

# Perspectives et éthique des systèmes autonomes de pensée artificielle

Joël Colloc

Normandie Univ, UNIHAVRE, UNICAEN, UNIROUEN, CNRS, IDEES, 76600 Le Havre France  
joel.colloc@univ-lehavre.fr,  
WWW home page: <http://www.cirtai.org/spip.php7rubrique389>

## Résumé

La faisabilité de systèmes de pensée artificiels autonomes nécessite de comparer la façon dont les êtres humains acquièrent leurs informations et développent leur pensée avec les capacités actuelles des systèmes d'information autonomes. Notre modèle utilise quatre hiérarchies : la hiérarchie des systèmes d'information, la hiérarchie cognitive, la hiérarchie linguistique et la hiérarchie numérique de l'information qui combine l'intelligence artificielle, la puissance des modèles d'ordinateurs, des méthodes et des outils utilisés pour développer des systèmes d'information autonomes. La question de la capacité des systèmes autonomes à élaborer une forme de pensée artificielle se pose actuellement avec les conséquences éthiques sur la vie sociale et la perspective du mouvement transhumaniste.

**Mots clés :** pensée artificielle, holisme, éthique, systèmes autonomes, conscience artificielle.

## Introduction

Les progrès récents des sciences cognitives, des neurosciences, de l'informatique et de la robotique relancent le projet de la création d'un être artificiel autonome capable de penser. Cet article étudie la faisabilité d'un tel projet en comparant la manière dont les humains acquièrent, exploitent leurs informations et élaborent la pensée avec les capacités actuelles des systèmes d'information autonomes. Notre comparaison est fondée sur quatre hiérarchies : la hiérarchie des systèmes d'information, issue de la systémique, fournit un indicateur de complexité et d'autonomie; la hiérarchie cognitive décrit l'acquisition sub-symbolique et l'émergence de notre expérience personnelle, notre savoir, tandis que la hiérarchie langagière construit le discours décrivant les connaissances acquises en termes d'objets concrets et abstraits sur l'environnement et sur soi-même. La hiérarchie informationnelle numérique repose sur les concepts nécessaires, des modèles, des méthodes et des outils nécessaires à la réalisation de systèmes d'information autonomes. Enfin, nous discutons de la nature et de l'éthique de la pensée artificielle.

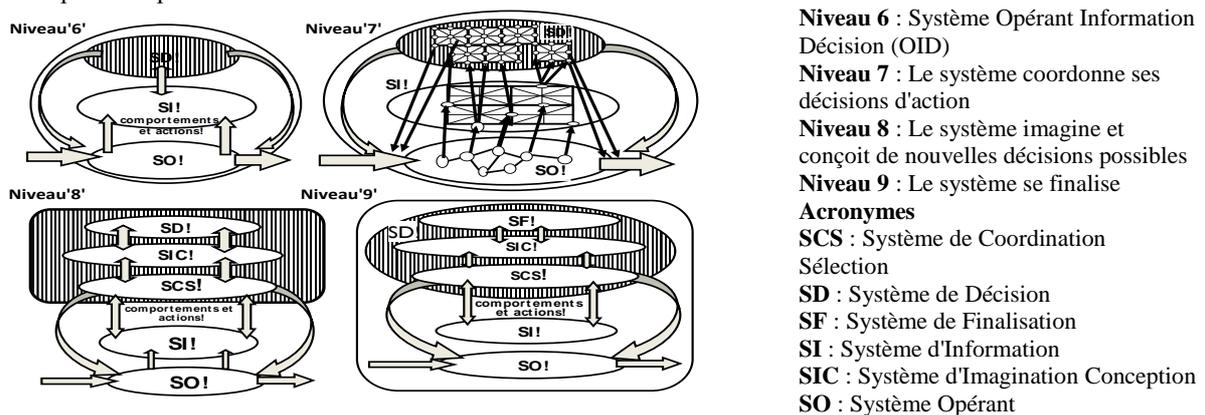


Figure 1. Les niveaux 6 à 9 du modèle de système complexe de J-L. Le Moigne © (Le Moigne, 1990)

## 2. La hiérarchie des systèmes d'information

Jean-Louis Le Moigne a décrit l'évolution des systèmes d'information. Comme dans l'épigénèse, le niveau suivant ajoute de nouvelles caractéristiques au niveau précédent. Par souci de concision, seuls les niveaux de 6 à 9 sont présentés sur la figure 1. Au niveau 6, le système devient capable de mémoriser ses décisions O.I.D. (Système opérant (SO), Système d'Information (SI), et Système de Décision (SD)).

Au niveau 7, le système coordonne de nombreuses décisions ou actions à chaque instant  $t$  concernant son activité

interne, sa régulation et ses informations externes en provenance et vers son environnement (SCS). Au niveau 8, le système est doté d'un sous-système d'imagination et de conception (SIC). Au niveau 9, le système est capable de décider sur ses propres décisions et de déterminer les aspects positifs et négatifs de ses actions. Cette finalisation d'un système complexe (SF) est proche de la pensée humaine ce qui lui confère une autonomie de décision et lui permet de choisir ses propres objectifs. Les systèmes multi-agents autonomes (SMAA) appartiennent à cette catégorie.

### 3. Intégration des voies sensorielles et motrices : émergence des objets

La neurophysiologie des voies sensibles est décrite en détail dans (Arbib & Hanson 1988)(Gazzaniga et al., 2001) et (Purves et al., 1997). Les voies sensorielles (vision, audition, olfaction, goût, toucher...) ont globalement une organisation ascendante similaire : des capteurs périphériques, une transduction en signal électrique, la transmission (+/- longue) par les nerfs dans le tronc cérébral ou par des paires de nerfs crâniens (V,VII, IX, X), le relais dans une aire thalamique spécifique, une projection dans une zone corticale spécialisée et des aires associatives communes à plusieurs modalités sensorielles. À chaque étage, il existe de nombreux neurones récurrents formant une « rétro-propagation » qui assurent les boucles de rétroaction essentielles au fonctionnement du cortex cérébral qui est fondé sur la prédiction des futures perceptions attendues comme cohérentes avec celles qui viennent d'être mémorisées et ce à chaque niveau de détails : des éléments de perception éphémères, que nous appelons percepts jusqu'aux objets mémorisés durablement (figure 2 et 3). La motricité fonctionne de manière similaire par des voies descendantes et comporte aussi des boucles de rétroaction. Le fonctionnement du cortex cérébral est plutôt uniforme sur l'ensemble de sa surface. La vision d'une pomme active l'odeur du fruit, sa forme, sa consistance (en connexion avec la motricité) et le mot « pomme » ce qui nous permet d'en parler. Inversement, sentir l'odeur d'une pomme (sans la voir) active une image stéréotypée d'une pomme tel que nous souhaiterions la voir. Toutes ces prédictions sont possibles parce que nous mémorisons une perception globale et unifiée de la pomme reliée aux différents sens, aux modalités linguistiques qui lui correspondent (Hawkins and Blakeslee, 2005), (Minsky, 1988).

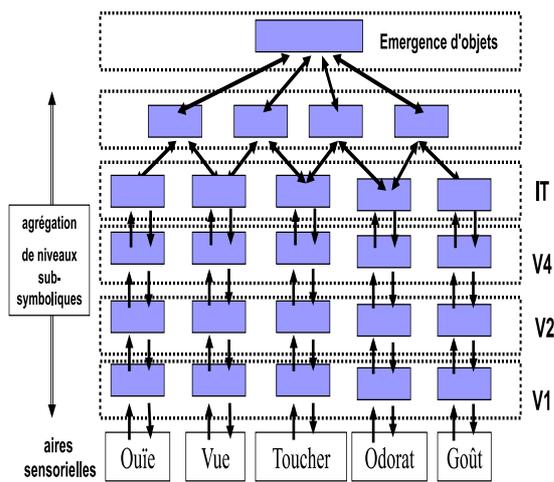


Figure 2. Émergence sensorielle d'objets connus (Hawkins et Blakeslee, 2005)

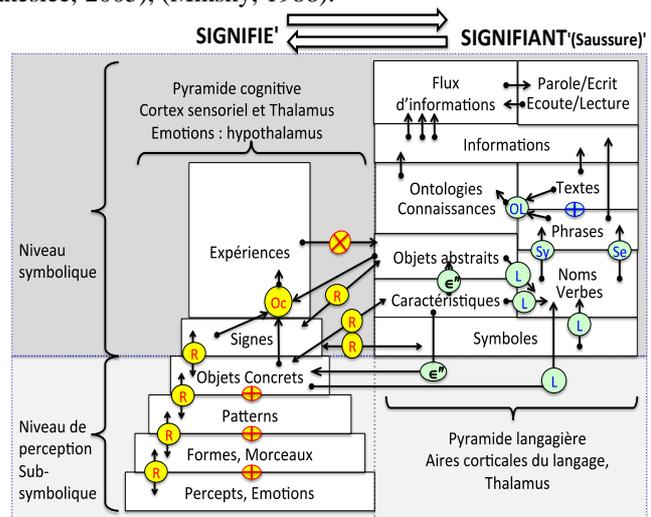


Figure 3. Double hiérarchie cognitive/langagière

La figure 2 montre les différentes couches du cortex V1, V2, V4 qui intègre les modalités sensorielles et les zones associatives du cortex inféro-temporal (IT) et qui sont activées quand une pomme est reconnue selon une des modalités. Cette activation est transmise par des voies descendantes dans les autres modalités sensorielles et les aires linguistiques qui vont pouvoir prédire et vérifier la concordance des perceptions aux différents niveaux de détails correspondants de manière de plus en plus fine pour confirmer ou infirmer qu'elles correspondent bien à celles attendues dans les autres sens. Nous pouvons ainsi savoir et distinguer si nous ressentons une vraie pomme ou seulement une représentation partielle comme dans le célèbre tableau de Magritte intitulé « Ceci n'est pas une pomme ». A cette organisation anatomophysiologique correspond une hiérarchie de l'information sensorielle transmise et intégrée dans les niveaux du cortex figure 3. L'hippocampe est situé au sommet de la pyramide mnésique et effectue des correspondances et l'interconnexion des différentes modalités sensorielles et mémorise les caractéristiques des nouveaux objets inconnus (Hawkins and Blakeslee, 2005). La figure 3 décrit un modèle abstrait du système d'information humain (Colloc, 2016). Dans la partie gauche se trouve la pyramide sensorielle et cognitive qui conduit à l'émergence des objets concrets et à la mémorisation de la connaissance et des expériences concernant les objets du monde. A chaque niveau, des opérateurs de reconnaissance ®

confirment ou infirment la cohérence dans le flux de perceptions. Les perceptions sont des éléments primaires perçus par les organes des sens. Des points et des lignes (verticales, obliques, horizontales...) établissent dans le champ de vision des formes élémentaires comme celles décrites dans la théorie des « geons » proposée par Hummel and Biederman en 1987 (Biederman, 1987), (Boucart, 1996). Les neurohormones secrétées principalement par l'hypothalamus et l'amygdale influencent l'interprétation et l'information sensorielle et régulent l'ensemble des équilibres du corps (la faim soif/ satiété, plaisir/douleur, les émotions...) (Vincent, 1996), (Vincent, 2002) (Damasio, 1995).

L'amygdale intervient dans les circuits de la récompense et de la motivation. Pour Alain Prochiantz, il n'y a pas d'organe de la pensée, « Il n'y a pas de pensée sans corps, et de corps sans pensée » (Prochiantz, 2001) ; il rejoint ainsi Francisco Varela (Varela et al., 1993).

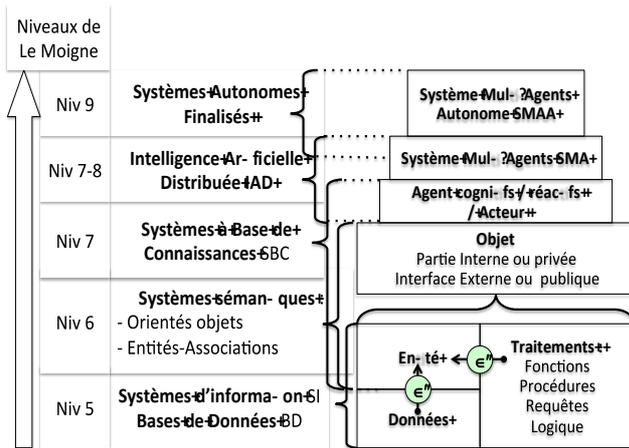
Les opérateurs cognitifs de composition  $\oplus$  et de reconnaissance (R) établissent des signes qui fondent l'expérience des formes et du comportement des objets concrets à différents niveaux successifs de composition (partie gauche de la figure 3). L'opérateur de synthèse  $\otimes$  prédit les caractéristiques ou le comportement des objets similaires et exploite les opérateurs suivants : l'opérateur de généralisation (G) qui regroupe ensemble les objets similaires dans une catégorie. L'opérateur de déduction (Dd) prédit les caractéristiques et le comportement des objets à rechercher à partir des formes et des ensembles de perceptions déjà obtenus. L'opérateur d'induction (In) recherche les propriétés prévisibles des objets inconnus en se fondant sur les ressemblances avec des objets déjà rencontrés et mémorisés.

L'opérateur d'abduction (Ab) détecte les propriétés manquantes ou anormales dans des objets habituellement bien connus. L'opérateur de subsumption (Sb) connecte un type d'objet plus spécialisé à un type d'objet plus général déjà connu. L'opérateur d'analogie détecte les similarités de structures, de caractéristiques et de comportement avec des objets déjà connus et mémorisés.

#### 4. La hiérarchie linguistique

Les objets concrets n'ont pas besoin d'être mis en mots pour être reconnus, le langage apparaît tardivement dans la maturation du système nerveux. Jean Piaget souligne que le langage est graduellement mis en place du 12<sup>ème</sup> au 18<sup>ème</sup> mois de la vie de l'enfant (Piaget, 1992). Le langage des signes est plus ancien que le langage parlé, il est mis en place plus tôt dans l'encéphale et précède le langage dans le flux verbal (Jacquet-Andrieu and Colloc, 2014). La perception, l'émergence, la reconnaissance des objets concrets et la pensée sont antérieurs à l'implémentation du langage, ils sont considérés comme sub-symboliques, ils résultent de la mémorisation des expériences de perception. La théorie de Haeckel « l'ontogénèse récapitule la philogénèse » trouve un nouvel éclairage par la découverte des gènes de développement Hox (Prochiantz, 2001). En particulier, le développement des voies sensorielles a lieu durant l'embryogenèse et continue après la naissance en présence des stimulations de l'environnement. L'absence de stimulation visuelle de chatons à la naissance empêche la maturation de la vision et cause une partielle ou totale agénésie des voies visuelles (Shatz, 1992). De la même manière, les autres fonctions intellectuelles seraient organisées selon une proto-carte intellectuelle (similaire à celle qui a été proposée par Brodman) qui détermine la migration des neurones et la maturation des voies sensorielles et motrices.

La pyramide linguistique (partie droite de la figure 3) établit les symboles, les caractéristiques, les noms des objets abstraits à partir des objets concrets ou d'autres objets abstraits préexistants. Elle détermine la sémantique, la syntaxe des phrases qui permettent d'établir les flux d'entrée (écouter/lire) et sortie (parler/écrire) des textes et des flux d'information (verbaux et non-verbaux). Les opérateurs linguistiques sont : l'opérateur lexical (L) en charge de nommer et d'associer un mot avec un objet concret ou abstrait et avec ses caractéristiques ou comportements. L'opérateur d'appartenance ( $\in$ ) (a-un) établit qu'une caractéristique ou un comportement appartient à un objet concret ou abstrait. L'opérateur sémantique définit la signification du mot (nom, verbe, adverbe...) qui correspond à l'objet concret ou abstrait qui est signifié selon la théorie de (de Saussure, 1916). L'opérateur syntaxique (Sy) détermine dans un langage donné les syntaxes possibles d'une phrase. Les objets du monde ont une signification par leur nom, leurs actions par des verbes et leurs caractéristiques par des adjectifs, l'ensemble ordonné par des règles de syntaxe d'une grammaire et des opérateurs comme la généralisation (est-un), causalité (causé-par), composition (est-partie-de), ordre temporel (précède, succède)... Le langage permet de créer de nouveaux objets abstraits mémorisés à leur tour et ensuite combinés avec les autres de manière incrémentale. Notre modèle est certes toujours incomplet et réducteur mais il montre les interactions entre la perception, la cognition, le langage et les émotions. Il essaie de réconcilier l'approche structuraliste de la linguistique initialisée par Ferdinand de Saussure avec les approches neurophysiologiques et neuropsychologiques actuelles (de Saussure, 1916).



### Légende

**Niv 5.** Les Objets encapsulent (€) des attributs, et les procédures nécessaires pour effectuer les calculs arithmétiques et logiques ainsi que les déductions des valeurs des données.

**Niv 6.** Les systèmes sémantiques implémentent des entités et des objets.

**Niv 7-8.** Les bases de connaissances distribuées sont fondées sur des agents cognitifs et la coopération d'ontologies.

**Niv 9.** Les systèmes multi-agents autonomes dotés de conscience (SMAAC) disposent d'une mémoire interne pour stocker des faits de conscience. Ils sont capables de choisir seuls leurs objectifs et de planifier leurs buts et leurs actions.

Figure 4. La hiérarchie des modèles de systèmes d'information numériques (Colloc, 2016)

## 5. la hiérarchie numérique de l'information

L'information repose sur un triptyque : données, information et connaissance (Abiteboul, 2012). Beaucoup de systèmes à bases de connaissances ont été développés comme MYCIN (Buchanan and Shortliffe, 1984) et SIAMED (Colloc, 1985). La logique était utilisée pour classer les objets comme cela a initialement été montré par Lewis Carroll (Carroll, 1896) et avec des langages comme LISP (McCarthy, 1960) ou PROLOG (Colmerauer, 1996). Les modèles sémantiques (orientés objets puis agents) fournissent des moyens pour modéliser de la connaissance et des ontologies de manière plus efficace (Shen et al., 2015). Le Web sémantique permet d'accéder à une quantité considérable de données et de connaissances concernant le monde et les personnes (niveaux 6-7). Les systèmes multi-agents (SMA) sont utilisés pour réaliser des simulations : une communauté de nombreux agents fournis peuvent coopérer pour résoudre des problèmes nécessitant la recherche de chemins dans des graphes complexes à l'aide de leur comportement stigmergique (niveau 7). Les systèmes multi-agents (SMA) sont capables de coordonner de multiples ontologies et agents connaissances pour fournir une aide à la décision en médecine (Shen et al., 2015). Alain Cardon a proposé un système psychique constructible qui s'appuie sur des connaissances en psychologie et qui est capable de générer de nombreux flux de pensées qui se développent comme une représentation dynamique, se déroulant dans la temporalité sous la forme d'ensembles organisés de processus qui élaborent un système artificiel psychique et ses interactions (Cardon, 2011). Ce système multi-agent autonome conscient (SMAAC) est capable de choisir ses objectifs et de déterminer les buts nécessaires pour les atteindre (Niveau 9).

## 6. Conclusion : les ordinateurs peuvent-ils penser ?

Les ordinateurs pourraient penser d'une manière très différente de celle des humains puisque les avions volent différemment mais bien plus vite que les oiseaux comme le souligne Jacques Pitrat (Pitrat, 1990). Le zen constitue un holisme total où le monde ne peut absolument pas être divisé en parties. Le dilemme est que pour chaque objet du monde, selon le maître zen Mummon : « *On ne peut l'exprimer avec des mots et on ne peut l'exprimer sans les mots* ». On ne peut que vivre ce paradoxe que par son expérience de l'objet, si on le nomme on est obligatoirement réducteur et si on ne le nomme pas c'est lui accorder aucune propriété aussi infime soit-elle, ce qui est inutile. Pour la pensée bouddhiste, se fier aux mots pour parvenir à la vérité, c'est comme se fier à un système formel toujours incomplet (Hofstadter, 1985). Pour Jiddu Krishnamurti, notre conscience est commune à l'ensemble de l'humanité : tous les êtres humains pensent et contribuent à l'élaborer. Il considère l'individualisme, l'égo comme un obstacle à la compréhension de la conscience avec de rares moments de clarté (« insight »). « La pensée est un mouvement dans le temps et l'espace. La pensée est mémoire, souvenir des choses passées. La pensée est l'activité du savoir, savoir qui a été rassemblé à travers des millions d'années et emmagasiné sous forme de mémoire dans le cerveau. » (Krishnamurti, 2005). Il décrit deux formes de pensée : - la première est une réaction de la mémoire qui contient la connaissance, le résultat de l'expérience acquise depuis le début de l'humanité (phylogénèse) et depuis notre naissance (épigénèse) dans une boucle :  $\{ \text{expérience} \rightarrow \text{connaissance} \rightarrow \text{mémoire} \rightarrow \text{pensée} \rightarrow \text{action} \}$  et ainsi de suite... nécessairement limitée dans le temps, elle est utilisée chaque jour, rationnelle, individuelle, avide de pouvoir et de progrès, soumise à la connaissance qui s'accumule, dans les mots qui divisent et cette division est responsable de toute la souffrance, de tous les problèmes du monde. L'autre forme de pensée : l'action-perception où en de rares occasions, nous portons simplement notre attention au monde, sans l'interpréter, sans nommer quoi que ce soit, vierge de tout préjugé et de toute connaissance et en particulier spontanément, en vivant ce moment sans y penser et sans

recourir à la volonté. Par exemple : la perception directe d'un magnifique paysage de montagne un matin avec l'utilisation de tout nos sens où nous nous oublions nous même et bannissons l'usage des mots (Krishnamurty, 2005). Une telle perception ne nécessite que l'utilisation de la partie inférieure gauche : la pyramide cognitive sub-symbolique (figure 3) et pas la pyramide linguistique. Le premier mode de pensée constitue un cercle vicieux qui est déjà implémenté dans les ordinateurs qui pourront penser mieux que les êtres humains ! Et rien ne peut empêcher qu'un ordinateur n'invente une nouvelle religion à l'origine de nouvelles souffrances pour l'humanité (Krishnamurty, 2005). L'apparition du langage durant son évolution aurait donc conduit l'être humain à perdre sa spontanéité de la perception immédiate du monde tel qu'il est. Les technologies numériques, qui renforcent la nature symbolique de notre relation au monde, sont, sans aucun doute, en train de constituer le sommet de notre ignorance. La réalisation de SMAAC dotés de pensée est maintenant possible. Ces systèmes ont accès à Internet et vont devenir rapidement beaucoup plus efficaces et puissants que les êtres humains avec des conséquences inquiétantes pour le futur de l'humanité. Toutefois, nous avons montré que les ordinateurs ne peuvent penser que selon la première modalité seulement. La principale caractéristique de l'humanité est d'être encore capable de penser également selon l'autre forme de pensée mais pour combien de temps encore avant de perdre cette faculté essentielle ?

## **Bibliographie**

- Abiteboul S. (2012). Sciences des données : de la logique du premier ordre à la toile. <http://books.openedition.org/cdf/529>, 2012.
- Arbib A. and Hanson A. R. (1988). Vision, brain and cooperative computation. The MIT Press, 1988.
- Biederman I. (1987). Recognition by components: a theory of human image understanding. *Psychological Review*, 94(2), 115-147.
- Boucart M. (1996). La reconnaissance des objets, Presses Universitaire de Grenoble.
- Buchanan et Shortliffe, (1984) B. G. Buchanan and E. H. Shortliffe. Rule-Based Expert Systems, The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley.
- Cardon A. (2011) Un modèle constructible de Système Psychique. Automates Intelligents.
- Carroll L. (1896) Symbolic Logic, Part I Elementary., London Macmillan and Co.
- Colmerauer A. (1996) The birth of Prolog
- Colloc J. (1985) Informatique Médicale (SIAMED) Application : Logiciel original d'antibiothérapie médicale. Med. Phd, Fac. Medicine, University Lyon 1, 1985.
- Colloc J. (2016) L'éthique des systèmes d'information autonomes vers une pensée artificielle in Enjeux du big data et identifications des données médicales, Les Cahiers du Numérique, ed. Lavoisier, 12 :1-2016, pp187-201.
- Damasio A.R. (1995) L'erreur de Descartes la raison des émotions. Odile Jacob
- Damasio A.R. (2003) Spinoza avait raison Joie et tristesse, le cerveau des émotions. Odile Jacob.
- De Saussure F. (1916) De Saussure F. Cours de Linguistique Générale. Grande Bibliothèque Payot.
- Gazzanica M. et al., (2001) Neurosciences cognitive - la biologie de l'esprit. DeBoeck Université.
- Hawkins J. et Blakeslee S. (2005) Intelligence, Campus Press
- Hofstadter D. (1985) Gödel Escher Bach, Les Brins d'une Guirlande Eternelle. InterEditions, 1985.
- Jacquet-Andrieu A. et Colloc J. (2014) On How to Define Anticipation in the Verbal Flow. *International Journal of Computing Anticipatory Systems*, 28:219-235.
- Krishnamurti J. (2005) La nature de la pensée. 2005.
- Le Moigne J-L. (1990) La modélisation des systèmes complexes. DUNOD-BORDAS.
- McCarthy J. (1960) Recursive functions of symbolic expressions and their computation by machine, part I. *Communication of the ACM*, 3(4):184-195.
- Minsky M. (1988) La société de l'esprit éd. InterEditions. InterEditions, 1988.
- Mountcastle V. (1978) The Mindful brain: Cortical Organization and the group-selective theory of higher brain, chapter An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System, pages 7-50.
- Piaget J. (1992) Biologie et Connaissance - Essai sur les relations entre les régulations organiques et les processus cognitifs Delachaux et Niestlé.
- Pitrat J. (1990) Métaconnaissance Futur de l'intelligence artificielle. HERMES.
- Prochiantz A. (2001) Machine-Esprit. Odile Jacob.
- Purves D. et al. (1997) Neuroscience. Sinauer Associates, Inc., 1997.
- Shatz C. (1992) La maturation du cerveau. *Pour la science*, 181:88-97, 1992.

Shen, Y., Colloc J., Jacquet-Andrieu A., Lei K. (2015). Emerging Medical Informatics with Case-Based Reasoning for Aiding Clinical Decision in Multi-Agent System, *Journal of Biomedical Informatics*, 2015, 56, 307-317

Varela F. et al. (1993) *L'inscription corporelle de l'esprit*, Seuil.

Vincent J-D. (1996) *La Chair et le diable* Odile Jacob.

Vincent J-D. (2002) *Biologie des passions* Odile Jacob.