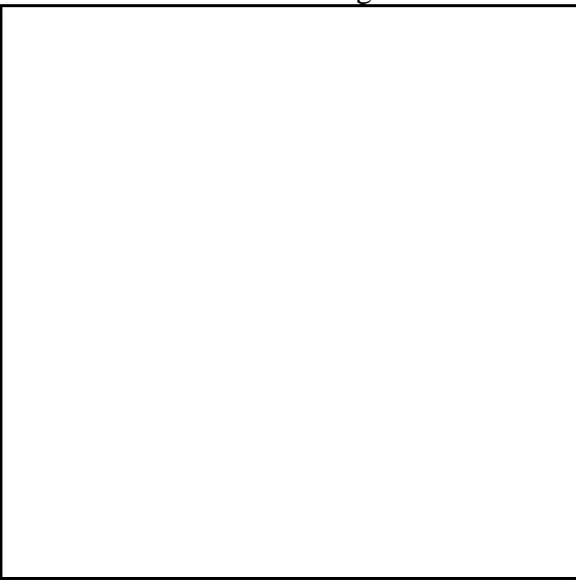
³⁴ Peirce lui-même ne fait pas du symbole un objet entièrement séparé «des cas existants de ce que le symbole dénote» : «(...) ce qui est général à son être dans les cas particuliers qu'il détermine. Il doit donc y avoir des cas existants de ce que le symbole dénote, bien qu'il faille comprendre ici par «existant», existant dans l'univers, qui peut être imaginaire, auquel le symbole renvoie. Le symbole sera, indirectement, par l'association ou une autre loi, affecté par ces cas particuliers.» (Peirce, 1978, p. 141).

³⁵ B. Bachimont (1996) développe une thèse similaire en distinguant deux principes irréductibles - celui de modélisation des connaissances, qui réfère aux objets de la langue, et celui d'effectivité, qui réfère aux processus d'automatisation des raisonnements -, et met en évidence l'aporie de l'IA classique quand au principe de modélisation.

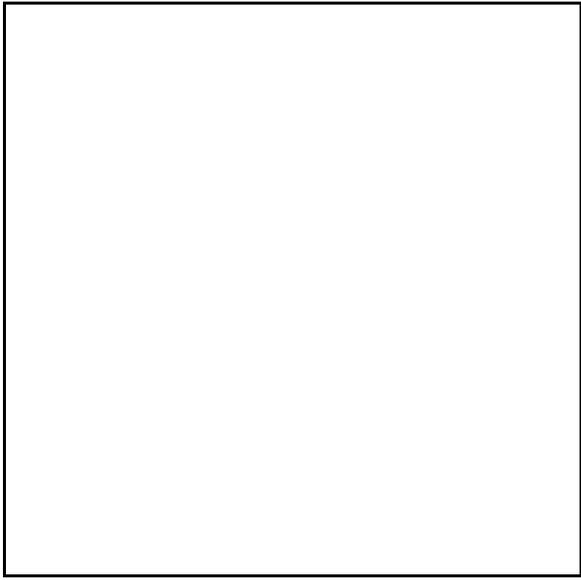
l'homme par des ingénieurs, qui le libèreront ainsi *ad vitam æternam* du joug de l'incertain, de l'imprévisible, de l'incomplétude (en un mot de son *imperfection*) qui font les souffrances de la vie, et le rendront à la paix éternelle :

« Mieux nous connaissons notre cerveau, plus nous pourrions l'améliorer (...). À mesure que ces inventions s'imposeront, nous les relierons à nos cerveaux, en insérant des milliers d'électrodes microscopiques dans le corps calleux, grand fuseau nerveux où circulent les données du cerveau. Toutes les parties du cerveau seront équipées de nouveaux accessoires. Enfin, nous remplacerons toutes les parties du corps et du cerveau et nous remédierons à tous les défauts et à toutes les blessures qui écourtent notre vie ». (Minsky, 1994, p. 123).

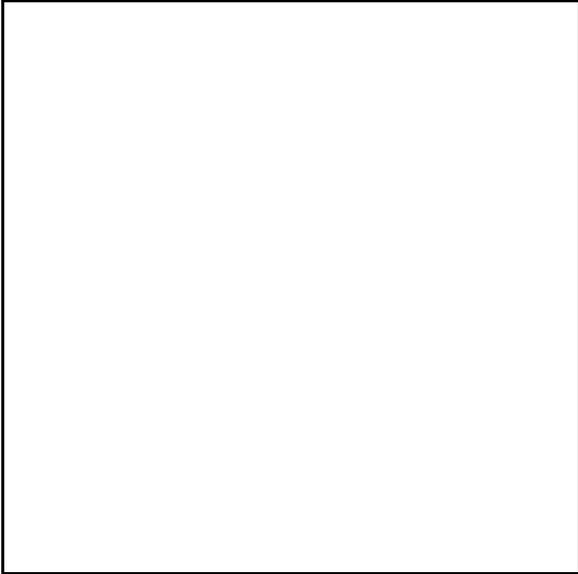
« Laisserons-nous la terre à des robots ? Oui, mais ils seront nos enfants. Nous devons nos esprits à la vie et à la mort de toutes les créatures qui nous ont précédés dans la lutte nommée évolution. Tout cet effort ne doit pas se transformer en un immense gâchis : c'est notre mission. » (Minsky,

1994,  p. 126).

A *contrario* de cet espace théorique conduisant à l'impasse ontologique que constitue le *robot* humain, l'espace sémiotique auquel nous référons est celui du *sujet* humain, fait de processus cognitifs possédant *tous* un *horizon temporel*, et dont les mécanismes irréductiblement non déterministes permettent de faire le lien avec, d'une part les générateurs de diversités biologiques conçus au niveau des cortex supérieurs (e.g.,



(Dehaene et Changeux,



1989)), d'autre part les effets de sortie du champ formel que constitue en particulier la métaphore³⁶ — véritable générateur de diversité sémiotique — dans l'usage de la langue.

³⁶ « Le discours sur la métaphore tourne autour de deux options : a/ le langage est, par sa nature et par origine, métaphorique, le mécanisme de la métaphore fonde l'activité linguistique et toute règle ou convention postérieure naît afin de réduire, de discipliner (et d'appauvrir) la richesse métaphorique qui définit l'homme comme animal symbolique ; b/ la langue (et tout autre système sémiotique) est un mécanisme conventionné, régi par des règles, une machine prévisionnelle qui dit quelles sont les phrases générables, celles qui sont « bonnes » ou « correctes » ou « dotées de sens » ; la métaphore, c'est la panne de cette machine, son soubresaut,

Cet espace sémiotique, rebelle aux théories unifiées de la cognition, n'empêche pourtant pas de théoriser. C'est au contraire l'espace de construction des théories permettant d'*interpréter le réel*. Dans cet espace, le déterminisme des théories ne croît que proportionnellement à la réduction de l'*horizon temporel* dont parle Prigogine et Stengers ou de la diminution du nombre de paramètres causaux pris en compte; les théories larges deviennent qualitatives ou spéculatives, au sens où elles ne peuvent être calculatoires ou décidables par un mécanisme formel.

Une autre distinction de l'espace sémiotique par rapport à l'espace symbolique formel est sa fonction de mémoire collective. Même si les mots trouvent *in fine* leur sens hors de l'espace des mots et dans l'espace corporel des sujets, ils ne remplissent leur fonction essentielle qu'au cours d'un échange (le discours soliloque est un effet secondaire). Ils sont appris par le jeune enfant au cours d'un processus déictique qui n'est autre qu'une *transmission*. La cognition spécifiquement humaine est rendue possible par l'*internalisation* d'un langage préexistant au sujet, mais ce langage n'a pas pour autant de transcendance : il est un cumul de codes et de savoirs³⁷ (toujours considéré comme un patrimoine) hérités à la naissance puis modifiés par chaque sujet. En retour, une fois «installé» dans le corps, le langage joue un rôle analogue à celui du fixateur dans le procédé de développement photographique. Par exemple, le rêve dont se souvient une personne au réveil n'a, sans sa formulation verbale (le fixateur), qu'une durée de vie de quelques minutes, voire de quelques secondes. Ce n'est que lorsqu'il est *formulé* qu'il se fixe à plus long terme. L'usage de la langue prolonge alors les capacités mnésiques individuelles par des capacités d'ordre mnémonique. L'espace sémiotique n'est donc pas seulement l'ensemble des systèmes de symboles, il est tout autant le lieu de la mémoire, d'abord inter-individuelle et culturelle dans l'espace social, puis individuelle et autobiographique chez le sujet.

II.3. La narrativité, du biologique au sémiotique et à l'identitaire

l'issue inexplicable, et dans le même temps, le moteur de son renouvellement. » (Eco, 1988, p. 140).

³⁷ « La vie de la culture est la vie de textes gouvernés par des lois intertextuelles où tout «déjà dit» agit comme règle possible. Le *déjà-dit* constitue le trésor de l'encyclopédie. » (Eco, 1988, p. 273).

Un mot est défini par d'autres mots sans nécessité de référer directement au dénoté : c'est là semble-t-il qu'apparaît la spécificité du langage humain par rapport aux pré-langages des animaux. Cette autonomie des signifiés par rapport aux référents n'est toutefois que transitoire. Tôt ou tard la chaîne des associations doit parvenir au corps du sujet qui effectue ces associations (*i.e.*, au moi biologique) sous peine de ne plus être *situé* et basculer dans des processus d'ordre psychotique.

Dans une formulation à rebours, l'accès au symbolique n'est pas un *plus* réservé aux êtres humains par rapport aux animaux, c'est une condition d'émergence du «je» dans le prolongement du processus narratif biologique. Nous parlons d'*émergence* et non pas d'*existence* du «je», car le «je» n'existe pas en tant que *substance* ; il n'est repérable qu'au cours d'un *processus* narratif, donc dynamique, lié notamment à la capacité cognitive de construire des signes linguistiques dans un espace social et historique :

« (...) le signe comme moment (toujours en crise) du processus de sémiotique est l'instrument par lequel le sujet lui-même se construit et se déconstruit perpétuellement. Le sujet connaît une crise bénéfique parce qu'elle participe de la crise historique (et constitutive) du signe. Le sujet est ce que les processus continuels de resegmentation du contenu le font être. En ce sens (même si le processus de resegmentation doit être forcément opéré par quelqu'un, même si l'on se doute bien que c'est toujours une collectivité de sujets) le sujet est parlé par les langages (verbaux ou non), pas par la chaîne signifiante mais par la dynamique des fonctions sémiotiques. Nous sommes, en tant que sujets, ce que la forme du monde produite par les signes nous fait être » (Eco, 1988, p. 61).

La langue dans laquelle s'exprime le sujet est le seul moyen qu'il possède pour *advenir*³⁸. L'identité d'une personne, au-delà de la déclinaison du code de sécurité sociale, est par principe non entièrement définissable (connaissable), ni par elle ni par d'autres qu'elle-même. La recherche d'une réponse à la célèbre question ontologique «qui suis-je ?» s'accompagne nécessairement d'une modification de l'identité du questionneur par le processus narratif, intérieur ou extériorisé, de la tentative de réponse. On ne peut appréhender l'identité du sujet que dans une dynamique : cette identité existe bel et bien puisqu'elle est la subjectivité même qui fonde le sujet et toutes les lois de protection de l'individu. Mais elle n'existe pas dans

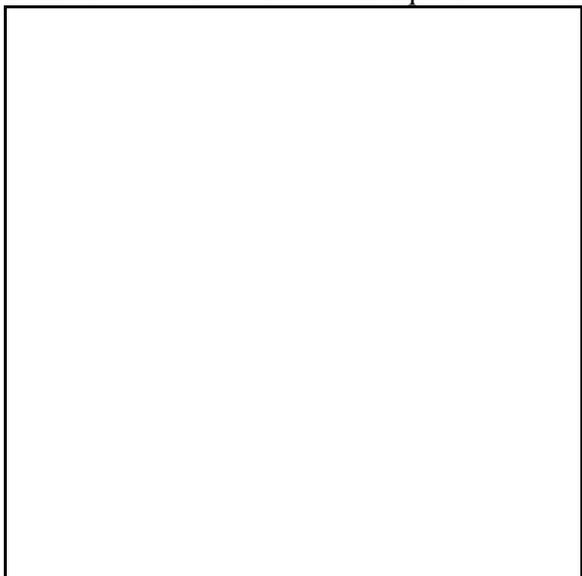
³⁸ « Ainsi, mon langage est la somme totale de moi-même, puisque l'homme est la pensée. » (C.S. Peirce, cité par Eco (1988), p. 62).

l'instantané. Il faut l'axe du temps, c'est-à-dire le *processus*, pour l'appréhender. Axe sur lequel s'inscrivent en-deçà de la fameuse barrière des 400 millisecondes les processus du moi biologique et au-delà les processus de la conscience³⁹.

CONCLUSION

L'enjeu d'un tournant que l'on pourrait qualifier de *sémiotique* de l'IA est la sortie de la *phase naïve* de l'IA selon laquelle la pensée humaine pourrait être modélisée par un système calculatoire sur des symboles-types, immuables, intemporels, en nombre fini, et dont les représentations mentales (les symboles-occurrences) sont des implémentations. Entre le point de vue cognitiviste qui cherche en permanence à reculer les bornes de sa naïveté, et les points de vue philosophique et neuropsychologique qui ne s'embarrassent d'aucune contrainte computationnelle, la question se pose de nouveau avec insistance de savoir quel est le sens de l'intelligence artificielle vis-à-vis de la compréhension de l'esprit humain, puisque telle est officiellement, par ses racines cognitivistes, l'une de ses vocations. Répondre rapidement à cette question revient à prendre des risques considérables qui consistent, d'un côté à rejeter en bloc l'automatisation de la pensée, la renvoyant dans le monde d'un esprit à la fois séparé et pilote de la matière, de l'autre côté à nier le monde à jamais non-directement observable, non-contrôlable, non-reproductible

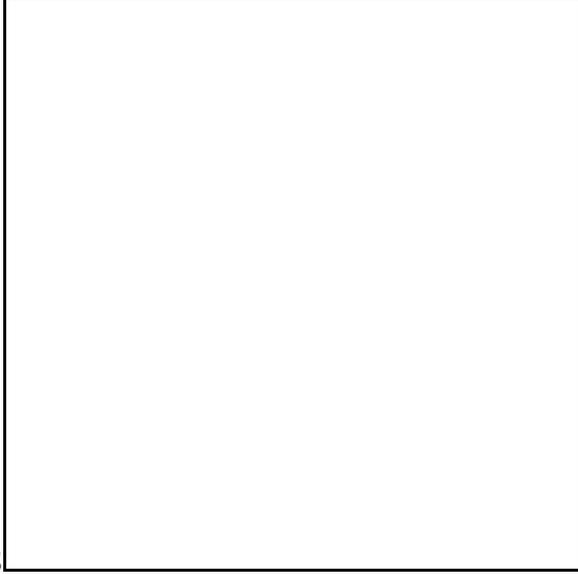
³⁹ Barrière fallacieuse en ce que les raisonnements les plus complexes, ceux qui mettent en jeu l'existence et le devenir social du sujet, sont fondés selon A. Damasio sur les marqueurs somatiques du corps (cf.



(Damasio, 1995, p. 215 et suiv.)).

au sens positiviste du terme, de la subjectivité, de l'expérience personnelle, des productions mentales a-perceptives, etc., et de considérer avec Minsky que l'automate humain, réplique parfaite (au sens de libérée de la souffrance) de l'homme, est au bout de la route de l'IA.

L'IA entretient, depuis ses origines, des liens étroits avec la philosophie, les neurosciences, la psychologie expérimentale et jusqu'à la psychanalyse (e.g. : Jorion, 1987 ; Colby et Stoller,



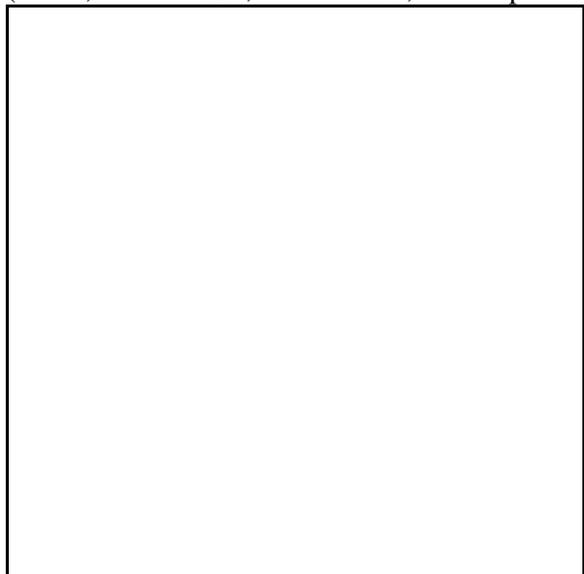
1988). Elle est de toute évidence utilisée comme un moyen de premier plan pour montrer soit ce qu'est l'«esprit» humain, soit ce qu'il n'est pas par contraste avec ce qu'elle propose. Pourtant, rien n'est plus douteux que de donner à l'IA la mission d'expliquer et modéliser l'esprit humain. De plus, elle reste pure spéculation sans la programmation et l'implémentation. La question se pose alors concrètement de trouver une voie non maximaliste et efficace de l'IA. Une voie qui évite de faire de l'IA le fer de lance d'un ontologisme simpliste, concluant ce qu'il avait posé en prémisses, à savoir qu'un esprit-programme originel organise son entour en se complexifiant lui-même au gré de son apprentissage permanent. De cet ontologisme est exclue la dimension anthropologique qui étudie le sujet humain dans ses relations, complexes et irréductibles, aux autres sujets humains⁴⁰ et fonde sa forme spécifique d'intelligence.

⁴⁰ « L'espèce humaine se caractérise par un développement singulier des relations sociales, que soutiennent des capacités exceptionnelles de communication mentale, et corrélativement une économie paradoxale des instincts qui s'y montrent entièrement susceptibles de conversion et d'inversion et n'ont plus d'effets

Une voie naturelle pour l'IA consiste alors à prolonger nos capacités de compréhension et d'action par l'outil informatique, auquel l'IA en tant qu'algorithmique avancée contribue autant que la programmation parallèle, la conception des réseaux, l'imagerie et la réalité virtuelles, etc. Mais le fin mot de l'histoire n'est certainement pas dans cette «bonne foi» dans l'outil, puisqu'un outil puissant n'a jamais fonctionné sans une réorganisation sociale.

Wittgenstein disait qu'il faut faire un pas analogue à celui de la théorie de la relativité pour intégrer dans nos conceptions ce qui nous fonde⁴¹. L'IA devrait alors y aider en tant que moyen d'aide à la conceptualisation-simulation de systèmes complexes, dont le système complexe bio-sociologique est celui qui nous concerne le plus, et dans lequel nous intervenons par l'IA même, dans ce qu'Hofstadter nomme une «boucle étrange». A ce titre, l'IA et l'ordinateur peuvent être considérés comme un domaine technique produit par une sémio-culture particulière, celle de notre société occidentale, pour tenter de trouver une homéostasie qu'elle n'a pas encore atteinte.

isolables que de façon sporadique. Des comportements adaptatifs d'une variété infinie sont ainsi permis et constituent la culture ; celle-ci introduit une nouvelle dimension dans la réalité sociale et dans la vie psychique. Cette dimension spécifie la famille humaine comme, du reste, tous les phénomènes sociaux chez l'homme. » (Lacan, 1938, 1984, p. 11-12 ; cité par



(Ogilvie, 1987, p. 69-70)).

⁴¹ « Ce qu'il faut à *nouveau* ici, c'est un pas, analogue à celui que fait la théorie de la relativité. » (Wittgenstein, 1969, 1976, p. 83).

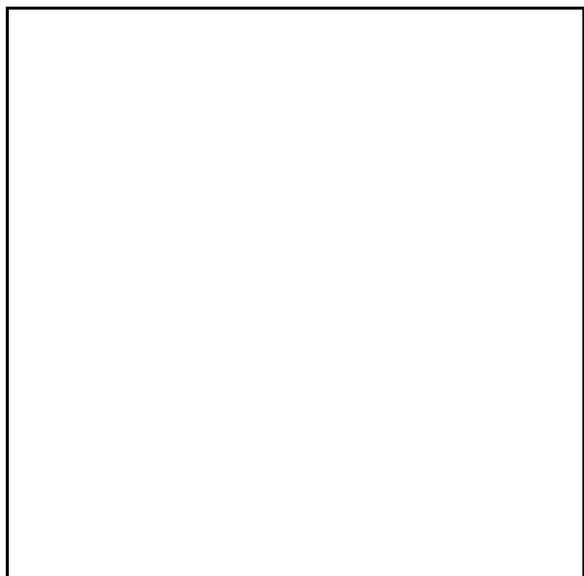
Remerciements. Plusieurs personnes m'ont apporté un soutien efficace et chaleureux par leurs critiques ou leurs corrections. Je les remercie vivement, à commencer par les membres du groupe *Théorie et Clinique des Pathologies de la Pensée* de la Maison des Sciences de l'Homme à Paris. Merci également au Pr. Denis Vernant dont je n'ai pu épuiser la richesse des remarques. Merci enfin à ma femme, Christine Bigallet, pour sa relecture attentive et quelques éclairages théoriques.

Bibliographie

- Agre P.E. (1993) Interview with Allen Newell, in Bobrow D. G. (éd.) *Artificial Intelligence in Perspective (special issue of Artificial Intelligence)*, MIT Press/Bradford Book. Cambridge, MA, pp. 415-449.
- Bachimont B. (1993) Nature, Culture et Artefacture : la place de l'intelligence artificielle dans les sciences cognitives, *Intellectica*, 17, 2, pp. 213-238.
- Bachimont B. (1996) Herméneutique matérielle : des machines qui pensent aux machines qui donnent à penser : critique des formalismes en intelligence artificielle, Thèse de Doctorat, École Polytechnique, Paris.
- Badiou A. (1982) *Théorie du sujet*, Le Seuil. Paris.
- Block N., Flanagan O. et Güzeldere G. (éd.) (1996) *The Nature of Consciousness*, MIT Press/Bradford Book, Cambridge, MA.
- Bourdieu P. (1982) *Ce que parler veut dire : l'économie des échanges linguistiques*, Fayard. Paris.
- Brooks R.A. (1991) Intelligence without representation, *Artificial Intelligence*, 47, pp. 139-159.
- Chalmers D. (1996) Qu'est-ce que la conscience ?, *Pour la Science*, 220, pp. 58-64.
- Colby K.M. et Stoller R.J. (1988) *Cognitive Science and Psychoanalysis*, Lawrence Erlbaum Associates. Hillsdale, New Jersey.
- Crick F. (1994) *L'hypothèse stupéfiante : À la recherche scientifique de l'âme*. Paris: Plon ; tr. H. Prouteau, *The Astonishing Hypothesis*, 1994, New York: Macmillan-Publ.-Co.
- Damasio A.R. (1995) *L'Erreur de Descartes : La raison des émotions*. Paris: Ed. Odile Jacob ; tr. M. Blanc, *Descartes'Error : Emotion, Reason, and the Human Brain*, 1994, Putnam-Books.
- Dehaene S. et Changeux J.P. (1989) A simple Model of Prefrontal Cortex Function in Delayed-Response Tasks, *Journal of Cognitive Neuroscience*, 1, 3, pp. 244-261.
- Demongeot J., Nérot O. et Jézéquel C. (1996) Mémoire de rappel dans les réseaux de neurones, in Rialle V. et Fisette D. (éd.) *Penser l'esprit : des sciences de la cognition à une philosophie cognitive*, Grenoble, France, pp. 407-419.

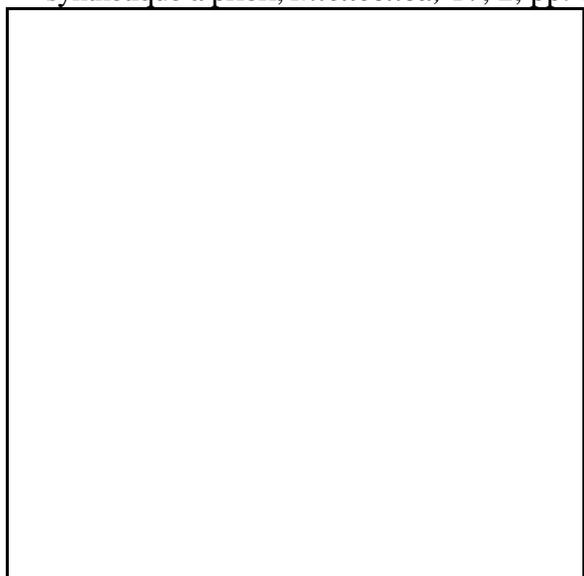
- Dennett D.C. (1988) *The intentional stance*, Bradford Book/MIT Press. Cambridge, MA.
- Dennett D.C. (1993) Allen Newell, Unified Theories of Cognition, *Artificial Intelligence*, 59, 1-2, pp. 285-294.
- Dennett D.C. (1993) *La conscience expliquée*. Paris: Odile Jacob ; tr. P. Engel, *Consciousness explained*, 1991, Boston: Little-Brown-and-Co.
- Desclés J.P. (1991) *Langages applicatifs, langues naturelles et cognition*, Hermès. Paris.
- Dreyfus H.L. (1993) La critique heideggerienne et l'approche husserlienne et searlienienne de l'intentionnalité, *Intellectica*, 17, 2, pp. 27-49.
- Ducrot J.J. (1984) *Jean Piaget, savant et philosophe. Essai sur la formation des connaissances et du sujet de la connaissance*, Droz. Genève.
- Eco U. (1988) *Sémiotique et philosophie du langage*, Presses Universitaires de France. Paris.
- Edelman G.M. (1992) *La biologie de la conscience*. Paris: Odile Jacob ; tr. A. Gerschenfeld, *The Remembered Present : A biological Theory of Consciousness*, 1989, New York: Basic Books.
- Hameroff S.R., Kaszniak A.W. et Scott A.C. (éd.) (1996) *Toward a Science of Consciousness*, MIT Press/Bradford Book, Cambridge, MA.
- Hofstadter D.R. et Dennett D. (1987) *Vues de l'Esprit : Fantaisies et Reflexions sur l'être et l'âme*. Paris: InterEditions ; tr. J. Henry, *The Mind's Eye : Fantaisies and Reflections on Self and Soul*, 1981, New York: Basic-Books.
- Jackendoff J. (1987) *Consciousness and the Computational Mind*, MIT Press / Bradford Book. Massachusetts.
- James W. (1890, 1981) *The Principles of Psychology*, Harvard University Press.
- Jorion P. (1987) Ce que l'intelligence artificielle devra à Freud, *L'Ane*, 31, pp. 43-44.
- Jorion P. (1990) *Principes des systèmes intelligents*, Masson. Paris.
- Kaufmann P. (éd.) (1993) *L'apport freudien : Éléments pour une encyclopédie de la psychanalyse*, Bordas, Paris.
- Lacan J. (1938, 1984) Les complexes familiaux dans la formation de l'individu. Essai d'analyse de la formation d'une fonction en psychologie, in Wallon (éd.) *L'Encyclopédie française*, Navarin. Paris.

- Lacharité N. (1996) Conflit de modèles en théorie de la représentation, in Rialle V. et Fisette D. (éd.) *Penser l'esprit : des sciences de la cognition à une philosophie cognitive*, Presses Universitaires de Grenoble. Grenoble, France, pp. 125-139.
- Miermont J. (1996) L'élaboration de l'espace sémiotique, in Rialle V. et Fisette D. (éd.) *Penser l'esprit : des sciences de la cognition à une philosophie cognitive*, Presses Universitaires de Grenoble. Grenoble, France, pp. 339-355.
- Minsky M. (1986) *The Society of Mind*, New York: Simon and Schuster ; tr. Jacqueline Henry, *La Société de l'esprit*, 1986, Paris: InterEdition.
- Minsky M. (1994) Laisserons-nous la terre à des robots ?, *Pour la Science*, 206, pp. 120-126.
- Newell A. (1980) Physical Symbol Systems, *Cognitive Science*, 4, pp. 135-183.
- Newell A. (1982) The knowledge level, *Artificial Intelligence*, 18, pp. 87-127.
- Newell A. (1990) *Unified Theories of Cognition*, Harvard University Press. Cambridge, MA.
- Newell A. (1993) Reflections on the knowledge level, *Artificial Intelligence*, 59, pp. 31-38.
- Newell A., Rosenbloom P.S. et Laird J.E. (1989) Symbolic architecture for cognition, in Posner M. (éd.) *Foundation of Cognitive Science*, MIT Press. Cambridge, MA, pp. 93-132.
- Newell A. et Simon H. (1972) *Human Problem Solving*, Prentice Hall. Englewood cliffs, NJ.
- Ogilvie B. (1987) *Lacan - Le sujet*, Presses Universitaires de France. Paris.
- Pacherie E. (1996) Le fonctionnalisme : état des lieux, *Intellectica*, 21, pp. 9-37.
- Peirce C.S. (1978) *Ecrits sur le signe*, (Rassemblés, traduits et commentés par G. Deledalle), Seuil. Paris.
- Petitot J. (1992) *Physique du sens*, Ed. du CNRS. Paris.



Petitot J. (1993)

Phénoménologie naturalisée et morphodynamique : la fonction cognitive du synthétique a priori, *Intellectica*, 17, 2, pp. 79-126.



Piaget J. (1992) *Biologie et*

connaissance, Delachaux et Niestlé. Neuchâtel et Paris.

Prigogine I. et Stengers I. (1992) *Entre le temps et l'éternité*, Flammarion. Paris.

Putnam H. (1992) *Définitions : pourquoi ne peut-on pas "naturaliser" la raison*. Paris: L'Éclat ; tr. C. Bouchindhomme.

Pylyshyn Z.W. (1990) Computing in Cognitive Science, in Posner M. I. (éd.) *Foundations of Cognitive Science*, MIT Presse/Bradford Book. Cambridge, MA, pp. 52-91.

Rastier F. (éd.) (1995) *L'analyse thématique des données textuelles*, Didier Érudition, Paris.

- Rastier F. (1996) Représentation ou interprétation ? Une perspective herméneutique sur la médiation sémiotique, in Rialle V. et Fisette D. (éd.) *Penser l'esprit : des sciences de la cognition à une philosophie cognitive*, Presses Universitaires de Grenoble. Grenoble, France, pp. 239-259.
- Sabah G. (1993) Vers une conscience artificielle, in Denis M. et Sabah G. (éd.) *Modèle et concepts pour la science cognitive*, Presses Universitaire de Grenoble. Grenoble, pp. 207-222.
- Searle J. (1987) Esprits, cerveaux et programmes, in Hofstadter D. et Dennett D. (éd.) *Vues de l'esprit - Fantaisies et réflexions sur l'être et l'âme*, InterEdition. Paris, pp. 354-373.
- Searle J. (1995) *La redécouverte de l'esprit*. Paris: Gallimard ; tr. C. Tiercelin, *The rediscovery of the mind*, 1992, Cambridge, MA: MIT-Press.
- Serres M. (1991) *Le Tiers-Instruit*, Françoise Bourin. Paris.
- Simon H. (1993) Allen Newell : the entry into complex information processing, *Artificial Intelligence*, 59, pp. 251-259.
- Sperber D. (1987) Un parti pris matérialiste, *L'Ane*, 31, Propos recueillis par M.V. Howlett et J. Miller, pp. 45-48.
- Thom R. (1990) *Apologie du logos*, Hachette. France.
- Varela F., Thompson E. et Rosch E. (1993) *L'inscription corporelle de l'esprit*. Paris: Le Seuil ; tr. V. Havelange, *The Embodied Mind: Cognitive Science and Human Experience*, 1991, Cambridge, MA: MIT Press.
- Vernant D. (1994) Du discours à l'action, *L'Enseignement philosophique*, jan-fév, pp. 10-25.
- Verschure F.M.J. (1993) Formal Minds and Biological Brains : AI and Edelman's Extended Theory of Neuronal Group Selection, *IEEE Expert*, 8, 5, pp. 66-75.
- Vincent T. (1996) "Pendant que Rome brûle" : *La clinique psychanalytique de la psychose de Sullivan à Lacan*, Éditions Arcanes. Strasbourg.
- Vincent T. (1996) Psychose, Réseau mnésique, Intentionnalité, in Rialle V. et Fisette D. (éd.) *Penser l'esprit : des sciences de la cognition à une philosophie cognitive*, Presses Universitaires de Grenoble. France, pp. 357-372.
- Virole B. (1995) *Sciences Cognitives et Psychanalyse*, Presses Universitaires de Nancy. Nancy.
- Visetti Y.-M. (1996) Fonctionnalismes 1996, *Intellectica*, 21, pp. 282-311.
- Winograd T. (1993) Heidegger et la conception de systèmes informatiques, *Intellectica*, 17, 2, pp. 51-78.
- Wittgenstein L. (1969, 1976) *De la certitude*, Tel, Gallimard. Paris.