UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ A
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR

MARCO ANTONIO LUCAS DE SOUZA

INTELLIGENCE ARTIFICIELLE ET PHILOSOPHIE: LES CRITIQUES DE H. L DREYFUS ET J. SEARLE À L'INTELLIGENCE ARTIFICIELLE

MARS, 1992

Resumé

Ce mémoire a pour but de montrer les rapports entre l'Intelligence Artificielle et la philosophie, à partir d'une investigation sur les bases philosophiques de l'Intelligence Artificielle et des travaux de quelques philosophes qui critiquent ce domaine de recherche. Il vise à comprendre comment l'Intelligence Artificielle en tant que domaine lié à la science et à la technologie peut constituer un sujet important pour la réflexion philosophique. Il s'agit d'un effort pour expliciter les rapports entre l'Intelligence Artificielle et la philosophie, en montrant la signification des travaux de J. R. Searle et de H. L. Dreyfus sur l'Intelligence Artificielle, aussi bien que celle de certains concepts et thèses importantes dans le domaine de l'Intelligence Artificielle.

L'idée conductrice du travail est que l'Intelligence Artificielle est fondée sur un *Logos*, lequel est constitué de théories scientifiques et philosophiques et de connaissances techniques. L'auteur en fait donc ressortir, en premier lieu, la signification historique et montre que même si elle est en rapport avec un *mythe*, L'Intelligence Artificielle se développe plutôt à partir d'un ensemble de connaissances diverses et de techniques (*logos*). Ensuite, il expose les critiques philosophiques de J. R. Searle et de H. L. Dreyfus sur l'Intelligence Artificielle, afin de mettre en relief, l'intérêt et le caractère philosophique de ce thème. Son propos est de montrer que lorsque ces deux auteurs font des critiques à l'Intelligence Artificielle, ils ont une cible précise: les présuppositions philosophiques sous-jacentes à ce domaine de recherche, à savoir la tradition philosophique représentationnaliste (empiriste et rationaliste) et le fonctionnalisme.

L'auteur conclut que 1) le fait que l'Intelligence Artificielle se fonde sur un logos n'exclut pas le fait qu'elle possède un caractère mythique lié à l'univers de l'imaginaire humain. 2) Le *mythe* de l'Intelligence Artificielle est lié à son *logos* mais pour définir et réaliser un projet dans ce domaine on a dû abandonner le mythe. 3) Les critiques de Dreyfus et Searle mettent en évidence les rapports entre l'Intelligence Artificielle et la philosophie, le caractère respectivement, anti-représentationnaliste et anti-fonctionnaliste de leurs critiques à l'Intelligence Artificielle montrent que ce domaine de recherche demande une prise de position philosophique par rapport à l'esprit et en particulier aux représentations. 4) Les critiques de Dreyfus et de Searle ont plusieurs points communs et lorsque ces deux auteurs s'efforcent de mettre en évidence les mythes et les limites de l'Intelligence Artificielle, ils se rapportent aux limites philosophiques qui sont derrières l'Intelligence Artificielle i.e. , les limites du représentationnalisme et du fonctionnalisme (*logos*) à la base de ces recherches.

REMERCIEMENTS

Je suis d'abord très reconnaissant à ma collègue Madalena Vange pour son appui ainsi que les critiques qu'elle ma apportées tout au long de ce travail.

Je tiens à remercier tout particulièrement mes professeurs brésiliens, madame Vera L. Vidal, monsieur Hilton Japiassu, Michel Thiollent, Mário L. Guerreiro, et madame Helena Garcia. Je leur dois ma gratitude pour la confiance et l'enthousiasme qu'ils ont montrés en soutenant mon initiative d'entreprendre des études plus avancés au Québec.

Je remercie mon directeur de recherche D. Vanderveken qui a accueilli mon projet avec grand intérêt. Sa rigueur de même que son sérieux académique m'ont toujours inspiré sympathie et respect.

J'exprime aussi ma reconnaissance aux professeurs Nicolas Kaufman et Maryvonne Longeart pour leur apport bibliographique et pour leur encouragement lorsque je me suis adressé à eux pour éclairer certains points importants de ma recherche.

Je ne peux pas oublier de remercier les professeurs Claude Savary et André Leclerc pour leurs remarques et pour leur contribution à la révision de la forme finale de ce travail, lequel a été rendu possible grâce aux bourses de la "Coordenação de Aperfeiçamento de Pessoal de Nivel Superior" CAPES, Ministère de l'Éducation du Brésil.

TABLE DE MATIÈRES

INTRODUCTION

Nous avons pris connaissance de l'expression "Intelligence Artificielle" au milieu des années 1980 lorsque, engagé dans un groupe de recherche en philosophie analytique au Brésil, nous avons travaillé sur les problèmes épistémologiques et logiques de la traduction automatique.

À l'occasion d'une rencontre de jeunes chercheurs promus par l'UFRJ* nous fûmes invités à présenter un travail[1] sur les processus de traduction automatique. Nous avons commencé notre exposé en mettant en rapport les processus de traduction naturelle et artificielle, en signalant, en même temps les contraintes sémantiques les plus courantes affrontées par ceux qui veulent réaliser une traduction automatique acceptable.

Jusque là tout allait bien, mais lorsque nous avons discuté les rapports entre la traduction automatique et l'Intelligence Artificielle et que nous avons essayé d'expliquer les objectifs principaux de ce domaine de recherche nous avons suscité un débat surprenant. Toute sorte de bons arguments philosophiques ont été soulevés pour montrer le manque de bon sens du projet des êtres ou de programmes capables d'agir de façon intelligente. Après cette discussion nous fûmes convaincus que l'Intelligence Artificielle est un thème philosophique par excellence.

A cette époque, l'Intelligence Artificielle était encore méconnue de certains cercles de philosophes et quelquefois, lorsqu'on touchait à ce sujet, on suscitait de petits sourires.

Ensuite nous travaillâmes avec monsieur Michel Thiollent auprès des ingénieurs dans une équipe consacrée au cognitivisme dans un centre de recherche scientifique (Coppe**-UFRJ). Là-bas, nous entrâmes en contact avec la littérature technique en sciences cognitives et en Intelligence Artificielle et nous constatâmes que le thème de l'Intelligence Artificielle suscitait toute sorte de questions philosophiques sur l'esprit et sur le langage, sur l'éthique, la logique etc. Pour tous ceux qui étaient intéressés à l'Intelligence Artificielle et aux recherches cognitives en général, les lectures interdisciplinaires en philosophie représentaient une base indispensable.

Nous avons compris, avec la lecture d'un article de J. Searle et du livre *What computers Can't Do* de H. Dreyfus, que par le biais d'une réflexion sur l'Intelligence Artificielle, nous

---

* Universidade Fedéral do Rio de Janeiro.

1 M. A. Lucas, "A Tradução computadorizada e a tradução humana" *Boletim de Filosofia*, UFRJ-IFCS, Nº 6, decembre, 74-81, R. J., 1986. presentée par occasion de la "II semana de Iniciação científica da UFRJ-1985".

** Coordenação dos Programas de Pós-Graduação de Engenharia

pouvons redécouvrir et analyser beaucoup de problèmes philosophiques. Cela constitue la motivation majeure qui nous a conduit à étudier l'Intelligence Artificielle.

La notion d"intelligence artificielle"

Il y a une distinction courante dans toutes les discussions sur la recherche en Intelligence Artificielle. Il s'agit de la différence entre l'*intelligence artificielle* et *l'intelligence naturelle*. Cela peut sembler trivial, mais il s'agit d'un point de départ qui permet de saisir immédiatement la difficulté de trouver une notion univoque d"intelligence" et, à plus forte raison, d"Intelligence Artificielle".

Le mot "Intelligence"[2] est employé, par le sens commun, pour designer une propriété essentiellement humaine. Il s'agit d'un prédicat qui distingue l'homme de l'animal. L'intelligence est toujours associée aussi à des êtres dotés d'un esprit, ou pour être plus précis, de certains états mentaux.

Si nous considérons les approches objectives (scientifiques) et subjectives (métaphysiques) utilisées pour caractériser l'intelligence ou l'esprit chez les êtres humains, nous en viennons à comprendre pourquoi il n'y a pas d'accord à propos de la façon d'analyser le fonctionnement du cerveau et de l'esprit. Il y a plusieurs critères et théories à notre disposition nous permettant d'analyser le fonctionnement cérébral, l'intelligence et l'esprit, mais il n'y a pas une théorie générale et acceptée capable de rendre compte des relations entre ces sujets.

L'être intelligent doit avoir certaines capacités qui ne sont pas très bien connues. Plusieurs théories cognitives sont actuellement proposées afin de rendre compte de certains aspects de l'intelligence et révèlent la complexité de ce phénomène. Pour comprendre l'intelligence, il faut expliquer plusieurs propriétés cognitives qui lui sont sous-jacentes telles que:

1) des facultés mentales comme le jugement, le raisonnement, la mémoire;

2) des capacités telles que l'attention, l'abstraction, la créativité, l'adaptation, la compréhension d'idées et de concepts;

3) les capacités d'appréhender l'environnement, d'assimiler des connaissances, d'utiliser le langage naturel, de distinguer le concret de l'abstrait;

4) la capacité d'apprendre et d'utiliser de nouvelles connaissances de manière à apprendre et à résoudre des problèmes.

---

2 Mot qui vient du latin *intelligere*, "comprendre".

Ce cadre très simplifié des propriétés cognitives reliées à l'intelligence est sûrement insuffisant pour rendre compte de sa complexité. Nous pourrions construire et rassembler un ensemble très vaste de nos capacités cognitives sans pouvoir toutefois comprendre complètement l'intelligence. Chaque élément mentionné comme l'attention, l'abstraction et l'apprentissage exige une théorie et chaque théorie doit tenir compte de tous les autres éléments mentionnés. Il y a encore plusieurs processus cognitifs et cérébraux qui sont derrière l'intelligence qui n'ont pas été suffisamment expliqués.

En ce que concerne la façon dont l'esprit fonctionne et comment il est lié cerveau, il y a plusieurs thèses, mais aucune ne constitue une théorie acceptée unanimement.

Nous ne connaissons pas très bien quels sont les processus qui se produisent au niveau neurologique donnant lieu à des événements mentaux et à des attitudes considérées intelligentes chez les êtres humains. Nous ne pouvons pas être complètement sûrs, non plus, que l'intelligence ou les états mentaux sont des caractéristiques ou des propriétés qui appartienent seulement à des êtres possédant une structure neuronale comme celle des humains.

En dépit des difficultés à comprendre l'esprit et les processus intelligents qui en résultent, on veut créer des êtres artificiels intelligents.

Le terme "artificielle"[3], est moins problématique que celui d"Intelligence", il désigne des choses fabriquées, inventées et enfin tout ce qui n'est pas produit par la nature, mais par une technique humaine quelconque. Le terme "artificielle" dans l'expression "Intelligence Artificielle" ne veut pas dire plus que ce qu'entend le sens commun: des études qui visent à produire une intelligence de synthèse au moyen des techniques et des théories.

À l'instar de Turing qui n'a pas voulu répondre à la question "les machines peuvent-elles penser?" (précisément parce qu'il a trouvé embarrassant d'analyser la signification du sens commun des termes "penser" et "machine") nous n'avons pas ici, la préoccupation d'analyser les termes "intelligence" et "artificiel" pour voir si les notions qu'ils designent sont incompatibles ou contradictoires. Nous n'avons pas, non plus l'intention de nier ni de prouver, dans ce travail, que l',Intelligence Artificielle[4] est possible.

Une des façons les plus générales de caractériser l'entreprise de la création des êtres artificiels intelligents (l'IA) est de présenter l'hypothèse suivante: si une machine digitale[5]

---

3 Du latin *artificiallis*, qui veut dire "fait avec art".

4 Dorénavant cette expression sera abrégé IA, tel que le font habituellement tous les auteurs dans les textes sur ce sujet.

5Lorsque nous utilisons les termes, "ordinateur", "machine", ou l'expression "machine digitale" dans ce mémoire nous voulons toujours parler des ordinateurs digitaux c'est-à-dire, des machines numériques capables de traiter des données exprimées sous forme discrète (nombres binaires) ou discontinue. Les ordinateurs digitaux ont une forme de traitement des informations différente des ordinateurs analogiques pour lesquels des variations continues de caractère mécanique représentent ponctuellement des données: comme par exemple dans une règle à calcul. Pour une discussion intéressante sur les systèmes digitaux voir J. Haugeland *Artificial Intelligence: The Very Idea*, Bradford Book, MIT Press, 1985, pp. 52-58.

est capable de manipuler de façon efficace certaines représentations (à partir de symboles) et peut effectuer des tâches complexes exigeant de l'intelligence, alors elle est capable d'être perfectionnée pour d'autres tâches complexes où l'intelligence humaine se fait nécessaire.

Même si nous rencontrons une définition générale caractérisant la recherche sur les machines intelligentes, il n'y a pas, parmi les chercheurs en I.A, une définition largement acceptée de l'expression "Intelligence Artificielle". Il existe plutôt différentes définitions lesquelles sont en rapport avec des projets de recherches dont les ambitions peuvent différer, pouvant même être incompatibles. Nous présentons ici quelques unes de ces définitions, à savoir:

1) l'IA est un domaine de recherche basé sur des travaux sur des réseaux neuronaux, qui vise à concevoir des systèmes capables de donner des outputs semblables à ceux produits par un système cognitif humain.

2) l'IA est une recherche qui rassemble plusieurs domaines scientifiques et techniques afin de faire faire par un ordinateur digital des tâches qui demanderaient de l'intelligence si elles étaient faites par un être humain.

3) l'IA est une "discipline" dans le domaine de l'informatique dont le but principal est de bien représenter en transformant en programmes informatiques l'expertise humaine dans plusieurs domaines.

4) l'IA est une recherche de caractère interdisciplinaire qui rassemble des techniques informatiques et des théories diverses visant à comprendre la cognition sous plusieurs aspects et à concevoir des théories en psychologie et des programmes capables d'agir de façon intelligente.

5) l'IA est une recherche liée à l'informatique qui a pour objectif de simuler sur des ordinateurs digitaux, certains processus cognitifs (en tant que processus de traitement d'informations) en rapport avec l'intelligence.

6) l'IA est un ensemble de techniques informatiques et de théories dans plusieurs domaines intéressés à la pensée, à la logique et à l'intelligence. Elle vise à concevoir des programmes capables de résoudre des problèmes, de traiter le langage naturel et de réaliser d'autres tâches en rapport avec l'intelligence.

7) l'IA est un domaine de recherche qui s'intéresse à l'aspect purement formel de l'intelligence (en tant qu'activité complexe d'opération sur des symboles) indépendamment de la constitution physique spécifique du système considéré comme intelligent[6].

Les premières préoccupations de l'IA, avant qu'elle reçoive ce nom au congrès de Dartmouth en 1956[7], étaient d'étudier des systèmes biologiques "auto-organisateurs minimaux", afin d'expliquer comment des conduites intelligentes pouvaient en émerger (approche ascendante).

Après ces expériences les chercheurs ont élaboré des modèles fonctionnels du comportement humain afin de simuler l'intelligence humaine. Ensuite la simulation de l'intelligence a donné lieu à une approche symbolique visant la représentation des connaissances dont le but était de fournir aux systèmes de l'IA des connaissances sur un domaine spécifique leur permettant d'agir de façon intelligente. (approche descendante). L'IA a modifié à plusieurs reprises l'orientation de ces travaux; pour cette raison, elle n'a pas une définition univoque.

Les chercheurs en IA n'ont pas toujours la même formation universitaire, quelques-uns sont informaticiens, d'autres philosophes, d'autres linguistes ou psychologues etc. L'IA n'est pas une science, mais un ensemble non unifié de règles, méthodes et théories provenante de plusieurs domaines tels que l'informatique, la neurologie, psychologie, la linguistique, la logique et la philosophie. Plusieurs auteurs définissent l'IA comme une spécialité qui peut concerner plusieurs des domaines mentionnés. C'est le cas de D. Andler qui expose ainsi son point de vue: "(...) Il faut d'emblée préciser que l'IA n'est pas— ou pas encore— une science, ni même une discipline. C'est, si l'on veut, une spécialité (...)"[8].

L'IA est constituée d'un ensemble de connaissances et de techniques, lesquelles ont leur unité en fonction d'un objectif commun en rapport avec l'activité de programmation, d'explication de certains processus cognitifs relatifs à l'intelligence humaine, ou à la résolution de certains problèmes complexes (par des moyens informatiques) dans plusieurs domaines de la connaissance.

---

6 La définition (1) correspondrait plus au moins à la définition de ce que nous appelons *approche ascendante* en Intelligence Artificielle, tandis que la définition (2) serait une simplification de certaines définitions de l'approche descendante de l'Intelligence Artificielle que nous irons présenter dans les prochains chapitres de ce mémoire.

7 P. MacCorduck, *Machines Who Think. A personal Inquiry into the History and Prospects of Artificial Intelligence*, W. H. Fremman & Company, NY, 1979, pp.96-100.

8 D. , Andler, in H.L. , Dreyfus *Intelligence Artificielle: mythes et limites*, Paris, Flammarion, 1984. , [avant-propos édition française], p. X, (traduit de l'anglais *What Computers Can't Do, The Limits of Artificial Intelligence*, Harper & Row, Publishers, NY, 1979, par Rose-marie Vassallo-Villaneau).

L'Intelligence Artificielle et la Philosophie

La conception d'une "intelligence artificielle" est une entreprise audacieuse qui trouve ses fondements dans les sciences informatiques et ses bases théoriques dans certains courants philosophiques. Les discussions sur l'IA suscitent toujours des réflexions importantes sur la cognition et soulèvent des questions philosophiques, par rapport au langage et à l'esprit, et d'autres thèmes effectivement du domaine philosophique et logique:

> The interests of philosophers and workers in AI intersect and overlap in many ways. Some philosophers have tried to use the resources of AI to shed new light on long-standing philosophical problems, and others have been vocal critics of the philosophical claims made by AI researchers. There has also been convergence. In carryng out specific projects, AI researchers have frequently been led into areas traditionally discussed and investigated by philosophers, providing new opportunities for collaborative exploration.
>
> The philosophical impact of AI has been greatest on the philosophy of mind. AI has suggest new answers to long-standing about the nature of mind, led to the reformulation of traditional problems, and given birth to new controversies of its own. The mind-body problem, the mechanism-free will debate, and disputes about the nature of understanding, intentionality, and intelligence have all been transformed in substantial ways by the advent of AI[9].

Nous croyons que les discussions et les critiques des philosophes sur l'IA ont une importance épistémologique considérable aussi bien pour l'IA que pour la philosophie. Indépendamment de la position prise par le philosophe à propos des machines Intelligentes.

L'IA permet de discuter plusieurs problèmes philosophiques traditionnels d'un point de vue actuel et suscite aussi des nouveaux problèmes et discussions en philosophie. Une autre chose importante à mentionner est la portée de l'intervention philosophique dans le cadre du développement de l'IA.

> Un philosophe comme Daniel Dennett, des chercheurs en IA comme R. Schank, J. McCarthy ou D. Marr s'accordent pour penser que la philosophie peut aider l'IA à progresser sur le plan des concepts, donc à terme sur le plan pratique, tandis qu'inversement l'IA oblige la philosophie à préciser ses choix,à dissiper des malentendus millénaires, à formuler et à tester de manière scientifique de nouvelles hypothèses. L'IA serait pour la philosophie un banc d'essai(...)[10]

Pour D. Andler, la philosophie est à la base de la recherche en IA, laquelle "recycle" le savoir philosophique pour en retirer des éléments importants à sa constitution.

> Parmi les textes produits récemment à l'interface de la philosophie et de l'IA, certains manifestent indéniablement une originalité, une finesse nouvelles. On ne

---

9 R., Van Gulick, *Philosophical questions*, in S.C., Shapiro, et D., Eckroth, *Encyclopedia of Aritficial Intelligence*, J. Wiley & Sons Inc., USA, 1987, p. 736

10 D., Andler *op. cit.* p.p. XIII-XIV.

> peut s'empêcher toutefois de se demander si ces textes ne doivent pas leurs qualités, plus qu'à l'IA en tant que discipline (introuvable), à la perspicacité proprement philosophique de leurs auteurs. Ils confirment en tout cas l'intérêt d'«essayer» certaines idées philosophiques, en quelque sorte, «sur» les concepts utilisés par l'IA, et indiquent la possibilité d'une fécondation mutuelle[11].

En parlant métaphoriquement, l'IA est devenue une sorte un catalysateur à l'intérieur de la philosophie, en même temps que les réactions qu'elle provoque produisent des effets sur elle-même. Les travaux de Searle et de Dreyfus que nous allons discuter à l'intérieur de ce mémoire sont un bon exemple de ce que nous venons d'affirmer. Nous avons choisi de discuter les travaux de ces deux philosophes car ils révèlent les rapports entre l'IA et la philosophie:

> Hubert Dreyfus voit dans l'IA une manière d'aboutissement de toute la tradition philosophique occidentale. Elle serait donc la science (élue) de cette philosophie, et cette philosophie se constituerait en philosophie de cette science. Mais le renversement final n'est-il pas dès lors inévitable? L'IA comme science de la philosophie, donc comme *science philosophique*, comme philosophie exacte, ultime avatar de la philosophie? L'idée se profile nettement dans certains textes de l'IA[12].

J. Searle lui aussi lorsqu'il discute les limites de l'IA est amené à prendre en considération des présuppositions de caractère philosophique qui sont la base théorique de cette recherche pour pouvoir exercer sa critique philosophique:

> Tandis que H. L. Dreyfus faisait de l'IA le point culminant et la conclusion dérisoire de la métaphysique traditionnelle, J. R. Searle fonde son appareil critique sur la philosophie classique; il veut démontrer que, compte tenu des moyens dont elle s'est dotée, l'IA ne pourra jamais répondre de ses prétentions, à savoir simuler l'esprit sur une machine.
>
> [Un peu plus bas, l'auteur de ce passage ajoute:]
>
> La distinction IA-faible/IA-forte n'était là que pour circonscrire ce qui, dans l'IA, a pu servir de prétexte et d'argument à une autre attitude philosophique, le fonctionnalisme[13].

Un point commun entre les critiques de ces deux auteurs est qu'elles laissent sous-entendre que l'IA n'est qu'un mythe.[14] Le mot "mythe" apparaît dans les travaux de ces deux auteurs sur l'IA, mais le sens de "mythe" n'y est pas précisé. L'usage du mot "mythe"

---

11 D., Andler *op. cit.* p.XIV.

12 *Idem.*

13 J.G. Ganascia, *L'âme-machine. Les enjeux de l'intelligence artificielle,* Éditions du Seuil, Paris 1990. p. 220.

14 H.L., Dreyfus *Intelligence Artificielle: mythes et limites*, Paris, Flammarion, 1984. voir aussi, Searle, J., "The Myth of the Computer" *New York Review of Books, 3-6,* avril, 29, 1982.

a été pour nous intrigant, plusieurs autres auteurs l'utilisent aussi pour caractériser certains éléments pré-historiques et historiques par rapport à l'IA.

Ainsi nous avons consulté plusieurs travaux critiques sur l'IA afin de préciser cette notion de "mythe"; le mot "mythe" apparaît à plusieurs reprises, sans que son sens soit discuté ou précisé. Nous avons constaté aussi que pour affirmer que l'IA est un mythe, Searle et Dreyfus et d'autres auteurs se penchent plutôt sur les bases théoriques et philosophiques de l'IA que sur son aspect vraiment mythique. Pour cette raison nous avons proposé de faire une distinction entre le *mythe* et le *logos* de l'IA.

Lorsque nous faisons l'opposition entre le caractère mythique et le *logos* de l'IA nous ne voulons aucunement dire que le *mythe* a un caractère pré-logique ou irrationnel.

Dans le sens dans lequel nous l'entendons dans ce mémoire la notion de *mythe* est en rapport avec un ensemble des conceptions liées à l'idée d'intelligence artificielle, lesquelles ne sont pas dérivées (comme c'est le cas du *logos*) de théories et de méthodes scientifiques, ni de conceptions philosophiques.

À partir de la distinction faite entre *mythe* et *logos*, nous espérons avoir montré qu'il n'est pas licite d'affirmer que l'IA est seulement un *mythe*, car elle repose sur des principes de rationalité hérités de la science et de la philosophie en Occident.

Le but et la structure de ce travail

Ayant pour but d'identifier les rapports profonds et plus récents entre l'IA et la philosophie et aussi de comprendre comment elle peut être une réflexion philosophique sur un domaine lié à la science et à la technologie nous nous proposons de comprendre dans ce mémoire la signification des travaux de Searle et de Dreyfus sur l'IA, aussi bien que celle de certains concepts et thèses importantes dans ce domaine.

Nous ne voulons pas discuter l'impact (au contraire de ce que propose Van Gulick)[15] que l'IA a eu sur la philosophie. Ce qui nous intéresse est de présenter quelques éléments de la recherche en IA que nous croyons d'intérêt épistemologique et qui ont été les plus saillants pour nous. En même temps, nous voulons faire ressortir quelques aspects relatifs aux fondements philosophiques de l'IA. Nous allons montrer, à l'aide des textes de H. L. Dreyfus, J. Haugeland et Searle que l'IA a un rapport actuel et lointain avec la philosophie.

L'idée conductrice qui parcour ce mémoire est que l'IA est fondée sur un *logos* qui est constitué de théories scientifiques et philosophiques et de connaissances techniques.

---

15 R. Van Gulick *op. cit.* , p. 736.

En résumé, l'objectif de ce travail est, précisément, de montrer les rapports entre l'IA et la philosophie, à partir d'une investigation sur les bases philosophiques de l'IA et des travaux de quelques philosophes qui critiquent ce domaine de recherche[16].

La stratégie employée pour arriver à notre objectif est 1) de bien marquer la distinction entre le *mythe* et le *logos* de l'IA 2) de bien déterminer les bases philosophiques du *logos* de l'IA 3) d'exposer les critiques philosophiques de Dreyfus et Searle à l'IA, lesquelles, en elles-mêmes, mettent en relief, l'intérêt et le caractère philosophique de l'IA.

La structure de notre travail est la suivante: dans le chapitre I nous allons faire ressortir la signification historique de l'IA, et montrer que même si l'IA est en rapport avec un *mythe*, elle se développe à partir d'un ensemble de connaissances et de techniques (*logos*).

Dans le chapitre II nous avons un objectif général qui est d'exposer les bases philosophiques de ce que nous appelons le *logos* de l'IA en montrant que l'IA hérite de la tradition philosophique représentationnaliste son modèle explicatif et sa compréhension de l'esprit. Dire que l'IA est héritière de la tradition représentationnaliste comme de plusieurs autres domaines scientifiques est trivial. Ce que nous voulons expliciter en montrant les rapports de l'IA avec cette tradition est que la raison d'être de l'IA dépend de l'assomption de certaines conceptions philosophiques sur la représentation et sur l'esprit.

Dans les chapitres III et IV nous voulons réitérer (à partir des critiques de H.L. Dreyfus et J. Searle à l'IA) les rapport entre l'IA et la philosophie en montrant que lorsque ces deux auteurs font des critiques à l'IA ils ont une cible précise: les présuppositions philosophiques sous-jacentes à ce domaine de recherche, à savoir, la tradition philosophique représentationnaliste (empiriste et rationaliste) et le fonctionalisme.

---

16 Nous avons écrit ce mémoire pour répondre à plusieurs questions personnelles mettant en rapport le développement du calcul mécanique et les conceptions rationalistes en philosophie, toutes ces questions peuvent être résumées dans une seule question: "Qu'est-ce que l'intelligence Artificielle et quel est son rapport avec la philosophie?"
Lorsque nous avons commencé à écrire ce mémoire nous avions un tout autre objectif qui était de découvrir un problème spécifique en rapport avec la tradition philosophique et avec l'IA et le discuter conceptuellement à partir des points de vue de Searle et Dreyfus. Cependant après une recherche exhaustive, afin de trouver des textes importantes capables de servir comme point de départ à notre recherche, nous n'avons pas trouvé un seule travail mettant en rapport de façon analytique les critiques de ces deux philosophes ou consacré aux bases philosophiques de l'IA. Nous avons constaté que plonger dans les problématiques internes de l'IA ou travailler sur des critiques philosophiques plus ou moins externes à cette recherche nous faisait perdre de vue notre prémière question. Ainsi nous avons réalisé une reconnaissance du terrain commun entre l'IA et la philosophie (ce travail) qui est en soi même une réponse à notre question initiale.
Nous croyons que notre effort dans ce mémoire de comprendre ce qu'est l'IA et son rapport avec la philosophie nous permet de passer à une critique conceptuelle (notre première objectif) avec beaucoup plus de familiarité. Nous avons défini les questions communes à l'IA et la philosophie et nous croyons que nous sommes en mesure de poursuivre, sur une base sûre, notre travail philosophique sur ce thème.

PREMIÈRE PARTIE

LE MYTHE ET LES BASES PHILOSOPHIQUES DE L'INTELLIGENCE ARTIFICIELLE

Je crois que le cerveau humain a une exigence fondamentale: Celle d'avoir une représentation unifiée et cohérente du monde qui l'entoure, ainsi que des forces qui animent ce monde. Les mythes, comme les théories scientifiques, répondent à cette exigence humaine. (...) La grande différence entre mythe et théorie scientifique, c'est que le mythe se fige. Une fois imaginé, il est considéré comme la seule explication du monde possible. Tout ce qu'on rencontre comme événement est interprété comme un signe qui confirme le mythe. Une théorie scientifique fonctionne de manière différent. Les scientifiques s'efforcent de confronter le produit de leur imagination (la théorie scientifique) avec la «réalité», c'est-à-dire l'épreuve des fait observables. De plus, ils ne se contentent pas de récolter des signes de sa validité, ils s'efforcent d'en produire d'autres, plus précis, en la soumettant à l'expérimentation. Et les résultats de celle-ci peuvent s'accorder ou non à la théorie. Et si l'accord ne se fait pas, il faut jeter la théorie et en prouver une autre.[17]

François Jacob

[17] F. Jacob "L'évolution sans projet" in *Le Darwinisme aujourd'hui*, Le Seuil, 1979, pp.145-147.

CHAPITRE I

Le mythe et le *logos* de l'intelligence artificielle

Présentation:

L'acceptation par les scientifiques de, et leur insistance sur, l'idée même de la recherche sur les machines intelligentes sont intrigantes, vu que jusqu'à maintenant, les machines de conception classique sont incapables d'opérer en dehors d'une programmation très précise. Cependant les chercheurs en IA n'hésiteraient pas à soutenir que leurs recherches sont cohérentes. Nous allons voir qu'il y a des raisons d'ordre historique, technique et scientifique à la base de l'IA. Ces raisons peuvent être considérées comme faisant partie d'une pré-histoire et d'une histoire de l'IA.

Nous allons commencer notre travail en prenant en considération l'idée poursuivie par l'homme depuis longtemps autour de la création d'êtres artificiels, selon des modèles tantôt mythiques, tantôt rationnels et anthropomorphiques. Dans cet esprit nous proposons une opposition entre le mythe et le logos de l'IA[18]. En ce que concerne son caractère historique, cette opposition est proposée dans le but strict de montrer, suivant Dreyfus, que l'idée d'une intelligence artificielle n'est pas une création humaine qui apparaît soudainement, mais quelque chose qui habite notre imaginaire depuis longtemps[19] et qu'il s'agit, surtout, d'une idée ancrée dans les conceptions scientifiques et philosophiques occidentales.

Pourquoi faisons-nous une distinction entre le mythe et le logos de l'IA? La réponse à cette question est que nous n'avons pas trouvé une unanimité autour de l'expression

---

18 Nous entendons par le terme "Logos" tout un ensemble de connaissances scientifiques, philosophiques et des techniques qui constituent les bases de l'IA dans sa phase non mythique. Nous utilisons le mot "Mythe" pour caractériser les constructions non scientifique de l'esprit en rapport avec l'IA. C'est-à-dire, des conceptions (se rapportant à l'imaginaire humain) sur des êtres artificiels capables d'imiter l'homme de plusieurs façons.
Nous utilisons de l'opposition entre "mythe" et "logos", courante dans l'histoire de la philosophie. Nous faisons cette opposition afin d'exposer les différences entre les croyances non justifiées en 'IA et tout l'éffort d'organisation et assemblage des connaissances diverses et des expériences dans ce domaine.
Il faut mentionner encore que nous exploitons le sens général du terme "logos" dans les acceptions de la philosophie grecque, c'est-à-dire, en tant que "raison" "discours" et aussi dans son sens étymologique originaire de "réunir", "rassembler" (cf. H. Japiassu, et D. Marcondes, Dicionário de Filosofia, Jorge Zahar Editora Ltda, Rio de Janeiro, 1990, p.154). Dans la philosophie grecque, le "logos"est en rapport avec la recherche d'un rassemblement d'éléments pour la conception d'un principe organisateur de toutes les choses, et par conséquent avec le développement d'un tout autre modèle de pensée. le terme logos de l'IA tel que nous le proposons, veut souligner dans ce même sens l'organisation, le rassemblement de plusieurs connaissances et techniques que expliquent la raison d'être de l'IA.
19 M. Longeart, "Intelligence artificielle, Mythe ou réalité?", in Carrefour, Société Philosophique de l'Outaouais", Hull, 1989, pp.149-152.

"Intelligence Artificielle": certains auteurs disent qu'une telle notion est liée à un projet utopique, d'autres, croient fermement qu'elle designe un effort théorique et technique sérieux qui est en rapport avec la pensée rationaliste et avec le développement scientifique et technique de notre civilisation.[20] Nous nous demandons à partir de cette constatation quelle est la raison pour laquelle la notion d"Intelligence Artificielle" est quelquefois associée à un mythe et d'autres fois à un *logos*.

Dans ce premier chapitre nous allons montrer l'ancienneté de l'IA en présentant certains éléments *pré-historiques* sur la conception de machines intelligentes. Nous allons montrer aussi les élément *historiques* (sur le plan intellectuel et technique) qui sont sous-jascents à la conception d'une intelligence artificielle. Pour réaliser cet objectif, nous allons présenter brièvement quelques-uns des programmes, projets et langages; nous parlerons brièvement de certaines connaissances importantes pour l'Intelligence Artificielle, sur l'analogie cerveau-machine, sur les approches ascendante (voie connexionniste) et descendante (la voie analytique et formelle d'orientation symbolique). Nous parlerons aussi très brièvement des difficultés technologiques de l'architecture informatique classique.

Encore dans la partie concernant le logos de l'IA nous mentionnerons quelques idées de savants comme Charles Babbage, George Boole, Norbert Wiener, Warren MacCulloch,Turing, parmi d'autres, qui apportent, avec leurs travaux scientifiques, une grande contribution à l'IA, même s'ils ont travaillé avant l'arrivée des premiers ordinateurs. Ils ont tous aidé de différentes façons à la réalisation mécanique du calcul, permettant ainsi le développement des premières thèses sur des machines intelligentes. Ces savants ont lancé de nouvelles hypothèses sur l'intelligence naturelle et artificielle, sur les langues naturelles et formelles et enfin, ils ont poussé la continuation d'un vieux projet représentationnaliste de tout comprendre par le moyen de représentations et même d'expliquer la pensée comme étant un mécanisme de calcul. Tous les éléments que nous allons présenter visent à montrer la pertinence de l'opposition entre *mythe* et le *logos* appliquée à une analyse historique et philosophique de l'IA.

---

20 cf. M. Longeart, ibdem.

1- Le mythe: la préhistoire de l'Intelligence Artificielle

Le mythe de l'I.A remonte à une période lointaine qui évoque déjà le dessein de l'homme de créer des répliques de lui-même et de son intelligence à partir des éléments de la nature. Cette phase mythique peut être saisie dans la littérature, dans la poésie de plusieurs peuples; elle est basée sur des principes quelquefois religieux.

Plusieurs exemples tirés de textes anciens sont souvent mentionnés, pour montrer l'ancienneté des modèles anthropomorphiques de création d'êtres artificiels intelligents. La référence principale est le chant XVIIIe de l'Iliade[21] de Homère qui décrit plusieurs automates avec des caractéristiques humaines:

> Dans le texte en question, Homère fait expressément état d'objets qui se meuvent par eux-mêmes (automaton). (...) dans le récit, deux servantes en or (...) Celles-ci, qui s'empressent de servir leur maître, sont «semblables à des jeunes filles vivantes », leur « faculté de penser est dotée de la raison>>, elles possèdent même « la voix et la force » et «savent produire des œuvres ». La conception de tels automates fort perfectionnés, fabriqués avec le métal le plus précieux, suppose la possibilité d'un transfert de caractères et fonctions humains (vie, raison, voix, force et capacité de servir et de travailler) à des artefacts. En langage contemporain, on dirait bien qu'il s'agit de robots possédant l'intelligence artificielle[22].

Au même titre que les automates construits par Hephaïstos, plusieurs autres passages de textes importants servent d'exemples et sont rapportés dans des documents historiques sur l'Egypte, Rome, ainsi que plusieurs autres civilisations anciennes. Au Moyen-âge, on trouve dans plusieurs récits des passages sur le thème des êtres artificiels doués de capacités humaines[23]. La plupart de ces modèles anthropomorphiques, comme le Golem de Prague, supposent une intervention divine pour leur création.

Certaines initiatives sont plus ou moins indépendantes d'une telle intervention divine. Bombastus von Hohenheim, plus connu sous le nom de Paracelse, qui a été le fondateur d'une médecine hermétique, propose la création d'une sorte d'homuncule "artificiel" à partir d'une formulation alchimique dont les ingrédients étaient un mélange bizarre fait de sperme

---

21 Cf. Homère, Iliade, Bibliothèque de la Pléiade, Gallimard, France, 1957. Traduction, introduction et notes de Robert Flacelière.

22 L. Couloubaritsis, "Du Logos à l'informatique, l'histoire d'une mutation", in L. Couloubaritsis et G. Hottois, Penser l'informatique, informatiser la pensée: Mélanges offerts à André Robinet, Éditions de l'Université de Bruxelles, Belgique, 1987, pp. 14-15.

23 P. MacCorduck, Machines Who Think. A personal Inquiry into the History and Prospects of Artificial Intelligence, W.H. Fremman & Company, NY, 1979 pp.3-9.

humain. Il semble que l'IA a un rapport plus étroit avec l'alchimie que ne le supposait Dreyfus[24].

D'autres références historiques et littéraires anciennes et plus récentes sur des êtres artificiels nous sont fournies par Pamela MacCorduck[25] nous montrant que la recherche actuelle sur les machines intelligentes est liée à une aspiration qui fait partie depuis longtemps de notre culture. Le mythe résulte des croyances non-justifiées. Il est un aspect de l'IA présent à l'intérieur de plusieurs textes sur l'IA. Il faut cependant remarquer que l'IA se développe plutôt à partir de croyances justifiées par des éléments théoriques de caractère scientifique,technologique et philosophique.

1.1- Le passage du mythe au logos

La médecine du XVIe siècle, a fait de nouvelles découvertes sur le fonctionnement des principaux organes du corps humain. La compréhension et la reconnaissance de l'homme comme étant un mécanisme dont les parties (par exemple,. le coeur et les poumons) sont comparées à des pompes et à des soufflets, a eu une répercussion énorme sur la construction des automates, qui sont devenus de plus en plus perfectionnés dans les siècles suivants.[26]

Cette partie de la préhistoire de l'IA, en rapport avec le développement scientifique et technique a aussi un lien avec la philosophie. Au XVIIe siècle, influencé par le perfectionnement des automates, par ces nouvelles idées scientifiques et par la technique, Descartes propose l'idée de "l'animal-machine". Toutefois, selon lui, même si la machine pouvait simuler quelques processus mécaniques humains elle resterait toujours incapable de simuler l'esprit de l'homme et de comprendre le langage naturel.

> [Des machines] (...) s'il y en avait qui eussent la ressemblance de nos corps et imitassent autant nos actions que moralement (...) nous aurions toujours deux moyens très certains pour reconnaître qu'elles ne seraient point pour cela de vrais hommes. Dont le premier est que jamais elles ne pourraient user de paroles, ni d'autres signes en les composant, comme nous faisons pour déclarer aux autres nos pensées. Car on peut bien concevoir qu'une machine soit tellement faite qu'elle profère des paroles, et même qu'elle en profère quelques-unes à propos des actions corporelles qui causeront quelque changement en ses organes: comme, si on la touche en quelque endroit, qu'elle demande ce qu'on lui veut dire; si en un autre, qu'elle crie qu'on lui fait mal, et choses semblables, mais non

24 H. L. Dreyfus, "Alchemy and Aritificial Intelligence", in The RAND Corporation, décember, Cal. ,1965, P- 3244. pp.82-86.
25 P. McCorduck, op. cit.
26 P. McCorduck, ibid.

> pas qu'elle les arrange diversement, pour répondre au sens de tout ce qui se dira en sa présence(....)[27].

Le modèle de l'animal-machine proposé par le philosophe français suggère celui de la machine humaine (homme machine). Descartes croyait les machines (les automates) très éloignées du modèle humain créé par Dieu. Les automates étaient pour lui des copies imparfaites de ce modèle divin, l'homme-machine. Cependant, le philosophe a écrit :

> Ce qui ne semblera nullement étrange à ceux qui, sachant combien de divers automates, ou machines mouvantes l'industrie des hommes peut faire, sans y employer que fort peu de pièces, à comparaison de la grande multitude des os, de muscles, de nerfs, des artères, des veines, et de toutes les autres parties qui sont dans le corps de chaque animal, considérons ce corps comme une machine, qui, ayant été faite de mains de Dieu, est incomparablement mieux ordonnée, et a en soi des mouvements plus admirables, qu'aucune de celles qui peuvent être inventées par les hommes[28].

Chez Descartes, "l'homme-machine" est basé sur une biologie mécaniste régie par Dieu. Pour lui la machine humaine en tant que création divine se situe au-dessus de l'animal et des automates[29].

> Dans son *Discours de la méthode*, il affirme expressément qu'il se contente de supposer que « Dieu formât le corps d'un homme entièrement semblable à l'un des nôtres, tant en la figure extérieure de ses membres qu'en la conformation intérieure de ses organes », sans le composer pour autant d'une autre matière que celle que son analyse précédente (sa physique) avait décrite, et sans mettre en lui au commencement aucune âme raisonnable, ni surtout aucune autre chose pour y servir d'âme végétative ou sensitive. Dieu excite seulement « en son cœur un de ces feux sans lumière» que Descartes avait lui même expliqué selon des lois mécaniques, et qui sont des « esprits » — analogues aristotéliciens de ce *pneuma* dont les Stoïciens thématiseront définitivement.(...) C'est sa référence aux automates qui l'aide en fait à rendre pertinente la parenté entre l'homme-machine qu'il décrit et l'homme réel, dans la mesure où ils s'accordent aux mêmes lois mécaniques[30].

Il faut avoir à l'esprit toutefois que, pour Descartes, les machines pourraient atteindre le degré de complexité de l'animal mais jamais celui de l'homme. En plus, il fait une

---

27 R. Descartes, Le discours de la méthode, (cinquième partie), Œuvres Philosophiques Tome I (1618-1637), Textes établis, présentés et annotés par Ferdinad Alquié, Éditions Garnier Frères, 1963, pp. 628-629.
28 R. Descartes, op. cit. , p. 628.
29 Pour Descartes, l'homme est mis dans une autre catégorie, distincte des automates, grâce à sa capacité d'utiliser une langue et donc sa raison. L'homme, quoique machine (son corps), ne pourrait jamais être complètement reproduit (corps et esprit) par des inventeurs d'automates.
30 L. Clouloubaritsis, op. cit. , p. 26.

distinction entre les substances mentales, *res cogitans*, et les substances physiques, *res extensa* qui sépare nettement les machines des êtres humains. Les machines sont dénuées de la substance mentale. Malgré son dualisme[31], Descartes, ainsi que Kant et Leibniz, ont rendu possible le modèle épistémologique de l'homme-machine qui permet de mieux comprendre l'homme en tant que système mécanique. Nous trouvons des explications mécanistes du corps chez Kant (*Critique du Jugement* 1790) et aussi chez Leibniz. Pour ce dernier, l'être vivant était aussi assimilé au mécanisme, selon une monadologie:

> Ainsi chaque corps organique d'un vivant est une espèce de machine divine, ou d'un automate naturel, qui surpasse infiniment tous les automates artificiels. Parce qu'une machine, faite par l'art de l'homme, n'est pas machine dans chacune de ses parties. Par exemple, la dent d'une roue de laiton a des parties ou fragments, qui ne nous sont plus quelque chose d'artificiel et n'ont plus rien qui marque de la machine par rapport à l'usage où la roue était destinée. Mais les machines de la nature, c'est-à-dire les corps vivants, sont encore machines dans leurs moindres parties jusqu'à l'infini. C'est ce qui fait la différence entre la nature et l'art, c'est-à-dire entre l'art divin et le nôtre[32].

Descartes apporte une contribution épistémologique importante pour la poursuite future du projet sur l'IA. Comme l'explique Lambos Clouloubaritsis:

> Ce qui n'était qu'implicite dans la conception ancienne de la technè, alors qu'il était manifeste pour l'activité humaine (où le corps est «organikon»), devient explicite dans la technique moderne depuis notamment l'émergence de l'automation, celle-ci supposant des fonctions et des activités pour réaliser la finalité. De ce point de vue, on discerne mieux la mutation accomplie, grâce à l'automation, par le transfert des fonctions propres à l'homme vers l'objet technique.
>
> Dans cette histoire, la pensée de Descartes atteste une étape décisive. En introduisant en effet le modèle de l'homme-machine, elle rend plus proche et plus évident le rapport entre l'homme et l'automate[33].

Si l'homme est sur le plan physique une machine, il devient plus facile de croire à la possibilité d'une "intelligence artificielle"; puisqu'à partir du modèle mécanique de l'homme, nous pouvons concevoir assez aisément des machines ayant les mêmes propriétés. Les discussions de Descartes sur les capacités des automates marquent le passage du mythe au logos de l'IA. *C'est la première fois qu'on se demande si les machines peuvent imiter l'homme.*

---

31 Car la distinction dualiste corps-esprit n'est pas compatible avec l'idée de création de machines intelligentes.
32 G.W. Leibnitz, La monadologie, § 64, éd. Boutroux, lib. Delagrave, 1966, p.p. 178-179.
33 L. Clouloubaritsis, op. cit., p. 25.

La distinction faite par Descartes entre l'homme et la machine vise à établir la distinction entre l'homme et l'animal. Mais en même temps, elle constitue déjà une des premières spéculations sur les possibilités de conception d'êtres artificiels doués de caractères humains.

Le XVIIIe siècle sera marqué par les travaux de deux savants qui ont consacré une partie de leur vie à la construction et à la réflexion sur les automates. Le premier, Jaques de Vaucanson, était un ingénieur de l'industrie de la soie; il fut célèbre par ses trois automates: le joueur de flûte traversière (1737), le joueur de tambourin et le canard (1738). Le second, Julien de La Mettrie, était un médecin et philosophe matérialiste dont la pensée est exprimée dans son *Histoire naturelle de l'âme* (1745) mais c'est dans son livre *L'homme-Machine* (1746) qu'il reconnaît définitivement, l'homme comme étant un mécanisme, une sorte de machine très bien élaborée.

> Je ne me trompe point; le corps humain est une horloge, mais immense, & construite avec tant d'artifice & d'habilité, que si la roue qui sert à marquer les secondes vient à s'arrêter; celle des minutes tourne & va toujours son train; (...) l'obstruction de quelques vaisseaux ne suffit pas pour détruire, ou suspendre le fort des mouvements, qui est dans le coeur, comme dans la pièce ouvrière de la Machine (...)[34].

Julien de La Mettrie considère, en effet, l'homme comme n'importe quelle autre machine. Cependant pour lui, comme pour Descartes et Leibniz, la machine humaine est la plus parfaite parmi toutes les machines.

> Le modèle d'homme-machine cartésien qui s'inspire de la pensée scientifique de l'époque de Descartes, réapparaît, chez Julien de La Mettrie, dans une conception non dualiste, mais moniste. Ce dernier se veut plus cartésien que Descartes en reprenant intégralement le mécanisme physiologique de Descartes et en lui donnant une nouvelle dimension psycho-physiologique [35].

La comparaison de l'homme avec les automates anticipe les premières discussions sur les possibilités intellectuelles et linguistiques des machines faites par A. Turing.

L'idée, issue du mécanisme et du rationalisme cartésien, d'étendre aux machines, à partir d'un modèle, des caractéristiques propres aux êtres humains, marque une nouvelle étape dans l'histoire de la technologie en Occident. Cette étape connaîtra, selon Haugeland, Couloubaritsis, M. Longeart et d'autres auteurs, son apogée avec les recherches en IA.

---

34 J.O. La Mettrie, L'homme-machine (1746), Jean-Jacques Pauvert, Hol., Ultrecht, 1966, p. 145.
35 G. Delaloye, présentation de l'ouvrage de J. de La Mettrie, (op. cit., p. 19),

> La fascination pour les automates et le rêve prométhéen de créer un être à notre image trouvent leur expression paradigmatique dans le mythe du robot. Heritier des mécanismes de Vancanson et de la machine à calculer de Pascal, le robot, machine pensante, est souvent interprété comme le triomphe du rationalisme[36].

La préhistoire de l'IA est caractérisée ici,par le fait que depuis l'Antiquité, l'homme conserve une sorte de mythe anthropomorphique, selon lequel il est possible de concevoir des êtres, semblables à lui, à partir de la matière inerte. Ce mythe est ancré dans la culture occidentale, faisant en sorte que l'idée d'une Intelligence artificielle ne paraît pas si étrange à nos yeux. Mais ce n'est pas pour cela que certains scientifiques reconnaissent l'IA comme un domaine de recherche sérieux et attaché à la tradition, c'est plutôt parce que l'IA est un domaine basé sur un logos qui est en rapport avec la tradition intellectuelle et le développement technologique en Occident.

2- Aspects théoriques et empiriques du *logos* de l'IA

Cette seconde phase, qui correspond à "l'histoire" de l'IA, nous la présentons comme une recherche liée à la tradition philosophique et scientifique de l'Occident. C'est plus exactement, en ce qui concerne l'aspect épistémologique, qui nous renvoie à certaines méthodes et théories empruntées aux disciplines scientifiques, que nous pouvons dire que l'IA est dans la continuité de cette tradition.

Quand les chercheurs en IA suivent la tradition scientifique basée sur la notion de représentation et lorsqu'ils prennent une position quelconque sur le rapport entre le corps et l'esprit, ils sont en train de travailler au tour d'un *logos* de l'IA[37]. Ce logos est caractérisé par l'utilisation des méthodes scientifiques et des instruments conceptuels de plusieurs domaines scientifiques subsidiaires à la recherche en IA.

L'IA est lié à des mythes mais elle s'est développée autour d'un ***logos***. Le logos se présente comme une nouvelle dimension à la fois rationnelle et empirique de l'idée toujours présente chez l'homme de créer un simulacre quelconque de son esprit et de son corps. À la suite de Descartes, qui marque le passage du mythe au logos, Turing a posé, d'une autre façon, la question sur la pensée des machines, mais il a pu posée cette question en vertu un

---

36 M. Longeart, op. cit. , p. 149.
37 Cette idée constituera le point de départ du prochain chapitre concernant les rapports entre l'IA et la philosophie.

contexte bien particulier en rapport avec le développement du calcul mécanique dont le point culminant est l'invention des ordinateurs digitaux.

> (...) pour que Turing puisse poser en 1950 la question cruciale «les machines peuvent elles penser?» il fallait disposer de trois résultats majeurs:
> 1) La démonstration par Turing en 1936 de l'universalité d'un certain type de machine qui sera appelé plus tard «machine de Turing». (...)
> 2) La modélisation par Claude Shannon en 1938, des circuits gâce à la logique des propositions.(...)
> 3) La modélisation par Warren McCulloch et Walter Pitts en 1943, des opération d'une cellule nerveuse et de ses connexions avec d'autres cellules (réseaux de neurones) grace à la logique des propositions. (...)[38]

Le *logos* représente, par contre, un abandon des modèles mythiques; il se manifesté par le triomphe de la conception philosophique de la raison conçue comme une sorte de calcul, et par le développement de la programmation et de la technologie informatique. Ces deux moments, l'un sur le plan des idées et l'autre sur le plan de la technique, constituent respectivement, les bases empiriques et rationnelles pour la constitution de l'IA en tant que domaine de recherche dont la respectabilité est reconnue depuis une trentaine d'années.

L'IA est en rapport avec l'histoire du calcul mécanisé, considérée aussi comme préhistoire de l'Informatique. L'histoire du calcul est marquée par un fait extérieur à elle, à savoir, la réussite de Joseph-Marie Jacquard qui a inventé un métier à tisser(1801). Son invention était équipée d'un mécanisme semblable aux cartons perforés de nos premiers ordinateurs. La machine à cartons perforés du métier à tisser de Jacquard, était capable de fonctionner presque sans intervention humaine, en simulant, pour certains tâches textiles, les gestes répétitifs et précis des tisserands humains.

Les premiers pas de Jacquard vers l'automatisme ont attiré l'attention du mathématicien britannique Charles Babbage qui a conçu le premier "ordinateur" à cartes perforées (1834) considéré comme l'ancêtre des ordinateurs contemporains. Sa calculatrice nommée *Analytical Engine*, restera comme un projet, d'ailleurs très bien documenté, car elle n'a jamais été construite à cause des limitations d'usinage du XIXe siècle. La machine de Babbage était composée d'éléments de mémoire et de plusieurs autres sophistications logiques et technologiques (devenues opérationnelles seulement longtemps après, grâce aux techniques électroniques). L'*Analytical engine* a été conçu avec une telle précision que cela a permis à la célèbre collaboratrice de Babbage, la mathématicienne Lady Lovelace,

---

38 M. Longeart, op. cit. , pp. 149-150.

d'écrire les premiers programmes (très semblables à certains programmes actuels) un siècle avant l'apparition des premiers ordinateurs digitaux.

L'IA est le fruit d'une longue recherche sur les machines abstraites et réelles basées quelquefois sur des modèles anthropomorphiques. Les recherches en IA sont liées dans la première phase à des études sur la structure et le fonctionnement du cerveau inspirées de la cybernétique.

La dernière guerre a obligé les chercheurs de plusieurs domaines à travailler ensemble et très vite. Les travaux à cette époque avaient pour but le développement des techniques balistiques basées sur des mécanismes d'autorégulation et la construction de calculatrices de plus en plus puissantes capables de contrôler ces mécanismes et aussi de déchiffrer des messages cryptographiques. Les résultats au niveau des théories mathématiques ont été fort impressionnants.

A la fin de la seconde guerre, on arrive à la fabrication des premiers vrais ordinateurs qui sont liés à une tradition qui commence avec l'abaque, passe par le métier à tisser de Jacquard et le «moteur» analytique de Charles Babbage et continue avec cinq générations successives d'ordinateurs:

1) les machines à base de tubes et de tambours magnétiques,

2) les machine à base de transistors et de mémoires à ferrite,

3) les machines à base de transistors et de circuits intégrés,

4) les machines à base de circuits LSI (Large Scale Integration).

5) les VLSI (Very Large Scale integrated) sur lesquels les recherches technologiques sont en cours[39].

Les développements des techniques électroniques du XXe siècle deviennent importants pour le développement et l'existence de l'IA, mais cela ne serait rien si on n'avait pas eu une base formelle assez solide, rendue possible grâce aux travaux de mathématiciens et logiciens tels que George Boole, Willian Stamley Jevons, Jacques Herbrand, Alan Turing, W. Pitts, Claude Shannon, et des cybernéticiens tels que Norbert Wiener et W. McCulloch.

Ces savants ont joué un rôle très important dans le développement des recherches en IA. Ils ont ouvert des chemins importants qui conduiront au perfectionnement des modèles théoriques sur les ordinateurs, à des méthodes de déduction automatique, à des traitements d'informations et à la modélisation neuronale. Tels éléments on été importants lors de la phase initiale de la recherche en IA.

---

39 Pour une analyse de l'évolution technologique des ordinateurs, voir Metropolis, N. Howlet, J. et G. C. Rota éds. , A *History of Computing in the Twentieth Century,* Academic Press inc. , N.Y. ,1980.

A la fin du XIXeme siècle, les scientifiques commencent à s'intéresser aux mécanismes régulateurs des machines industrielles. La recherche sur ces mécanismes a amené à la découverte des principes de rétroaction, lesquels associés aux processus homéostatiques des êtres vivants debouchera sur les études cybernétiques aux siècles suivants. Cette discipline aura une importance fondamentale pour la naissance de l'IA.[40] Pendant la seconde Guerre mondiale, Norbert Wiener a créé la cybernétique qui est le fruit d'un projet de recherche destiné, au M.I.T., à des fins militaires.

La cybernétique peut, si on veut, être considérée comme la discipline-mère de l'IA. L'IA a deux choses en commun avec la cybernétique: premièrement l'interdisciplinarité, deuxièmement le fait que les chercheurs en cybernétique seront les fondateurs de l'IA. En plus, l'IA a hérité de plusieurs idées de la cybernétique, ces idées seront importantes pour son développement. Parmi ces idées nous trouvons, premièrement, l'approche mathématique du cerveau et du comportement humain, qui conduisit à l'analogie entre le cerveau et l'ordinateur.[41]

2.1- Les deux approches principales en Intelligence Artificielle

L'IA se développe à partir de deux approches rivales: 1) l'approche ascendante (*bottom-up*), liée aux théories et hypothèses qui ont un caractère biologique, généralement associées à la cybernétique et à des théories évolutionnistes sur le fonctionnement neuronal. 2) l'approche descendante (*Top-Down*), dont les théories et hypothèses privilégient l'aspect formel et la construction de modèles logiques capables de rendre compte des raisonnements humains, des éléments formels du langage naturel et de la façon dont nous solutionnons des problèmes et bâtissons des représentations. Cette approche (***engineering approach***) est basée sur des heuristiques, sur la logique propositionnelle et sur le calcul des prédicats de premièr ordre[42].

Les tenants des approches ascendante et descendante ont pris initialement la voie de l'anthropomorphisme cerveau-ordinateur. Par exemple, même ceux qui croyaient plutôt a une approche formelle pour concevoir de machines intelligentes comme A. Turing

---

40 Cf. R. Lignonnière, Préhistoire et histoire des ordinateurs, ed. Robert Laffont, Paris,1987, pp.100-110.

41 L'étude des systèmes d'autorégulation trouvés à l'intérieur du corps humain et des animaux. ont suscité l'intérêt de neurobiologistes, de mathématiciens et de psychologues. Cette vision cybernétique du comportement animal et humain fait naître l'analogie cerveau-machine.

42 Pour une distinction de ces deux approches et leur rapport avec l'IA et la science cognitive, voir D. C. Dennett , Brainstorms, Philosophical Essays on Mind and Psychology, Bradford Books, Publishers, Inc. Montgomery, Vermon, 1978. Actuellement, avec un certain succès des travaux du néo-connexionnisme, on peut remarquer un regain d'intérêt pour l'approche ascendante; nous parlerons de cela plus loin.

("Intelligent Machinery") prenaient la voie de l'analogie cerveau-machine comme orientation de leurs recherches. J.G. Ganascia ajoute cependant que:

> S'il y a toujours simulation de la physiologie de l'individu, cette simulation déplace son objet. De l'activité du corps en déplacement, on en est venu à s'intéresser uniquement au cerveau, organe dont on sait si peu que toute simulation se limite à un modèle extrêmement grossier. De fait, l'analogie avec le cerveau n'a qu'une fonction opératoire. Il ne s'agit pas de reproduire fidèlement les mécanismes physiologiques, mais simplement de tirer parti de cette image pour dégager les fonctionnalités majeures qui doivent être simulées[43].

Pour les chercheurs de l'approche ascendante il n'y avait aucun problème à établir des analogies entre la structure physique du cerveau humain et celle des ordinateurs. D'un autre côté, pour ceux de l'approche descendante, il s'agissait d'établir des analogies entre la structure formelle sous-jacente à la pensée et les structures formelles des programmes d'ordinateurs.

Nous allons donner maintenant un aperçu historique de l'IA en présentant de façon très résumée [44] les principaux travaux des approches descendante et ascendante. Nous allons prendre en considération la chronologie suggérée par Dreyfus dans ***What Computers Can't Do***.

2.1.1- L'approche ascendante

Dans les années 1950, on assiste à un développement des recherches sur les systèmes de câblage, c'est-à-dire sur les connexions neuronales des êtres primitifs dans l'échelle biologique. Les premiers chercheurs de l'approche ascendante étaient intéressés au système de "câblage" ou de connexion du cerveau humain.Ils pensaient qu'il suffirait de comprendre la façon binaire dont le cerveau traite les informations et d'essayer de créer des répliques électroniques des réseaux neuronaux.

A l'origine de ces recherches qui donnent naissance à l'I.A nous retrouvons comme nous venons de le mentionner le cybernéticien et créateur de la cybernétique Norbert

---

43 Cf. , J.G. Ganascia,op. cit. pp. 23.

44 Cette présentation est loin d'être complète en vertu de toutes nos limitations de temps dans ce mémoire. Nous n'avons pas ici l'intention d'expliquer les particularités théoriques et techniques de chaque programme mentionné, mais nous voulons, tout simplement, présenter quelques travaux le plus représentatifs dans le cadre historique des recherches en IA. C'est avec cet objectif que nous avons fait une brève présentation de chaque programme ou projet, où nous esquissons quelques uns de ses éléments théoriques. Il y a pour chaqu'un des programmes ou projets de recherche présentés, dans cette partie de notre travail, un ouvrage de référence cité, dans lequel nous pouvons trouver des informations de caractère théorique et technique qui sont plus complètes et précises.

Wiener qui a étudié les systèmes biologiques et électroniques[45], et Warren MacCulloch. Particulièrement intéressé par la philosophie, la neurophysiologie et l'informatique MacCulloch, propose une théorie sur le fonctionnement du système neuronal, fondée sur la logique, champ où il a été beaucoup aidé par le mathématicien Walter Pitts[46].

Par leurs études, ces deux derniers scientifiques ont contribué à l'apparition de l'analogie entre le fonctionnement du cerveau et celui des ordinateurs digitaux[47]. Ils ont soutenu l'hypothèse selon laquelle le cerveau fonctionnerait selon des principes logiques en mode binaire[48]. En étudiant mathématiquement le fonctionnement des automates et les connexions neuronales, ils ont réalisé des expériences avec des réseaux électroniques de neurones artificiels ayant pour modèle des réseaux neuronaux biologiques.

> Far more important than MacCulloch's ideas was his certainty that mind could be known and described in scientific terms. He was an enormous influence on Minsky, much greater than, say Turing, whose work Minsky knew somewhat, but regarded as somehoww off two one side. It was MacCulloch's certainty that convinced Minsky and many others that machine intelligence was possible — and to get going on the problem[49].

L'approche de MacCulloch et al. est dite ascendante, car les chercheurs de cette orientation croyaient que l'on pourrait simuler sur des machines les réseaux de cellules nerveuses, par exemple de la fourmi, de la grenouille et du chat et remonter ainsi dans cette échelle biologique jusqu'aux réseaux neuronaux humains. C'est-à-dire, à partir de l'étude des cellules nerveuses des êtres vivants plus simples, ils pensaient en arriver à la simulation des réseaux neuronaux plus complexes.

L'approche ascendante peut être caractérisée par l'analogie cerveau-machine, laquelle persiste jusqu'à nos jours (voir le néo-connexionnisme).Cette analogie peut être expliquée de façon très simple: à la manière d'un ordinateur, le cerveau organise les informations qu'il reçoit en forme d'entrée (input), les traite, et y répond en termes de sortie (output) par le moyen d'opérations logiques susceptibles d'être programmées.

D'après l'analogie établie, les cellules vivantes du cerveau sont comparables aux transistors ou aux simples circuits intégrés élémentaires d'un ordinateur. L'activité

---

45 N. Wiener, *Cybernetics*, The M.I.T. Press, Cambridge, Mass., 1961.
46 W.S. McCulloch, "The Brain as a Computing Machine", *Electrical Engineering*, 1949, vol. LXVIII, 492-497.
47 Nous allons voir que l'approche ascendante stricto sensu qui inaugure des analogies de ce genre s'intéressait surtout à l'étude évolutive des systèmes biologiques; selon les premiers chercheurs connexionnistes cette étude pourrait apporter plusieurs contributions à la conception de systèmes informatiques.
48 Voir W.S. McCulloch et W. Pitts "A Logical Calculus of the Ideas Immanent in Nervous Activity", Bulletin of mathematical Biophysics, Vol. 5, 1943. (source: M. Longeart op. cit. p.152).
49 P. McCorduck, op. cit., p. 77.

neuronale se fait par une propagation binaire d'informations sous la forme de signaux électriques. Cette information est traitée différemment selon les états internes du système. Selon MacCulloch, toute l'activité neuronale peut être comprise avec précision par des moyens logiques capables de représenter la présence ou l'absence dans le cerveau de signaux électriques au niveau neuronal.

L'analogie cerveau-machine repose théoriquement de façon indirecte sur une découverte logique faite au XIXie siècle par George Boole (1815-1864). Ce logicien a conçu un système de logique mathématique, plus précisément une algèbre de la logique où les propositions logiques peuvent être codées sous forme de 1 et 0 et manipulées comme des nombres ordinaires. Il croyait que l'étude mathématique dépendait d'un traitement formel des symboles sans avoir besoin de prendre en considération leur signification[50].

Comme nous le savons, l'algèbre classique nous permet d'établir des relations entre différentes opérations mathématiques Nous trouvons ce type de modèle algébrique dans les formulations logiques proposées par George Boole. Ce logicien utilise l'algèbre pour établir des relations entre les fonctions logiques élémentaires, comme nous l'indique Dreyfus.

> Boolean algebra is a binary algebra for representing elementary logical functions. If "a" and "b" represent variables, "." represents "and," "+" represents "or and "1" and "0" represent "true" and "false" respectively, then the rules governing logical manipulation can be written in algebraic form as follows:
> a+a=a a+0=a a+1=1
> a.a=a a.0=0 a.1= a [51].

Claude Shannon, parmi d'autres chercheurs de son époque, a eu l'intuition que cette algèbre booléenne pourrait être employée pour analyser des systèmes d'information complexe comme les systèmes de communication à distance[52]. Le calcul algébrique de Boole permettait d'analyser les états ouvert-fermé d'un circuit électrique sous la forme de la disjonction, de la conjonction et de la négation logiques. Cette analyse booléenne des systèmes de communication fut un facteur très important pour le développement de

---

50 Pour comprendre les motivations logiques de l'auteur mentionné voir G. Boole, An Investigation of the Laws of Thought on which are Founded the Mathematical Theories or Logic and Probabilities, Collected Logical Works, Open Court, Chicago, 1940, V.II . Voir aussi H. L. Dreyfus (1979), op. cit. , p. 315. qui nous fournit cette référence bibliographique et indique que la première édition de cette ouvrage date de 1854.

51 H.L. Dreyfus, (1979), p.70.

52 Cf. R. Ligornniere, Préhistoire et Histoire des Ordinateurs, éd. Robert Laffont, Paris 1987, pp. 219-222. Voir aussi C. E. Shannon and W. Weaver, The Mathematical Theory of Communication. , University of Illinois Press, Urbana, 1962, cité par Dreyfus (1979), op.cit. , p.165.

l'informatique et par conséquent important pour les conceptions des premières théories basées sur les analogies entre le fonctionnement binaire du cerveau et des ordinateurs[53].

L'analogie cerveau-machine est un thème fréquent dans les textes sur l'IA; elle a des antécédents très lointains. Searle nous donne son témoingnage:

> In my childhood, we were always assured that the brain was a telephone switchboard. ('What else could it be?') I was amused to see that Sherrington, the great British neuroscientist, thought that the brain worked like a telegraph system. Freud often compared the brain to hydraulic and electromagnetic systems. Leibniz compared it to a mill, and I am told that some of the ancient Greeks thought the brain functions like a catapult. At present, obviously, the metaphor is the digital computer[54].

L'analogie cerveau-machine est une façon anthropomorphique de considérer les ordinateurs à des époques où les chercheurs étaient impressionnés par la capacité de manipulation symbolique des machines et par les propriétés binaires du cerveau.

L'approche logique du fonctionnement neuronal et les théories cybernétiques proposées par MacCulloch, ont eu un énorme succès dans les années 1940 et 1950: "Dans les mains de John Von Neumann, à Princeton, ces idées allaient devenir un des principaux catalyseurs de l'invention de l'ordinateur. A cette époque, il utilisait des tubes à vide pour représenter les neurones de MacCulloch-Pitts"[55].

L'analogie cerveau-machine, telle qu'elle apparaît dans les écrits de J. von Neumann, servait surtout à attirer l'attention sur les principaux éléments d'un système informatique. Neumann a été un des premiers chercheurs de l'ère de l'informatique à se demander, inspiré par les idées de MacCulloch[56], comment concevoir des ordinateurs à l'image des connexions du système nerveux humain.

A l'origine de cette analogie entre les systèmes neuronal et informatique, nous devons citer aussi Frank Rosenblat qui a suscité l'optimisme autour des idées connexionnistes, à partir de ses théories et de son expérience avec le *Perceptron* [57]. Les travaux de Frank

---

53 L'algèbre binaire de G. Boole est traditionnellement employée pour le calcul dans les projets des circuits électroniques de l'ordinateur . Ces procédures de calcul ne suivent pas la méthode décimale, mais sont faites en base 2, écrit donc avec deux chiffres 0 et 1, qui sont les deux valeurs d'un "bit" (binary digit) correspondantes à la présence ou absence d'une impulsion électrique dans le circuit de la machine.

54 J. Searle, Minds, Brains, and Science, Harvard University Press, Cambridge, Mass. , 1984 , p.44.

55 F. Varela, Connaître, Éditions du Seuil, Paris, 1989. (Traduit de l'anglais: Cognitive Science. A Cartography of Current Ideas, 1988, par Pierre Lavoie). p.p. 30-31.

56 von Neumann n'était pas, cependant, totalement d'accord avec les idées de MacCulloch selon lesquelles le traitement des informations par le cerveau serait complètement semblable à celui des commutateurs on/off des ordinateurs. Cf. H. L. Dreyfus (1979), op.cit. , p.160.

57 Il s'agissait d'un prototype d'ordinateur inspiré du système nerveux humain. Il était capable, avec beaucoup de limitations, de reconnaître certaines formes.

Rosenblat (1958) visaient la conception de répliques électroniques des réseaux de neurones du cerveau humain à partir de l'imitation du fonctionnement des câblages des neurones. Le *Perceptron* était une machine en réseau conçue uniquement pour réaliser des expériences dans ce sens:

> It originally had three levels. The first was a grid of photocells corresponding to the retine of the eye, which reacted to light stimulus. Below this level, associated units collected the impulses transmitted from the photocells, to which they'd been randomly wired, and those in turn signaled to response units. Because we know that animal, including humans, are born knowing some things, Rosenblat and his coworkers modified the Perceptron in such a way that not all of it was randomly wired, which improved its performance in recognizing and "learning," say, a letter [58].

Rosenblat croyait, déjà à cette époque, qu'à court terme l'ordinateur serait capable de résoudre des problèmes non numériques complexes, de prendre conscience de soi et de se reproduire grâce à des chaînes de montage informatisées. Cependant le Perceptron a connu beaucoup de difficultés techniques et théoriques. Il eut d'importantes réactions contraires aux idées proposées par Rosenblat, principalement de la part de Seymour Papert et Marvin Minsky, deux jeunes chercheurs impliqués dans la création et le développement de l'IA dans les années qui suivirent les recherches de Rosenblat.

Les chercheurs de l'approche ascendante se sont finalment rendus compte du fort parallélisme du cerveau humain par rapport à leurs modèles. Le plus complexe des systèmes informatiques est très simple comparé à la complexité du système nerveux humain. Ce dernier a une structure auto-organisable caractérisée entre autres choses par la complexité synaptique.

2.1.2- L'approche descendante et ses trois phases

L'approche descendante a pris la place de l'approche ascendante dans les années 1960 lorsque Minsky, et Seymour Papert[59], ont discrédité complètement l'approche ascendante. L'approche descendante est aujourd'hui l'approche principale en IA. Contrairement aux chercheurs de l'approche ascendante, ceux de l'approche descendante n'essaient pas d'imiter

---

58 P. McCorduck, op.cit p.87.

59 Cf. M. Minsky et S. Papert, Perceptrons, MIT Press, Mass. , 1969. Selon Dreyfus (1979), (op. cit. , p. 328). Cette ouvrage était dédié aux "psychologues et biologistes qui veulent savoir comment le cerveau calcule la pensée (how the brain computes thoughts)."

les réseaux neuronaux. Ils croient qu'il faut adopter une approche symbolique au problème de la programmation en IA.

L'approche descendante est appelée aussi approche analytique et formelle. Elle est caractérisée par la présupposition que les processus décisionnels qui sont en jeux dans les comportements intelligents peuvent être modélisés selon des régles analytiques et formelles en profitant de la capacité des ordinateurs digitaux. L'approche analytique et formelle s'est développée en trois phases présentées ci-dessous par ordre chronologique.

1) Première phase: la simulation cognitive, la recherche sur le traitement sémantique de l'information.
2) Deuxième phase: les travaux sur les micro-mondes.
3) Troisième phase: La recherche autour de la représentation des connaissances.

2.1.2.1- La Première phase de l'approche descendante

La simulation cognitive et la recherche sur le traitement sémantique de l'information.

Quelques-uns des chercheurs de l'approche descendante, comme Turing (dans sa phase descendante, 1950)[60] et d'autres après lui, pensaient, inspirés par le béhaviorisme, qu'il suffirait de faire en sorte que la machine puisse suivre les méthodes que l'homme emploie lorsqu'il fait quelque chose demandant de l'intelligence. Il suffirait d'observer et d'étudier le comportement de l'homme et de le transformer en programmes. Si la machine était capable d'imiter le comportement humain, elle serait capable d'être intelligente. René Moreau nous parle de cette perspective descendante:

> Puisque ces hommes de grand talent voyaient dans la résolution de problèmes, principale utilisation des machines qu'il concevaient, la manifestation supérieure de l'intelligence, la réflexion des chercheurs informaticiens s'orienta tout d'abord vers ce domaine. Comme ils auraient commis un crime de lèse-majesté en mettant en doute l'anthropomorphisme cerveau-ordinateur, il leur semblait tout naturel de faire suivre à la machine les mêmes méthodes que celles utilisées par l'homme pour résoudre les problèmes. Il se disait alors: « Observons comment un mathématicien fait pour résoudre un problème de géométrie à partir de cercles

60 G.J. Ganascia, op.cit., pp. 26-32.

qu'il trace au tableau et qui ne sont même pas ronds et nous saurons comment faire résoudre des problèmes de géométrie par la machine ».[61].

Les chercheurs de l'approche descendante s'intéressèrent, pendant la phase de la simulation cognitive, à la représentation de processus, propriétés et relations en rapport avec l'intelligence. Ils ont conçu des modèles cognitifs permettant de concevoir des programmes capables de résoudre des problèmes dans le domaine de la logique et de traiter des langues naturelles. Ils utilisaient aussi dans plusieurs expériences des règles heuristiques pour la résolution de problèmes, comme l'affirme Dreyfus: "(...Cognitive Simulation(CS) — [is] the use of heuristic programs to simulate human behavior by attempting to reproduce the steps by which human beings actually proceed"[62].

La recherche visant la simulation cognitive comptait sur des modèles représentationnels basés sur la logique. Mais il ne s'agissait pas, nous le répétons, d'imiter l'intelligence en construisant des répliques des connexions neuronales, mais de simuler ou reproduire le comportement du cerveau humain par des moyens formels et informatiques.

Nous présenterons maintenant quelques travaux importants de la première phase de l'approche descendante, comprenant:

La recherche en traduction automatique

La "traduction automatique" est une des premières recherches réputées en simulation cognitive, pendant sa première phase (les années 1950 et 1960.) L'IA reçoit des contributions de la linguistique. Cette recherche a rassemblé des linguistes, des cybernéticiens et d'autres chercheurs intéressés par la traduction automatique[63]. Les chercheurs de cette phase ont travaillé d'abord sur le dictionnaire informatisé (Anthony Oethinger 1954) et ensuite sur des travaux plus généraux comme les programmes d'ordinateur touchant la syntaxe et la sémantique des langues naturelles.

La recherche en traduction automatique est marquée d'abord par des recherches en cryptographie utilisant des méthodes statistiques sur des gros ordinateurs. A ce moment, la croyance était que la traduction mécanique des langues était un cas plus complexe de

---

61 R. Moreau, in J.C. Perez, De nouvelles voies vers l"intelligence artificielle, Masson, Paris, 1988, préface de l'oeuvre, p.5.
62 H.L. Dreyfus (1979), op. cit., p. 85.
63 Pour un aperçu de la phase initiale du projet sur les machines à traduire voir: A.D. Booth, and Locke, éd., Machine Translation of languages, NY, J. Wiley & Son; Londre, Chapman & Hall, 1955 en particulier l'article de Weaver, W., "Translation", pp.15-23.

déchiffrement cryptographique. Pour les premiers chercheurs en traduction automatique, la traduction par ordinateur était un problème d'ordre formel simuler la tâche du traducteur humain était simplement une question de la rendre compatible avec les algorithmes de programmation. Cela explique pourquoi ces recherches étaient attachées à des études très importantes en syntaxe faites à l'époque par Y. Bar Hillel, Z. Harris et Noam Chomsky [64].

Les travaux de Y. Bar Hillel ont été largement utilisés comme méthodes efficaces pour l'analyse du langage naturel et des langages de programmation par les informaticiens et chercheurs en IA. Bar Hillel a joué aussi un rôle important historiquement, car il a reconnu, dans les années 1960, que dans le contexte théorique de l'époque la traduction automatique était condamnée à l'échec.

Quand cette recherche a été presque complètement abandonnée, au milieu des années 1960, toute la recherche en IA. a subi un refroidissement. C'est seulement au début des années 1970, qu'on commence à récupérer et à faire le point sur les principaux problèmes et solutions trouvées dans les principaux domaines d'études sur le langage naturel. Les années 1980 représentent une phase de maturité pour les travaux en langage naturel, car les chercheurs ont beaucoup mieux délimité le champ de leurs expériences, en tirant les leçons de leurs erreurs antérieures.

Le Logic Theorist

Le Logic Theorist est considéré comme un des premiers programmes informatiques conçus en l'IA. Il était une preuve, selon leur auteurs Allen Newell et Herbert Simon, des capacités formelles illimitées des machines. Selon eux, ce programme permettait d'espérer que dans l'avenir, les ordinateurs seraient capables de faire preuve d'intelligence.

Newell et Simon ont choisi la démonstration des théorèmes comme objet de leur travail car ils croyaient qu'en exploitant le champ des raisonnements sur des problèmes logiques et simulant les capacités formelles de l'esprit humain, on pourrait permettre aux ordinateurs de résoudre toute sorte de problèmes qui demandent des capacités intelligentes humaines. D'après Newell et Simon le programme Logic Theorist pouvait démontrer 38 des 52 théorèmes trouvés dans les *Principia Mathematica* de B. Russell et A. N. Whitehead[65].

---

64 Influencés par les travaux de ces linguistes les chercheurs furent persuadés que la pensée humaine pourrait être simulée par des programmes d'ordinateur. L'ordinateur, par sa capacité de traitement des symboles, pourrait traiter les langues naturelles et rendre compte facilement de l'échange de l'information sémantique exigé dans l'activité de traduction des langues.

65 H.L. Dreyfus (1979), op. cit., p.93.

Pour démontrer un théorème, un être humain doit raisonner et faire des choix. Il doit également être capable de déterminer quelles sont les règles et quels sont les postulats adéquats pour arriver à faire des inférences valides. Pour simuler faiblement cette attitude Newel et Simon ont conçu le Logic Theorist selon une méthode algorithmique[66] et y introduisent des règles empiriques, et des stratégies globales.

Cependant ils ont compris, immédiatement, que la simulation des propriétés formelles de l'esprit est trop complexe pour être programmée de façon efficace. La méthode algorithmique conduisait toujours à des bonnes solutions, mais serait toujours vulnérable à l'*explosion combinatoire*[67].

Pour éviter qu'un même nombre de règles et postulats introduites dans le Logic Theorist conduise à l'augmentation croissante des combinaisons possibles, Newell et Simon ont utilisé la méthode heuristique[68] qui a été utilisée pour éviter la complexification des problèmes lors de la résolution des calculs par le Logic Theorist La méthode heuristique avait cependant un inconvénient, car elle ne permettait jamais de savoir si on pourrait ou non résoudre un problème. Ces auteurs ont reconnu, en suite, que la méthode heuristique devrait être abandonnée à cause dees limitations de cette approche à résoudre des problèmes plus complexes.

Le GPS

La compréhension des difficultés du Logic Theorist conduit les chercheurs à la création du GPS, General Problem Solver ou Système Général de résolution de problèmes qui était basé sur des méthodes moyens-fins, de nature heuristique. Ce programme permettait de prouver quelques théorèmes de logique et de résoudre certains problèmes par la méthode de l'escalade[69].

---

66 a méthode algorithmique est basée sur des procédures finies permettant de réaliser des opérations élémentaires pas à pas pour résoudre un calcul ou un problème en informatique ou mathématique. Selon cette méthode la solution à un problème ne peut être donné qu'après l'examen de tous les énoncés et données concernants ce problème.

67 L'explosion combinatoire est la croissance exponentielle du nombre de choix possibles de stratégies pour la résolution d'un problème.

68 Selon la méthode heuristique, pour résoudre un problème, nous devons appliquer des règles empiriques, basées sur des stratégies humaines ou toute autre procédure capable de permettre de restreindre le champ de recherche pour solutionner un problème. Les heuristiques ont un désavantage, elles ne permettent pas de garantir avec certitude que la solution du problème sera atteinte. Par contre, cette approche permet d'obtenir des solutions plus rapidement que la méthode algorithmique.

69 Selon cette méthode, le choix des chemins de résolution remonte dans une hiérarchie permettant de sélectionner le chemin le plus avantageux pour arriver à un but lors de la résolution d'un problème.

La stratégie du GPS, qui visait toujours à éviter l'explosion combinatoire, consistait à décomposer un problème en sous-problèmes, en proposant des sous-stratégies pour résoudre ces sous-problèmes. H.A. Simon discute, dans un article important intitulé "Modeling Human Mental Processes"[70], les limites et les capacités déductives du GPS et il note qu'on avait franchi quelques barrières quant à la compréhension des lois universelles de raisonnement.

Les jeux en tant qu'objet des recherches en Intelligence Artificielle

L'utilisation de l'ordinateur comme manipulateur de symboles a créé l'expectative que l'homme pourrait en très peu de temps reproduire sur ces machines toutes sortes de tâches qui exigent de l'intelligence. Parmi ces tâches, nous trouvons les jeux. Les jeux, en tant qu'activité ludique, ont attiré l'attention de plusieurs chercheurs sérieux comme C. Shannon, A. Newell, H. A. Simon[71] qui croyaient que les jeux pourraient constituer un chemin intéressant à exploiter pour tester les capacités de raisonnement logique de leurs programmes.

Les chercheurs en IA ont choisi les jeux, car ils se prêtent beaucoup plus facilement à l'analyse et à la formalisation sous forme de programmes d'ordinateur que les autres activités intelligentes humaines. Les jeux obéissent à des règles et ont des buts précis. Cependant, des jeux comme les échecs exigent un ensemble énorme de connaissances spécifiques et la capacité d'utiliser efficacement toutes ces connaissances pendant le déroulement d'une partie. Ces caractéristiques, qui relèvent de la capacité humaine de jugement et de l'expertise, sont très intéressantes pour la résolution de problèmes en IA[72].

La reconnaissance de formes par ordinateur

Il est important de mentionner encore la recherche sur la reconnaissance des formes par ordinateur car elle se révèle fondamentale pour le développement de l'IA, principalement

---

70 Voir H.A Simon, in the RAND corporation, feb., 1961 pp. 2122, cité par H. L. Dreyfus (1979), op. cit,. p. 93.et C.E. Shannon, "A chess-Playing Machine" in World of Mathematics J. R. Neumann Simon and Shuster eds., 1956, A. Newel, J. C. Shaw and H.A. Simon, "Chess Playing Programs and the Problem of Complexity", Feigembaum, E. A. and Feldman, J. Computers and Thought, J. ed. NY. McGraw-Hill, 1963. pp.39-70.

71 C.E. Shannon, "A chess-Playing Machine" in World of Mathematics J. R. Neumann Simon and Shuster eds., 1956, A. Newel, J. C. Shaw and H.A. Simon, "Chess Playing Programs and the Problem of Complexity", Feigembaum, E. A. and Feldman, J. Computers and Thought, J. ed. NY. McGraw-Hill, 1963. pp.39-70.

72 Cf. D. Michie, Reflexions sur l'intelligence des machine, Masson, Paris, 1990 pp.15-69. Traduit de l'anglais On machine Intelligence, 2e edition, Ellis Horwood Ltd, Londre, 1986.

en ce qui concerne la résolution de problèmes. La reconnaissance des formes permet à la machine d'avoir une vue globale du problème et de la situation dans laquelle le problème se pose. Parmi les auteurs importants dans ce domaine, il faut citer, O. Selfridge, U. Neisser[73], Murray Eden[74], Leonard Uhr et Charles Vossler qui sont des chercheurs reconnus pour leurs travaux de haut niveau technique.

Ces deux derniers ont développé un travail sur la reconnaissance des lettres de l'alphabet qui inaugure toute une méthodologie sur la reconnaissance des formes. En résumé, le point intéressant du travail de Uhr et Vossler est la considération de la forme d'une lettre de l'alphabet par rapport à une forme idéale décomposée en éléments plus simples. Le programme, développé par eux, permet d'effectuer la reconnaissance des lettres à partir de ses parties mises en rapport avec un *pattern*[75]. L'ordinateur peut ainsi reconnaître automatiquement de quelle lettre il s'agit, tout en balayant le caractère qui lui est présenté et en reconnaissant ses caractéristiques de base.

L' ANALOGY

Classifié par Dreyfus comme étant dans le domaine de la recherche sur le traitement sémantique de l'information, l'ANALOGY[76] est un autre programme complexe sur la reconnaissance de formes. Il est capable d'identifier des analogies entre patrons géométriques distincts, tels que ceux présentés pour les tests d'intelligence utilisés par les psychologues. Il s'agit d'une des premières expériences dans ce domaine, où on emploie des heuristiques et des règles de transformations logiques en rapport avec un système de repérage global de formes pour résoudre des problèmes abstraits de façon artificielle[77].

---

73 G. O. Selfridge and U. Neisser, "Pattern Recognition by Machine", dans Computers and Thought , Feigenbaun, E. A. and J. Feldman, éd. NY. McGraw-Hill, 1963, pp. 238-250.
74 M. Eden, "Other Pattern Recognition Problem and Some Generalizations in Recognizing Patterns Studies", in Living and Automatic Systems, Paul A. Kolers and Murray Eden, éd. MIT Press, Cambridge, Mass. 1968. Cité par Dreyfus (1979), op. cit,. p.98.
75 L. Uhr and G. Vossler, " A Pattern Recognition Program that Generates, Evaluates and Adjusts its Own operations"in Feigembaum, E. A. and Feldman, J. , éd. , Computers and Thought, McGraw-Hill, NY, 1963, pp.271-353.
76T.G. Evans, "A program for the Solution of a Class of Geometric Analogy Intelligence Test Question" dans Minsky, M. ed. Semantic Information Processing, Cambridge, Mass. M.I.T. , 1969, p.p.346-347.
77Cf. M. Boden, Artificial Intelligence and Natural Man, Basic Books, Inc. , Publishers, 1987, 2e édition, pp. 319-322. et aussi Dreyfus (1979), H. L. , op. cit. , pp. 137-139.

Le STUDENT

Le programme STUDENT (1965) visait la compréhension de phrases simples en anglais; son auteur Daniel G. Bobrow a élaboré un système composée de questions et de réponses simples en anglais. Dans ce programme, 1) les phrases simples du langage naturel sont fragmentées par la mise en évidence de certains mots représentatifs tels que "fois", "de", "égale", 2) les fragments de phrases en rapport avec "fois", "de", "égale", sont exprimés en termes de relations algébriques représentées par des variables[78] L'ordinateur ainsi programmé peut résoudre les équations selon des règles et répondre à des questions simples formulées en anglaistouchant un domaine extrêmement restreint. Le modèle sémantique proposé pour le programme de Brobrow respecte une seule relation: l'égalité et cinq fonctions arithmétiques[79].

Bobrow dit que son programme peut comprendre des phrases en anglais, mais il attribue un sens très faible au terme "comprendre". "In the following discussion, I shall use phrases such as 'the computer understands English". In all such cases, the "English" is just the restricted subset of English allowable as input for the computer program under discussion"[80].

Le programme STUDENT est souvent mentionné pour montrer les limites formelles de la recherche sur le langage naturel.

Il faut mentionner aussi les travaux sur les mémoires sémantiques de Ross Quillian, ce chercheur a beaucoup influencé la recherche en IA. Il s'inspire de la mémoire humaine pour s'attaquer à la compréhension du langage naturel par ordinateur. Il est intéressant de noter que la mémoire sémantique est basée sur la formulation de règles de façon à assembler des notions sémantiques et à les récupérer par après. L'objectif est de permettre au programme de représenter le sens des phrases et d'attribuer un sens à des mots et à des phrases nouvelles. La conception des mémoires sémantiques compte sur des mémoires de stockage d'informations sémantiques, spécialement conçues, et un système de repérage sémantique. Les mots sont enregistrés dans la mémoire de l'ordinateur sous la forme de réseaux sémantiques.

---

78 H.L. Dreyfus (1979), op. cit., pp. 132-137.
79 Cf. H. L. Dreyfus., *op. cit.*, p.133.
80 D. G. Bobrow, "Natural Language Input for a Computer Problem Solving System" *Semantic Information Processing*, M. Minsky éd., M.I.T. Press, Cambridge, Mass. 1969, p.235. Cité par Dreyfus *op. cit.*,. p. 134.

On pourrait programmer le réseau sémantique d'un programme en termes (a) de noeuds représentant des objets, des concepts et des événements et; (b) de liens ou arcs reliant les noeuds et spécifiant la nature des relations entre les objets et concepts. Dans le réseau sémantique les noeuds plus complexes peuvent être constitués de sous-réseaux et tous les mots sont définis par rapport aux autres:

> It further seems likely that if one could manage to get even a few word meanings adequately encoded and stored in a computer memory, and a workable set of combination rules formalized as a computer program, he could then bootstrap his store of encoded word meanings by having the computer itself "understand" sentences that he had Written to constitute the definitions of *other* single words. That is, whenever a new, as yet uncoded, words could be defined by a sentence using only words whose meanings had already been encoded, then the representation of this sentence's meaning, which the machine could build by using its previous knowledge together with its combination rules, would be the appropriate representation to *add* to its memory as the meaning of the new word[81].

Pour Quillian, la mémoire doit être un réseau sémantique d' interconnexions. Le concept de "mémoire sémantique" est en rapport avec le concept de "réseau sémantique" et sert de base à la représentation des connaissances à partir de méthodes heuristiques.

2.1.2.2- Deuxième phase de l'approche descendante

Les travaux sur les micro-mondes.

La recherche sur les micro-mondes a ouvert de nouvelles perspectives dans le domaine de la robotique et de la recherche sur le traitement informatique du langage naturel. Cette recherche visait le développement de méthodes pour la résolution de certaines tâches caractéristiques des systèmes cognitifs humains et la représentation de connaissances dans des domaines restreints appelés micro-mondes ou micro-domaines. L'hypothèse de cette approche descendante sur le micro-domaines est que la cognition est une forme de traitement symbolique qui peut être programmée. L'idée est que les machines ainsi que les hommes sont capables de manipuler des informations symboliques à partir de représentations internes, d'opérations logiques complexes et de l'utilisation de bases de

---

81 R. Quillian, "Semantic Memory" in *Semantic Information Processing*, M., Minsky, ed. M.I.T. Press, Cambridge, Mass., 1969, p.235. Cité par Dreyfus, *op. cit.*, p. 144.

connaissances structurées[82]. Les activités de l'esprit peuvent ainsi, à un certain degré, être représentées sous forme de programmes d'ordinateur.

La robotique.

A la fin des années 1960 une nouvelle branche de l'IA est née. Il s'agit de la recherche sur les robots[83]. Ces machines qui ont un corps analogue au corps humain, deviennent fascinantes comme sujet de recherche car elles faisaient déjà partie, comme on l'a déjà dit, d'une sorte de mythe de notre culture et partie aussi de l'imaginaire de l'homme occidental [84]. La création de domaines de recherche comme celui de la robotique augment l'espoir d'énormes progrès en IA[85]. Cependant la plupart des robots déjà conçus ne correspond pas à la définition forte qu'on vient de donner au terme "robot". Les chercheurs qui ont travaillé dans le domaine de la robotique[86] ont constaté qu'il y avait encore un long chemin à parcourir. Pour réaliser des machines capables d'une autonomie proche de celle des êtres humains, beaucoup de progrès doivent être faits dans d'autres domaines de l'IA tels que la représentation des connaissances, la vision, le langage naturel etc. La robotique, dont le nom vient de la science-fiction, est un thème stimulant pour les chercheurs en IA. Terry Winograd s'inspire de la robotique lors de son expérience avec le SHRDLU (1970)

Le SHRDLU.

Le SHRDLU[87] est un des programmes les plus complexes jamais créés. Il simule un micromonde composé d'objets, tels que des pyramides, des cubes, une boîte vide, et manipule ces objets avec un bras de robot qui est simulé sur ordinateur. Cette expérience vise à montrer que la compréhension du langage naturel par ordinateur n'est pas une tâche impossible. Elle représente un pas important pour l'IA, car c'est la première fois qu'un

---

82 Une base de connaissances est une partie de la mémoire d'un système composé, em général de règles sur des expertises un domaine donné; elle est dite structurée car elle est organisée de façon à ce que les données soient regroupées selon leurs caractéristiques et leurs relations avec les autres.

83Le mot "robot" a été crée par l'écrivain tchèque Karel Capek en 1923 pour nommer les personnages d'une pièce de théâtre intitulé *Les robots universels de Rossum* . *cf* McCorduck, P., (1979), p.5.

84 Le terme "robot" sert à designer des êtres artificiels conçus pour être capables de montrer certains comportements intelligents, d'exploiter, avec une complète autonomie, l'environnement, d'apprendre et enfin d'avoir un comportement adaptatif.

85 Cf. H. L. Dreyfus, *op. cit.*,. pp. 24-26, 80.

86 Pour une discussion interressante sur ce théme voir Albus, J. S. , *Brains, Behavior, and Robotics* , BYTE Publications, inc. EUA, 1981. Voir aussi, Michie, D. *op. cit.* pp.71-155.

87 T. Winograd, " A procedural Model of Language Understanding" *in Computer Models of Thought and Language* Shank, R and Colby, K. eds.W. H. Freeman Press, San Francisco, 1973. cf. Dreyfus, *op. cit.*, p. 5.

programme d'ordinateur intègre syntaxe et sémantique et un ensemble d'axiomes pour traiter le langage naturel, en produisant des actions dans un environnement ou micro-domaine[88]. Les recherches de Winograd ont été importantes pour les recherches ultérieures sur le langage naturel en IA.

Le ARCHES.

Il ne faut pas oublier le programme ARCHES développé par Patrick Winston (1970); c'est un des premiers programmes sur l'apprentissage par ordinateur. Ce travail est exposé dans un ouvrage de Winston sur la vision par ordinateur[89]. Selon Wiston, son programme l'ARCHES est basé sur des heuristiques, il peut reconnaître une arche (objet architectonique) à partir de concepts architecturaux élémentaires pré-programmés qui constituent les descriptions formelles des éléments nécessaires à la composition des objets appartenant à la classe des arches.

2.1.2.3- Troisième phase de l'approche descendante:la recherche portant sur la représentation des connaissances.

Le concept de *Frames* ou Cadre

La recherche sur la représentation des connaissances est différente de celle sur les micro-mondes, car il ne s'agit pas de travailler avec des domaines restreints de la connaissance, mais sur le développement de méthodes de représentation des connaissances et des faits courrants du monde considérés lors du raisonnement sur un problème donné. La recherche sur la représentation des connaissances courantes a été beaucoup influencée par les recherches de Marvin Minsky (1975), sur sa notion de *frame*. Selon Marvin Minsky un frame ou Shéma est une conception nouvelle de structure de données[90] qui fonctionne comme un réseau de noeuds ayant tous des attributs. Ces attributs sont en rapport avec d'autres réseaux de noeuds. Les noeuds sont des objets de type *frame*[91]. qui font partie de

---

88 Cf. H. L., Dreyfus, *op. cit.*, pp. 12-14.

89 Cf. P. H. Winston, "Apprentissage de déscriptions de classes à partir d'examples" *in* P. H. Winston *Intelligence artificielle*, pp. 382-406.

90 En IA une structure de données designe la façon dont les connaissances son regroupées en fonction de leur valeur et des éléments communs qu'elles mettent en relation.

91 Un frame (ou cadre) peut être un objet ou un concept structuré, par structuré on entend les données (cases ou slotsI=éléments d'un frame) qui sont identifiées par rapport à ses attributs et à son rôle lorqu'il est associé à d'autres données dans un programme. Le slot est un composant du frame qui fournit les descriptions atomiques précises sur un

frames plus larges et complexes. Ce système permettait à l'ordinateur de faire des inférences à partir de données non explicites tout simplement en parcourant le réseau de noeuds chargé en mémoire. L'auteur nous l'explique plus longuement ainsi:

> A frame is a data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party... We can think of a frame as a network of nodes and relations. The "top levels" of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals—"slots" that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet.
>
> Much of the phenomenological power of the theory hinges on the inclusion of expectations and other kinds of presumptions. A frame's terminal are normally already filled with "default" assignments[92].

Dans l'article " A Framework for Representing Knowledge" dont on vient de citer un passage, Minsky révèle que son inspiration pour la conception des frames relève de la philosophie, plus précisément de la phénoménologie de Husserl. Il dit qu'il s'est aussi inspiré de R. Quillian et de son concept de réseau sémantique. Le concept de frame constitue un des outils méthodologiques les plus importants pour la recherche en IA.

Les Systèmes Experts

Le programme DENDRAL conçu par Edward Feigenbaum sous l'inspiration des conceptions de frame et des travaux de A. Newell et Simon sur la représentation des connaissances.[93] permet de résoudre des problèmes de chimie dans le cas des expériences sur la spectographie de masse. Il aide par des moyens inférenciels à analyser les structures possibles d'un composé chimique inconnu. Ce programme a eu beaucoup de succès dans le domaine de l'IA. Il a stimulé la réalisation de nouvelles expériences dans la même direction, telles que le programme de recherche CADUCEUS de l'Université Carnegie Mellon, et le MYCIN de Stanford. Le premier date des années 1970 et consiste en une initiative pour rassembler d'importantes connaissances dans un vaste réseau sémantique capable d'aider d'aider à faire des diagnostics médicaux. Le second est un programme aussi

---

concept ou sur un objet du frame. Pour plus de détails voir P.H. Winston, Intelligence artificielle, Inter Editions, Paris, 1988, pp. 263-269.

92 M. Minsky, "A framework for Representing Knowledge" in The Psychology of Computer Vision, P. H. Winston, ed., McGraw-hill, 1975, NY. 1975, p. 212. Cité par Dreyfus (1979), H. L., op. cit,. p.35.

93 Dans ces travaux les connaissances sont représentées à l'aide de règles ou de déclarations logiques du type modus ponens ("si-alors") qui permettent aux programmes de traiter les informations en fonction de faits qui sont déclarés à l'entrée.

en rapport avec des expertises en médecine, mais son but est d'aider au traitement et le diagnostique des maladies infectieuses du sang.

Nous devons mentionner aussi d'autres systèmes-experts importants tels que EMYCIN, PROSPECTOR HERSAY II, qui vont attirer de plus en plus l'attention sur les systèmes-experts.[94]

Le concept de *Script*

Les recherches de R. Schank en IA sur la représentation des connaissances et l'analyse conceptuelle sont liées à l'intérêt de cet auteur pour les questions sémantiques. Ce chercheur, dont la formation linguistique et mathématique a eu une forte influence sur son travail, croit qu'il est possible de dégager la signification des discours et d'expliquer comment les gens peuvent appréhender les structures conceptuelles du langage naturel.

Pour Schank, la signification d'un discours a un rapport intrinsèque avec l'ensemble de concepts qu'il évoque dans l'esprit du locuteur et de l'interlocuteur, car le rapport entre les deux parties dans une situation conversationnelle, ne se limite pas à un échange des mots, mais concerne les sens des phrases prononcées. Une situation conversationnelle implique une dépendance conceptuelle qui fait en sorte que ce qui est dit, lu ou entendu repose sur des structures conceptuelles fondamentales lesquelles représentent la base sémantique de ce qui est dit, lu ou entendu. Schank croit qu'il est possible de créer des programmes capables de "comprendre" les phrases d'une langue naturelle. Pour faire cela, il suffit de ramener les phrases à des concepts élémentaires pouvant être représentés par ordinateur. Dreyfus explique ainsi son idée: "Schank believes he can start with isolated stereotypical situations described in terms of primitive actions and gradually work up from there to all of human life"[95].

Le programme d'analyse conceptuelle de Schank compte sur une base de données qui contient tous les mots qui peuvent figurer dans certaines phrases dans un contexte donnée. Cette base est conçue pour permettre 1) à la machine de décomposer les phrases selon son aspect conceptuel et d'analyser le tout en décrivant chaque phase de son analyse. Le

---

94 Le système expert est un programme capable de travailler à partir des expertises humaines: il est constitué 1) d'une base de faits (base de données qui emmagasine des ensembles des faits sur un domaine); cette base est une espèce de mémoire de travail à laquelle on peut être accéder à tout instant), 2) d'une base de connaissances (qui rassemble toutes les connaissances et stratégies de résolution de problèmes d'un expert dans un certain domaine de la connaissance) et (3) d'un moteur d'inférence (partie du système qui contient sa logique). Étant donné un problème, le moteur d'inférence exploit la base de connaissance et la base de faits et peut par des inférences construire des raisonnements indispensables à la résolution de ce problème. Pour plus de détails voir P. H. , Winston, (1988) op. cit. pp.-144-198.

95 H. L. Dreyfus (1979), op. cit. , p.40.

programme de Schank procède ensuite à 2) une identification du sens de la phrase de façon explicite. Cette étape concerne la représentation de la phrase. L'étape suivante 3) consiste en une analyse de la phrase, fournissant premièrement une paraphrase de la phrase de départ à partir des données obtenues lors de l'étape de représentation déjà accomplie. Ensuite, le programme procède à des processus logiques d'inférence.

Pendant l'étape d'analyse (3), les phrases sont divisées en éléments sémantiques liés à des concepts élémentaires, concernant, par exemple, l'action correspondante aux phrases. Il s'agit ici de déterminer les notions sémantiques primitives qui designent les concepts d'action:

> To carry out this project, Schank invented an event descriptions language consisting of eleven primitive acts such as: ATRANS—the transfer of an abstract relationship such as possession, ownership, or control; PTRANS—the transfer of physical location of an object; INGEST—the taking of an object by an animal into the inner workings of that animal, etc[96].

ATRANS, PTRANS, INGEST etc. , fonctionnent comme des marqueurs des relations dans une phrase et servent à orienter le programme vers la désambiguisation des phrases. INGEST, par exemple, représente un marqueur qui conduit le programme à chercher dans le texte des indices relatifs à la nourriture ou ce qui peut être avalé, digéré, mordu, mâché, etc.

Le système de compréhension du langage naturel conçu par Schank a fait l'objet de plusièrs de contributions de la part du psychologue Robert Abelson. Avec celui-ci Schank a conçu un noveau concepte de représentation de connaissances appelé *Script*, qui ressemble sous plusieurs aspects à l'idée de Shéma ou Frame développée par Marvin Minsky. Schank définit ainsi son nouveau concept:

> We define a script as a predetermined causal chain of conceptualizations that describe the normal sequence of things in a familiar situation. Thus there is a restaurant script, a birthday-party script, a football game script, a classroom script, and so on. Each script has in it a minimum number of players and objects that assume certain roles within the script. Each primitive action given stands for the most important element in a standard set of actions[97].

---

96 dem.
97 R. Schank, "Using Knowledge to Understand", Theoretical Issues in Natural Language Processing, Cambridge, Mass. , june, 10-13, 1975 p. 131. Citation et Italiques faites par Dreyfus (1979), H. L. , op. cit. , p.41.

Le *Script* est constitué d'une suite d'événements qui correspondent à des conceptualisations qui décrivent des types de situation où les choses sont placées dans un enchaînement caractérisant cette situation, par exemple:

Scène 1: entrée dans un restaurant.

Scène 2: commander un plat.

Scène 3: manger le repas.

Scène 4: quitter le restaurant.

Les Scripts permettent de comprendre une situation réeles en fonction des éléments sémantiques en question et de déduire le sens des phrases dans une situation en fonction du contexte. Par exemple, le programme SAM (Script Applier Mechanism)[98] développé par Feigenbaum et son équipe appliquant le concept de script est capable d'interpréter des histoires ayant un certain degré de complexité

3- Les limitations techniques de l'approche descendante et le retour de l'approche ascendante

Les travaux théoriques et les expériences de l'approche descendante sont liés à des formes classiques de programmation qui ont pour base des méthodes algorithmiques et heuristiques et sont compatibles avec l'architecture de von Neumann.

Selon l'architecture proposée par Neumann, tout ordinateur doit avoir quatre composants élémentaires:

(1) unité de contrôle,

(2) unité arithmétique et logique (UAL)[99],

(3) mémoire,

(4) unité d'entrée-sortie.

Les composants (1) et (2) constituent la partie principale de la machine appelée UCT[100] ou Unité Centrale de Traitement.

---

98 Ne pas confondre avec SAM, "Semi-automated Mathematics", un programme conçu pour la preuve de théorèmes en mathématique. Cf. H. H. L. Dreyfus (1979), op. cit., p.301.

99 L'UCT est un élément de calcul important qui constitue le processeur central des ordinateurs. Comme son nom l'indique il fait toutes les opérations arithmétiques et en plus il est responsable du calcul des opérations logiques dans le système.

100 L'UCT, c'est le composant de base de l'ordinateur responsable de l'interprétation et de l'exécution de toutes les instructions entrées dans le système informatique traditionnel.

A un niveau de programmation plus bas les formes de programmation sont toutes compatibles avec l'architecture de Neumann qui est basée ainsi sur trois principes:

(1) toutes les instructions doivent être codées en séquences de 0 et 1 compatibles avec les circuits de la machine;

(2) les instructions et toutes les autres informations nécessaires à l'opération demandée doivent être stockées en mémoire;

(3) le programme enregistré (2) doit fonctionner de manière à ce que, pendant son déroulement, il puisse chercher directement les instructions dans la mémoire, laquelle peut garder aussi d'autres programmes et instructions en rapport avec celui-ci.

L'architecture de von Neumann impose que les données et les instructions soient chargées en mémoire. Mais il y a une séparation entre les données et les programmes. Les données sont programmées pour occuper des cellules de mémoire et les instructions vont indiquer l'adresse et les changements du contenu de ces cellules de mémoire Tout le traitement de l'information est orchestré par un processeur digital sériel qui compose l'UCT.

L'architecture de von Neumann caractérise très bien la façon dont la plupart de nos ordinateurs actuels fonctionnent. Elle oblige l'existence d'une voie unique entre l'UCT et la mémoire de l'ordinateur. Cette voie, appelée bus, représente une limitation technologique pour la conception des programmes. Le bus limite, en certains aspects, la conception de nouveaux logiciels dans le domaine de l'IA, car les données suivent toujours cette voie unique, ce qui fait que toute l'information est traitée de façon fragmentaire.

Le fait que l'architecture classique permette seulement une opération de calcul à la fois diminue beaucoup la vitesse de calcul des machines actuelles; cette vitesse est limitée en vertu de la lenteur de la vitesse d'accès (par l'intermédiaire du bus) à la mémoire[101]. Cette limite, connue en informatique sous le nom de "goulot de Neumann", fait en sorte que toutes les techniques de programmation en IA se conforme elles aussi aux limites de la conception informatique classique de l'architecture de Neumann.

> Ce « goulot d'étranglement de von Neumann » impose des limites radicales lorsque la tâche à accomplir demande un grand nombre d'opérations séquentielles (comme l'analyse de l'image ou la prédiction météorologique). La recherche constante d'algorithmes de traitement parallèle est restée sans succès car elle va à l'encontre de l'orthodoxie computationnelle[102].

---

101 Nous expliquerons ci-après, par le moyen d'un exemple, les limitations de l'architecture de von Neumann.
102 F. Varela, op. cit,. p.55.

Supposons qu'un ordinateur classique doive identifier, par le moyen d'un processus de vision artificielle, trois objets: un cube dessiné sur du papier, un cube réel et une paire de souliers. De tous ces objets, la paire de souliers est évidemment celui qui a le plus de traits caractéristiques ou traits invariants, lesquels nous permettent de l'identifier parmi les autres. Mais c'est exactement cette propriété quantitative et qualitative qui va, paradoxalement, constituer la difficulté pour la machine séquentielle.

Un système avec des yeux artificiels commandés par un ordinateur de structure classique devra *grosso modo* consulter souvent ces cellules de mémoire et envoyer des informations à l'UCT par l'intermédiaire d'un *bus*[103] de données qui ira traiter les informations visuelles, les traduire en chiffres binaires, les envoyant de nouveau à la mémoire par le *bus*.La "perception" du système en question dépend alors de toute une série de calculs, qui demandent un acheminement séquentiel des données entrées, la mémoire et l'UCT.

Tous ces aller-retours, faits séquentiellement entre l'UCT et la mémoire, prennent un temps énorme et ralentissent le traitement des données; car en vertu de son architecture, le système séquentiel prendra beaucoup plus de temps pour reconnaître la paire de souliers que les autres objets.

Pour les êtres humains, plus un objet concret possède de traits caractéristiques plus il est facile de l'identifier:

> De multiples représentations coexistent dans notre cerveau au sein d'une même classe d'objets:nous distinguons facilement un chat, d'un chien ou d'une chèvre, mais nous percevons aussi immédiatement qu'il s'agit d'un vrai chat, d'une statue de chat ou d'un chat en peluche...Le stylo avec lequel nous écrivons peut se trouver près ou loin de notre oeil, orienté ou éclairé n'importe comment; c'est à autant d'images différentes qu'il donnera lieu sur notre rétine, et cependant, dans tous les cas, nous le reconnaissons, nous l'identifions aisément même s'il se présent sous un angle sous lequel nous ne l'avions jamais vu auparavant. Nous reconnaisons en quelque sorte de choses jamais vues, par combinaison de notions empruntées à des choses déjà vues[104].

Le but des recherches sur d'autres architectures "non classique" est de doter les machines de cette capacité humaine à saisir globalement un problème ou un objet, et

---

103Le bus des données est constitué par un ensemble de câblages de l'ordinateur à l'intérieur desquels circulent sequentiellement des informations sous forme binaire.
104 G. Lazorthes, Le cerveau et l'ordinateur, Editions Privat, Toulouse, 1988, p.115.

d'autres tâches qui nous sont caractéristiques, lesquels à l'heure actuelle représentent une difficulté insurmontable quand elles sont programmées sur des machines séquentielles.

Cela conduisit les chercheurs à essayer de concevoir d'autres architectures informatiques qui n'exigent pas de *bus* qui permettent une programmation plus souple et plus rapide. Les chercheurs veulent concevoir de moyens technologiques pour que les programmes puissent être traités de façon accélérée par plusieurs processeurs.Ils essayent de créer des schémas d'interconnexions entre ces processeurs, c'est-à-dire, de concevoir des topologies ou des configurations informatiques en réseaux de plus en plus optimales et qui soient mieux adaptées aux applications de l'IA.

Les architectures parallèles requièrent ces configurations informatiques en réseaux, afin de mettre en rapport plusieurs processeurs pour transmettre et traiter de façon parallèle les informations. La recherche sur le parallélisme vise à rendre les programmes plus efficaces, tout en évitant les problèmes des "goulots de Neumann".

3.1- La voie connexionniste: le retour de l'approche ascendante.

Le connexionnisme représente un retour de l'approche ascendante dans les années 1970. Il est marqué par un refus de l'orientation symbolique proposée par les recherches de l'approche descendante. Les raisons de l'opposition à l'approche symbolique sont renforcées par le fait que les recherches en neurobiologie ont montré, dès le début des premiers travaux connexionnistes, à l'époque de la cybernétique, que le cerveau ne travaille nullement selon des processus logiques entierement séquentiels et ne fait pas appel à à des mémoires adressées[105].

> Durant les premières années de la cybernétique, déjà, la prédominance de la logique comme approche principale en sciences cognitives était remise en question. (...), par exemple, le fait qu'on ne trouve pas de règles ni de processeur logique dans un cerveau réel et que l'information n'y est pas stockée à des adresses précises fut l'objet de discussions importantes. Il apparaissait plutôt que le cerveau fonctionne à partir d'interconnexions massives, sur schéma distribué, de sorte que la configuration des liens entre ensembles de neurones puisse se modifier au fil de l'expérience[106].

---

105 Les chercheurs néo-connexionnistes travaillent plutôt avec la conception de mémoire répartie, selon laquelle la mémoire doit être répartie par tout dans la structure du réseau neuronal du système connexionniste et non pas localisée selon des adresses comme dans les systèmes sériels traditionnels comme ceux conçus selon l'archicteture de Newman.
106 F. Varela, op. cit. , p. 53.

Le connexionnisme peut être considéré comme une continuation des projets de recherche réalisés en l'IA[107] par la voie ascendante. Il vise à résoudre les questions fondamentales, du point de vue théorique et technique afin d'aboutir à des formes d'intelligence entièrement artificielles à partir des recherches sur le parallélisme de masse, sur l'auto-organisation, sur le traitement coopératif (*Cooperative Action*)[108], sur les mémoires associatives et sur les nouvelles versions de perceptrons[109].

Le connexionnisme, ou mieux, le néo-connexionnisme, car, comme nous venons de dire, il s'agit d'un retour à l'approche ascendante, est influencé par les travaux de Frank Rosenblat et de MacCulloch-Pitts. Les récents résultats de travaux des connexionnistes ont créé un lien étroit entre les recherches en IA et les neurosciences. Cela explique pourquoi ces chercheurs sont appelés des "neuro-informaticiens". Le connexionnisme semble être prometteur selon certains auteurs, parce qu'il permet de développer certains branches de la recherche en IA et à obtenir certains résultats intéressants dans de secteurs tels que la vision, la reconnaissance de la parole, l'apprentissage, etc. ,qui présentaient jusqu'ici des limitations insurmontables pour la recherche descendante.

Les néo-connexionnistes sont en train d'étudier les structures complexes d'organisation des réseaux neuronaux. Le parallélisme des connexions neuronales du cerveau a été choisi, par eux, comme modèle pour des nouvelles architectures informatiques, considérées comme beaucoup plus appropriées au développement de l'IA que les modèles séquentiels basées sur l'architectures classique.

Les chercheurs néo-connexionistes ont constaté, dans leurs études sur le fonctionnement neuronal, que le traitement de l'information par le cerveau est au niveau synaptique,beaucoup plus lent que le traitement séquentiel et linéaire fait par les ordinateurs, car la vitesse de transmission des synapses est plus lente que celle des relais électroniques. La vitesse de traitement séquentiel est mesurée en *nanosecondes*[110] soit un million de fois plus rapide que la vitesse de traitement d'information au niveau neuronal. Mais la lenteur du cerveau au niveau synaptique est compensée par l'efficacité incomparable due à son parallélisme.

---

107 Traditionnellement, on oppose l'IA au connexionnisme par le fait qu'il s'agit de deux approches concurrentes. Cependant, nous pouvons considérer les recherches connexionnistes comme une continuation des projets de l'intelligence artificielle sur des bases non symboliques. Cette continuation est justifiée dans le sens que les chercheurs ascendants utilisent (de la même façon que leurs collègues de l'approche descendante), des techniques informatiques basées sur des représentations .

108 Cf. M. Boden, op. cit. , p. 483.

109 Idem.

110 Autrement dit, 'Unité de mesure de la vitesse de traitement d'information d'un ordinateur digital correspond à 10\ s\up4(-9)de seconde.

> Furthermore, all contemporary computers are serial processors, or nearly so; that is, only a comparative handful of components are actively doing anything at any one time—the rest are just waiting passively. But there is every reason to believe that brains function massively in parallel; millions or even billions of processes go on simultaneously. On the other hand, advanced semiconductors operate roughly a million time faster than neurons; and there is at least an approximate trade-off between speed and degree of parallelism.
>
> Despite these difficulties in comparison, I think it safe to say that the brain is many orders of magnitude more complex than any present artifact[111].

En plus de chercher à comprendre l'énigme du parallelisme du cerveau. Les chercheurs en sciences cognitives d'orientation connexionniste essayent d'expliquer la capacité humaine à travailler avec (et en dépit) de l'imprécision, de manipuler différents concepts de temps tout en prenant en considération le temps réel; l'habilité que les êtres humains ont de manipuler aussi des concepts flous comme petit-grand, mince, éloigné, et notre capacité d'établir des rapport causaux. Tout cela semble dépendre d'une certaine capacité globale à considérer des événements qui sont en rapport avec le parallélisme du cerveau humain. Les connexionnistes conçoivent des modèles qui visent à rendre compte de plusieurs processus cognitifs humains, tels que la mémoire associative, la capacité de bâtir des généralisations à partir d'exemples, la reconnaissance d'images complexes, et enfin le traitement d'informations brutes du monde réel, tels que, bruits, formes, couleurs, etc.

Les travaux sur les automates cellulaires ont favorisé le développement des théories sur le traitement global de l'information par les systèmes connexionnistes. Ils ont ainsi une influence sur le progrès et le succès du néo-connexionnisme. René Moreau, chercheur intéressé par l'IA, nous explique de façon resumée ce qu'est un "automate cellulaire" et comment il permet de traiter globalement des données:

> Un automate cellulaire est composé d'un réseau de cellules tel que chacune d'elles est un ordinateur élémentaire. Toute cellule peut prendre un nombre limité d'états. Pour changer d'état elle observe l'état des cellules qui lui sont adjacentes et applique une loi qui est la même pour toutes les cellules de l'automate et qui tient compte de l'état des cellules voisines, il y a donc bien connexionnisme. Toutes les cellules calculent au même instant l'état dans lequel elles seront placées à l'instant suivant. Ce calcul est donc fortement parallèle: s'il y a n cellules il y a n calculs simultanés. Un automate cellulaire peut donc être considéré comme un véritable réseau d'ordinateurs élémentaires (les cellules) chacun communiquant seulement avec son voisinage, d'où le nom de connexionnisme[112].

---

111 J. Haugeland, Artificial Intelligence: The Very Idea, Bradford Book, MIT Press,1985, p. 170.
112 J.C. Perez, De nouvelles voies vers l"intelligence artificielle, Masson, Paris, 1988, p.7.

Le réseau d'automates est utilisé dans les expériences sur l'apprentissage, la reconnaissance des formes et la compréhension du langage naturel. Il s'agit là de domaines moins développés aujourd'hui en IA. Les connexions d'un réseau d'automates servent de base à l'information qui y est codée sous forme de connaissances: "(...) l'information (connaissance) est 'CODEE DANS LES CONNEXIONS' en concepts définis par leur position dans le réseau"[113] Les chercheurs visent une optimisation du comportement du réseau, de façon à leur permettre d'évoluer vers les tâches associatives d'apprentissage, de reconnaissance et de compréhension.

Les chercheurs connexionnistes n'emploient pas de méthodes symboliques; il utilisent, par contre, les mathématiques pour fonder leurs hypothèses et développer leurs systèmes cognitifs ou réseaux de neurones artificiels, lesquels ont des rapports avec les systèmes neuronaux biologiques étudiés par les neurobiologistes.

> Un des aspects les plus intéressants de cette approche différente en sciences cognitives est que les symboles, au sens conventionnel, en sont exclus. Dans le cadre de l'approche connexionniste la computation symbolique est remplacée par des opérations numériques, par exemple les équations différentielles qui gouvernent un système dynamique. Ces fonctions sont plus fines que les opérations sur les symboles: le résultat d'une seule computation symbolique discrète serait obtenu dans un modèle connexionniste au moyen d'un grand nombre d'opérations numériques qui régissent un réseau d'unités simples. Dans un tel système, les éléments significatifs ne sont pas des symboles mais plutôt des schémas complexes d'activité entre les multiples éléments qui constituent le réseau[114].

En résumé, pour les connexionnistes inspirés du fonctionnalisme, la cognition n'est pas seulement une caractéristique des systèmes biologiques; tout système capable de présenter de l'émergence d'états globaux dans un réseau de composants simples est un système cognitif. Cette approche ne considère pas les éléments significatifs du système cognitif comme des symboles, mais plutôt comme des schémas complexes des processus cognitifs entre éléments constituants d'un réseau neuronal. Les significations dans un système connexionniste dépendent de l'état global du système et non pas des valeurs symboliques spécifiées au préalable. Elles résident dans un niveau sub-symbolique.

> Au niveau sub-symbolique, les descriptions cognitives sont construites à partir de constituants qu'à un niveau supérieur on appelerait symboles discrets. Le sens, toutefois, ne réside pas dans ces constituants en soi, mais dans les schémas d'activité complexes émergeant d'une interaction entre plusieurs d'entre eux.

---

113 J. C. Perez, op. cit. , p.65.
114F. Varela, op. cit,. p.p. 77-78.

> Cette différence entre le sub-symbolique et le symbolique nous ramène à la question de la relation entre les différents niveaux d'explication dans l'étude de la cognition. Comment l'émergence sub-symbolique et la computation symbolique peuvent être reliées ?[115].

Les systèmes connexionnistes conçus sont constitués de multiples éléments simples qui, lorsqu'ils sont reliés, constituent un système en réseau donnant lieu à l'émergence de propriétés cognitives globales semblables, dans certains cas, à celles des systèmes cognitifs naturels.

Selon l'approche de l'émergence un système cognitif artificiel a les caractéristiques suivantes:

1) Posséder une configuration souple des liens entre l'ensemble de neurones. Les connexions peuvent être modifiées en exploitant l'auto-organisation du système.

2) Travailler selon des règles de changement graduel des liens neuronaux à partir d'un état initial plus ou moins arbitraire.

3) Appliquer quelquefois des modèles d'apprentissage du type règle de Hebb (Hebb a suggéré que l'apprentissage était lié à des changements dans l'organisation neuronal.)

4) Exploiter des activités correlées entre les neurones: si deux neurones sont activés au même temps ils renforcent un lien dans le réseau neuronal, autrement les liens sont affaiblis.

5) Avoir une configuration de liens neuronaux qui est inséparable de l'histoire des transformations par lesquels réseau a passé.(apprentissage)

Les systèmes cognitifs artificiels opèrent selon un processus d'auto-organisation, ou mieux, synergétique, qui permet que les neurones artificiels puissent agir de façon *coopérative* en faisant émerger un résultat global.

> (...) grâce à la nature configurationnelle du système, une coopération globale en émerge spontanément lorsque les états de chaque «neurone» en cause atteignent un stade satisfaisant. Un tel système ne requiert donc pas d'unité centrale de traitement pour contrôler son fonctionnement. Ce transfert de règles locales à la cohérence globale est le cœur de ce qu'il était convenu d'appeler l'auto-organisation pendant les années de la cybernétique. Aujourd'hui, on préfère parler de propriétés émergentes ou globales, de réseaux dynamiques, ou non linéaires, de systèmes complexes, ou encore même de synergétique[116]

---

115 F. Varela, op. cit,. p. 80.
116 F. Varela, op. cit. , p. 61.

Il est important de mentionner à propos du connexionisme la contribution des études sur les modèles parallèles de mémoire associative[117] menées par Geoffrey Hinton et James Anderson. Ces deux chercheurs se sont intéressés au parallélisme de masse, c'est-à-dire, la création de machines à haut degré de traitement parallèle. Il est interressant de mentionner à ce sujet la machine de Boltzmann[118] dont la principale caractéristique est de pouvoir apprendre au moyen d'exemples, est un système parallèle basé sur des interconnexions neuronales. Perez nous en donne un aperçu:

> Dans la "BOLTZMANN machine", chaque processeur est connecté à des centaine d'autres et ces processeurs sont à états binaires. Le réseau est de type connexionniste, (...). Actuellement la "BOLTZMANN machine" est réduite à une centaine de processeurs,(...) Les micro-processeurs utilisés sont limités à 64 connexions tandis qu'un neurone humain peut être connecté à 50000 synapses (!) Ce qui est nouveau, c'est que chaque processeur calcule son nouvel état non sous forme DETERMINISTE mais PROBABILISTE, ce qui permet d'introduire une part de HASARD dans le processus[119].

Une autre chose à remarquer avant de finir notre exposé sur le néo-connexionisme est que les réseaux neuronaux connexionnistes sont des systèmes parallèles de traitement d'information dont la mémoire n'est pas localisée comme dans les systèmes sériels. Au contraire, la mémoire se répand partout dans le système. Une caractéristique importante de cette mémoire est qu'elle ne travaille pas selon le modèle binaire du tout ou rien, mais plutôt selon des critères de renforcement ou d'affaiblissement des connexions entre les neurones. Chacun des neurones du réseau réagit individuellement sans ralentir tout le système. Ils sont tous capables de produire l'agencement des informations, de mémoriser celles qui sont entrées et de les traiter de façon autonome selon ses propres besoins.

Le réseau neuronal créé par les chercheurs connexionnistes est constitué d'un système de liens entre les « neurones ». Comme dans les systèmes biologiques, il y a un passage constant d'informations entre les connexions neuronales qui constituent le système. Pendant une expérience avec le réseaux, on remarque que l'information circule plus ou moins facilement selon l'état global du système. La résistance au passage d'une information peut signifier un manque d'éléments pour son traitement, et que le réseau attend d'autres informations pour donner une réponse.

117G.E. Hinton et J.A. , Anderson "Parallels Models of associative Memory", Hillsdale, N. J. , Lawrence, Erlbaum, 1981. (ref. bibliographique M. Boden, op. cit p. 524).
118 Cf. G. Hinton ; T. Sejnowsky et D. Ackley, " A Learning Algorithm for Boltzman Machines", Cognitive Science, 1985, 9, p. 147-169. Voir aussi J. C. Perez, op. cit. , p.82. Les chercheurs néo-connexionnistes rendent hommage Ludwig Boltzmann (1844-1906) un des créateurs de la thermodynamique et de la mécanique statistique.
119J.C. Perez, op. cit. , p.82.

Plus une voie, ou un lien entre les «neurones», est utilisée, plus d'éléments d'information et de mémoire elle doit contenir. Cela facilite le flux des informations concernées et le filtrage des informations nouvelles. La mémoire et la capacité de traitement du réseau dépend de la configuration totale du système. La mémoire n'est pas ici un élément à part; elle est intrinsèque à la configuration globale du réseau. Chaque connexion constitue une unité de mémoire qui est en rapport avec les autres unités de mémoire.

Il faut remarquer que chaque cellule du réseau neuronal connexionniste, ne reçoit pas arbitrairement une adresse comme dans le cas des systèmes classiques de traitement digital d'information. L'adresse de chaque « neurone » est donnée par son contenu. Les informations sont agencées, repérées et rangées selon le contenu et les besoins du « neurone ». Cela a un effet global sur le système qui devient capable de réagir de façon fortement parallèle, imitant en quelque sorte les processus associatifs utilisés par le cerveau[120]

Pour finir cette section sur le néo-connexionisme, nous allons tout simplement compléter notre liste mentionnant d'autres recherches et chercheurs[121] représentatifs de cet courrant:

(a) Le système WIZARD créé par Igor Aleksander en 1985 qui comporte un réseau neuronal où chaque neurone correspond à une unité de mémoire RAM (*Randomic Acces Memory*). Il s'agit d'un système permettant de reconnaître des visages humains, tâche qui est d'ailleurs considérée très difficile dans le domaine des recherches sur la vision artificielle.

(b) Le projet "Cellular Automata Machine" (1984) coordonné par Tomaso Toffoli. Il s'agit d'une recherche qui vise à développer les automates cellulaires et à aboutir dans un court délai au succès avec un autre projet affilié dont le nom est "Connection Machine"[122]. Ce dernier projet à son tour vise à étendre le développement de machines dotées d'un certain "sens commun" et de capacités de raisonnement très fines.

(c) Les travaux de D. Waltz et de J. B. Pollack de l'Université de l'Illinois qui constituent une recherche assez intéressante sur la désambiguation du langage naturel au moyen de systèmes parallèles.

---

120 Cf. M. Boden, op. cit. , p. 483.
121 J. C. Perez, op. cit. , pp.65-89. Les recherches néo connexionistes sont résumées de façon très claire dans cet ouvrage. Nous l'utilisons tout au long cette partie de notre travail sur le sujet.
122 Pour un bon aperçu sur cette recherche, voir Personnaz, L. Dreyfus, G. Guyon, I. , "Les machines neuronales", La Recherche, n° 204, novembre, 1988, vol.19, pp.1362-1371.

(d) La recherche développée par J. J. Hopfield du California Institute of Technology est influencée par l'approche descendante. Cependant elle est aussi une recherche ascendante *stricto sensu*, car ce chercheur fait actuellement des études sur les circuits neuronaux élémentaires des escargots. Son travail utilise l'approche paralléliste et heuristique en même temps.

(e) Les travaux de Abu-Mustafa et Psaltis peuvent représenter ce qu'il y a de plus pointu en termes de recherche connexionniste. Ces deux auteurs travaillent sur des réseaux neuronaux et sur un système holographique pour la reconnaissance des formes. L'holographie constitue un thème peu exploité encore en I.A, mais qui se révèle très prometteur pour l'avenir de cette recherche[123]

La recherche en IA sur le traitement parallèle de données à partir des réseaux neuronaux est considérée par quelques auteurs comme la meilleure alternative à la lenteur des modèles informatiques classiques basés sur l'architecture de von Neumann. On dispose actuellement d'un bon support technologique pour construire des machines massivement parallèles. Par contre, en dépit de la disponibilité, de nos jours, de plusieurs modèles parallèles, on n'a pas encore défini une architecture que soit assez souple et efficace permettant la création de logiciels "parallèles".

Les limites des machines parallèles consistent en ce qu'il est très difficile et quelquefois impossible de programmer ces machines, qui, d'ailleurs, sont encore au stade expérimental. Les machines parallèles issues des laboratoires de recherche sont créées pour faire des tâches très spécifiques. Le problème à résoudre doit correspondre à son architecture. C'est pour cette raison que plusieurs chercheurs comme Winograd, MacCarthy, Schank et d'autres sont réticents face à cette approche innovatrice. Cela représente une certaine résistance ou si on veut une précaution due au fait que la plupart des recherches faites en IA ont été toujours menées sur des machines sérielles à l'archicterure de Newman, et par conséquent, sur des paradigmes de programmation en général compatibles avec cette l'architecture.

Les machines connexionnistes semblent adaptées à certains problèmes mais laissent à désirer ou sont complètement inadéquates quand il s'agit de traiter des données par des moyens strictement logiques et à caractère déductif.

---

123 Cf. Y., Abu-Mustafa, et D., Psaltis, "Des ordinateurs optiques à l'image du cerveau", Pour la Science, mai, 1987, pp. 71-80.

Est-il possible à l'heure actuelle de savoir si les approches ascendante et descendante sont complémentaires? Certains chercheurs comme F. Varela répondent que oui. Pour eux, ces deux approches peuvent être reliées:

> La réponse la plus évidente à cette question est que ces deux approches sont complémentaires, l'une ascendante et l'autre descendante, ou qu'elles pourraient être conjuguées dans un mode mixte, ou encore utilisées à des niveaux ou stades différents[124].

Selon Varela l'approche symbolique descendante et le sub-symbolique ascendante sont des approches complémentaires de la recherche en IA.

Cependant on doit considérer le fait que les recherches connexionnistes sur le parallélisme ont à peine commencé. Plusieurs auteurs, impressionnés par le développement du connexionisme, croient qu'il y a un avenir pour des machines qui soient à la fois sérielles et parallèles. Certains auteurs croient qu'il va falloir une coopération entre les capacités spécialisées des machines connexionnistes et le potentiel formel des systèmes symboliques. Cela veut dire, en effet, qu'une coopération entre l'approche connexionniste et l'approche descendante, basée sur la computation symbolique, s'impose.

Conclusion:

L'idée qu'un ordinateur digital adéquatement programmé peut avoir des états mentaux, c'est-à-dire, qu'il est possible de créer une intelligence artificielle, est en rapport avec un ensemble de connaissances et techniques. (*logos*).

L'IA est basée sur des modèles neuronaux et sur des modèles formels de cognition, elle est constituée d'un ensemble de théories et techniques lesquelles présentent parfois un caractère anthropomorphique et peuvent être parfois divergentes (approche ascendante-descendante). Les techniques de l'IA visent en général à programmer sur des ordinateurs digitaux certaines tâches en rapport avec la vision (reconnaissance de formes), l'utilisation du langage naturel, la résolution de problèmes etc. Ceux qui travaillent dans ce domaine s'intéressent à l'aspect formel et/ou biologique qui sont derrière les processus cognitifs humains, dans la mesure où cela permet de concevoir de nouveaux modèles et techniques capables de rendre possible la programmation et la compréhension de la façon dont les êtres humains acquièrent et traitent les informations.

124F. Varela, op. cit. , p. 80.

Nous ne sommes pas en mesure d'évaluer si l'IA se développera vers de nouvelles conceptions d'ordinateurs (en dehors des architectures traditionnels) ou vers des nouvelles techniques de programmation. Une chose est certaine, l'IA veut s'éloigner de ces mythes et éviter, comme le signale M. Longeart, toute "utopie mystificatrice"[125] en cherchant à s'affirmer en tant que logos.

Certains auteurs comme Dreyfus et Searle associent l'IA à une sorte de mythe. Le mythe est caractérisé par le fait que les chercheurs de l'IA oublient les limitations théoriques et techniques de leurs recherches en affirmant que les machines peuvent réaliser, *literalement*, certains tâches pour lesquelles on requiert de l'intelligence humaine. D'un autre côté, l'IA peut être associée à la notion de mythe lorsqu'elle est associée aux origines préhistoriques de l'idée de la conception et de la construction d'êtres artificiels doués de facultés humaines.

L'opposition entre mythe et logos peut être appliquée à une analyse épistémique de l'IA, le mythe représenterait la préhistoire et le logos l'histoire proprement dite. Le mythe n'offre pas d'éléments épistémologiques importants pour l'analyse critique de l'IA, tandis que le logos, comme nous allons voir dans le prochain chapitre, est en rapport avec la tradition scientifique et philosophique en Occident.

Descartes marque le passage du mythe au logos. Avec lui nous trouverons un modèle de rationalité qui, ajouté à l'objet technique (développement technologique), prendra la place de l'artifice mythique, sans qu'il y ait eu une rupture entre Logos et Mythe. Le logos de l'IA est en rapport avec toute une mythologie fondatrice qui est à la base de l'intérêt humain de créer des répliques de soi-même et de son esprit. C'est exactement cela qui est exprimé dans le passage suivant de Gérard Guièze:

> Si l'univers technique est bien le reflet de la modernité, son sens se constitue bien en rapport avec toute une mythologie fondatrice d'une sensibilité par exemple à l'aspect prometteur de la technique, rendue possible par la substitution d'un univers de l'artifice à l'univers de la nature. S'interroger sur la technique, c'est bien s'interroger sur son essence, mais pour tenter aussi d'y déchiffrer des concepts fondamentaux qui régissent son existence. D'où l'intérêt de toute une mythologie de l'artefact où la vérité de la technique s'annonce à travers une représentation de la Nature comme lieu de réalisation de l'être[126].

---

125 cf. M. Longeart, (1989) p. 150.

126 G. Guièze, "Les énoncés de Minds and machines comme image de l'objet technique" (Présentation) in Pensée et machine, Éditions du Champ Vallon, France 1983, p. 23 (traduit de l'américain Minds and Machine, Prentice-Hall Inc. ,1964, par Patrice Blanchard).

Le développement scientifique et technologique a été la cause principale qui a rendu possible d'envisager la création d'une intelligence artificielle. Cependant c'est dans le sein de la philosophie occidentale que les scientifiques vont trouver l'appui fondamental pour rendre parfaitement rationnelle la conception de machines intelligentes. L'idée qu'on peut tout représenter par des moyens formels de caractère mathématique ou logique fournissent les bases pour la constitution du logos de l'IA.

C'est en tant que résultat de tout un ensemble de principes de rationalité, d'idées scientifiques et de connaissances technologiques que le thème de l'intelligence Artificielle est aujourd'hui objet d'investigation scientifique. Sans la consolidation d'un modèle représentationnel de la pensée et sans le développement théorique de l'informatique, il est fort probable que l'IA (ou n'importe quelle enterprise ayant les mêmes buts qu'elle.) serait restée dans sa phase mythique, liée aux légendes ou à ce qu'on connaît aujourd'hui sous le nom de science-fiction.

Ils se servent de figures visibles et ils raisonnent sur ces figures, quoique ce ne soit point à elles qu'ils pensent, mais à d'autres auxquelles celles-ci ressemblent. Par exemple c'est du carré en soi, de la diagonale en soi qu'ils raisonnent, et non de la diagonale telle qu'ils la tracent et il faut en dire autant de toutes les autres figures. Toutes ces figures qu'ils modèlent ou dessinent qui portent des ombres et produisent des images dans l'eau, ils les emploient comme si c'étaient aussi des images, pour arriver à voir ces objets supérieurs qu'on n'aperçoit que par la pensée.[127]

Platon

[127]Platon, République, VI, 510 c-510 e , Traduit par E. Chambry, Garnier Paris, 1958, p.247.

CHAPITRE II

Les rapports entre l'Intelligence Artificielle et la philosophie

Présentation:

Les fondements philosophiques du logos de l'Intelligence Artificielle

Emmanuel Kant, préoccupé par la portée du jugement téléologique, se demandait, à la suite de René Descartes, comment on pourrait attribuer des fins à des êtres (sans intelligence) qui sont définis par les lois générales de la matière et du mouvement[128]. Aujourd'hui, les questions mettant en rapport les capacités l'humaines et de la machine ne cessent d'être posées. Plus on essaye de rendre les machines intelligentes en les faisant résoudre de problèmes humains plus on est obligé de réfléchir sur la pensée humaine, comment les êtres humains sont capables d'agir de façon intelligente:

> L'homme-machine? Question qui a toujours fasciné les philosophes et qui pourrait aujourd'hui s'énoncer ainsi: L'inscription matérielle dans le cerveau d'un système computationnel isomorphe à une machine de Turing Universelle rend-elle compte de notre capacité de perception, de l'utilisation du langage, de l'aptitude à concevoir, à croire, à désirer? Plus généralement, peut-on considérer cette machine abstraite comme une théorie formelle des processus de pensée dont le système nerveux assurerait la «matérialisation»?[129].

L'IA rassemble des connaissances de plusieurs domaines tels que l'informatique, la neurologie, la psychologie, la linguistique, la logique et la philosophie[130] Les rapports de

---

128 E. Kant, *Critique du Jugement*, 1770, 2º partie, 1º section, § 64-65 traduction G. Gibellin, 4e édition, Librairie Philosophique J.Vrin, 1962.

129 M. Borillo, "Une machine spéculative: Informatique, intelligence artificielle et recherche cognitive", *in Revue Internationale de Philosophie*, 1/1990, nº 172, p.47.

130 L'informatique fournit les éléments de base à l'IA en termes de matériel (hardware) et de methodologie de travail baseés sur la manipulation symbolique de données. La neurologie est aussi importante, car les découvertes dans ce domaine ont permi et permettent de développer des modèles neuronaux capables de servir de base théorique à de nouvelles conceptions en IA. La Psychologie a contribué à mettre en évidence plusieurs données importantes sur le comportement intelligent pouvant être programées sur une machine. Actuellement la psychologie cognitive, contribue au développement de l'IA, avec ses travaux sur le traitement de l'information par les systèmes cognitives humains au même temps qu'elle utilise des programmes de l'IA comme des outils à la modélisation des processus cognitives qu'elle étudie. La linguistique est à la base des premiers travaux sur la traduction automatique et donne une contribution importante aux

l'IA avec la philosophie sont importants. Plusieurs auteurs associent l'IA à la philosophie dans le sens où elle est en rapport avec l'entreprise de l'homme de comprendre sa propre pensée par l'intermédiaire de spéculations philosophiques, de la recherche logique ou des démarches scientifiques:

> HOW IS IT POSSIBLE for a physical thing—a person, an animal, a robot—to extract knowledge of the world from perception and then exploit that knowledge in the guidance of successful action? That is a question with which philosophers have grappled for generations, but it could also be taken to be one of the defining questions of artificial intelligence. AI is, in large measure, philosophy. It is often directly concerned with instantly recognizable philosophical questions: What is mind? What is meaning? What is reasoning and rationality? What are the necessary conditions for the recognition of objects in perception? How are decisions made and justified? [131].

D'autres auteurs voient l'IA plutôt comme un sujet scientifique d'importance pour les philosophes:

> Intelligence artificielle, logique et informatique sont aujourd'hui étroitement solidaires. L'honnêteté humaine qui ne s'est pas tenue au courant des développements rapides de ces trois disciplines risque de ne pas voir la portée philosophique de ceux-ci. Or, jamais une réflexion philosophique n'a été aussi nécessaire si on veut luter "contre la transparence que les produits de la raison opposent à la raison elle-même", selon la forte expression de Jules Vuillemin.
>
> L'Intelligence Artificielle apporte aujourd'hui des matériaux à la réflexion philosophique. Il n'y a pas antinomie. Mais la complémentarité entre la philosophie réflexive et la philosophie qui prend pour objet un domaine aussi étranger à son essence que l'Intelligence Artificielle. Paul Ricour l'a dit de façon très explicite: "la réflexion de la pensée sur elle-même se découvre...médiatisée par la réflexion sur des objets que la philosophie n'a pas engendré, mais qu'elle se trouve hors d'elle-même. Il s'agit ici, fondamentalement, des contextes

---

recherches sur le langage naturel. Les méthodes de réprésentation logiques inspirent la création des premiers programmes et langages de l'IA tels que le LOGIQUE THEORISTE, LISP et PROLOG. Ces méthodes constituent aussi une base importante pour la représentation des connaissances en IA. L'application de la logique à l'IA se montre très opportune du point de vue épistémologique, car en agissant de façon *métalogique*, sur les programmes de l'IA, la logique permet d'obtenir des bases formelles assez puissantes pour la représentation des connaissances. Elle permet aussi, de prouver la consistance, la complétude et la décidabilité des formalismes et d'analyser la cohérence théorique des programmes d'ordinateur produits dans ce domaine de recherche. Pour un aperçu sur les recherches logiques en IA voir A. Thayse, *Approche logique de l'intelligence Artificielle*, Tome 1, "De la logique classique à la programmation logique", Dunod Informatique, Paris, 1988. Il est intéressant de remarquer les travaux du groupe Léa Sombé en France qui développe actuellement une recherche sur des raisonnements logiques en IA à partir des informations incomplètes. Sur ce sujet voir encore le tome 2 de l'ouvrage qu'on vient de mentionner(Dunod Informatique, Paris, 1989) qui donne un aperçu des contributions logiques à l'IA à partir de logiques non-classiques jusqu'à la logique des bases de données. Voir aussi *Raisonnements sur des informations incomplètes en Intelligence Artificielle*, Toulouse : Teknea, 1989. Cet dernier livre présente quelques résultats de la recherche logique en IA du groupe Léa Sombé.

131 D.C. Dennett, "When philosophers Encounter Artificial Intelligence", *in* Stephen R. Graubard, éd. , *The Artificial Intelligence Debate, False Starts, Real Fondations*, MIT Press, Mass. ,1989, p. 283.

> scientifiques dont le philosophe doit apprendre les règles propres de constitution[132].

Nous pensons qu' il y a un rapport important entre l'IA et la philosophie, ce rapport tient au fait que certaines thèses philosophiques et celles défendues en IA se ressemblent beaucoup. Nous pouvons affirmer que l'IA est ancrée dans notre tradition philosophique et que pour cette raison, même si la philosophie n'a pas été impliquée dans les premières expériences sur les machines intelligentes elle a toujours été sous-jacente aux idées fondamentales qui orientent les travaux en IA. Nous ne croyons pas que la philosophie soit si étrangère aux discussions en IA.

L'IA en tant que *logos* est largement discutée en sciences humaines et dans plusieurs domaines scientifiques et techniques. Les préoccupations des philosophes et celles de ceux qui travaillent en IA sont, à plusieurs égards, convergentes. Quelques philosophes cherchent à appliquer les ressources méthodologiques et logiques qu'ils trouvent dans les travaux en IA pour discuter et essayer de résoudre plusieurs problèmes traditionnels au sein de la philosophie[133], comme par exemple la question des rapports entre l'esprit et le cerveau. D'autres se concentrent sur la critique des méthodes et fondements épistémologiques des thèses de l'IA.

La conception des programmes dans le domaine de l'IA exige tout un travail théorique, lequel est en rapport avec certains problèmes, épistémologiques, logiques et linguistiques d'intérêt philosophique tels que les discussions sur l'arrière-plan des connaissances humaines, sur les notions d'action, de probabilité, causalité et intentionnalité. Comme exemple des questions qui peuvent intéresser à la fois les philosophes et ceux qui travaillent dans le domaine de l'IA nous pouvons citer celles sur les rapports entre les raisonnements et la causalité (la notion de causalité), les raisonnements pratiques, les rapports entre les actions et l'intentionnalité, les rapports entre l'implication conversationnelle et la pensée et enfin les considérations sur les actes du discours et leur logique. L'IA reçoit et donne des contributions, non seulement à l'épistémologie et à la philosophie de l'esprit, mais à bien d'autres domaines de

---

132 P. Gochet, (Préface au numéro dédié à l'IA) *Revue Internationale de Philosophie*, Université de Liège. 1/1990, n° 172, Difusion Presses Universitaires de France, Hassocks, Harvester Press, 1979, 244 p. , pp. 3-4.

133 L'IA a eu surement un impact sur la philosophie en particulier sur la philosophie de l'esprit, domaine où elle a permis de donner de nouvelles réponses aux questions classiques sur la nature de l'esprit, permettant, dans un certain sens, la reformulation de certaines questions philosophiques sur ce sujet. L'IA stimule la discussions de problèmes et de sujets philosophiques tels que l'intentionnalité, les rapports entre le corps et l'esprit, la nature sémantique des processus cognitifs de compréhension, etc.

la philosophie comme la phénoménologie, l'herméneutique, la philosophie de la science, la philosophie de l'action et la philosophie du langage[134].

Afin de donner suite à la discussion du premier chapitre et en même temps d'analyser un aspect important du rapport entre l'IA et la philosophie nous nous proposons de discuter brièvement les fondements philosophiques de ce que nous appellons le logos de l'IA. Nous avons défini le "*Logos*" de l'IA comme étant tout un ensemble de connaissances technologiques, d'idées scientifiques et philosophiques qui constituent les principes de rationalité présupposés par la notion d'"Intelligence Artificielle".

Pour analyser les fondements philosophiques du logos de l'IA et montrer les rapports entre l'IA et la philosophie, nous nous proposons de montrer que l'IA est héritière de la tradition représentationnaliste selon laquelle toute connaissance peut être représentée au moyens de règles générales univoques. (section 1 de ce chapitre). Nous allons montrer aussi, tout en situant l'IA par rapport à quelques théories sur l'esprit, que parler sur les machines intelligentes exige l'adoption d'une théorie philosophique sur l'esprit, capable d'exprimer l'idée d'une pensée mécanique qui soit cohérente et réalisable. (section 2 de ce chapitre).

1- La tradition représentationaliste et l'Intelligence Artificielle

Depuis Platon l'homme cherche une théorie capable de lui permettre de s'expliquer et de se connaître soi-même et son esprit. La recherche d'une telle théorie a été presque toujours basée sur une sorte de représentation en tant que calcul, comme la géométrie, les mathématiques ou la logique. Cette même entreprise se poursuit lorsqu'on essaye d'élaborer une mécanisation de la pensée c'est-à-dire, l'effort de concevoir certains mécanismes capables de reproduire le raisonnement humain et même la pensée au moyen de représentations formelles.

L'IA est un projet de mécanisation de la pensée qui présente un intérêt philosophique. Nous avons vu que la question de la mécanisation de la pensée fût déjà posée par Descartes lorsqu'il s'interroga sur les capacités des automates de son époque. Après Descartes la

---

134 Cf. J. MacCarthy and P. J Hayes, "Some Philosophical Problems from the Standpoint of Artificial Intelligence", *In* B. Meltzer and D. Michie éds. , *Machine Intelligence*, vol. 4, Halstead, NY, 1969, pp. 463-502.

question de la mécanisation de la pensée réapparaît sous une nouvelle perspective: Turing propose une façon originale de répondre à la question sur l'intelligence des machines.

Nous avons constaté qu'il a toujours existé un effort philosophique, de répondre à la question sur la possibilité de mécanisation de la pensée. L'homme continue à chercher une théorie capable d'expliquer le caractère mécanique de la pensée décrite en termes physiques ou formels.

D'une façon ou de l'autre la pensée a été toujours analysée par son caractère représentationnel par des philosophes tels que Leibniz (1646-1716), T. Hobbes (1588-1679), J. Locke (1632-1704) et D. Hume (1711-1776), lesquels considéraient la pensée soit comme un phénomène physique soit comme un mécanisme de calcul. Les idées de ces philosophes sont derrière la conception mécanique de l'esprit, selon laquelle la pensée obéit à des lois physiques exactement comme les corps célestes et que l'esprit est un processus computationnel ou mécanisme de calcul.[135]

1.1- Le modéle représentationnaliste inspiré de la physique

La notion de représentation et l'appel à des règles sont des outils théoriques pour les chercheurs en IA. Depuis que les Grecs ont développé la logique et la géométrie nous connaissons des conceptions selon lesquelles notre pensée peut être assimilée à une sorte de calcul. C'était, par exemple, une conception existante dans la philosophie moderne. Les philosophes empiristes et rationalistes considéraient eux aussi possible l'explication du raisonnement et même de la pensée par des moyens formels. Cette idée sera développée par d'autres philosophes et mathématiciens contemporains qui se consacrent définitivement à la notion de mécanisation de la pensée. (C'est en partie sur ces éléments que se sont fondées les bases philosophiques du logos de l'IA.).

Les rapports de l'IA avec la tradition philosophique font en sorte que l'IA et la philosophie sont des sujets étroitement liés. Nous allons parcourir quelques éléments saillants de cette tradition fondée sur la confiance dans les modèles de réprésentation formelle. Pour comprendre la tradition il faut voir comment les énoncés vrais en sciences sont, de plus en plus soutenus par des représentations: symboles conventionnels et règles.

Ceux qui appartiennent à la tradition métaphysique représentationnaliste conçoivent la pensée comme une sorte de mécanisme de calcul. Par l'expression "tradition métaphysique

---

135 *Cf* M. A. Fischler et O. Firschein, *op.cit.* p.9

représentationnaliste" nous entendons tous les efforts d'explication basés sur des réprésentations qui sont à la base du développement de la pensée scientifique occidentale. La tradition métaphysique représentationnaliste n'est rien de plus qui la tradition philosophique qui valorise les règles, les démonstrations, et toute sorte de conventions et de notations capables de permettre d'expliquer et réprésenter formellement les phénomènes de la nature et de la pensée. Le rationalisme et l'empirisme logique, le mécanisme, le fonctionnalisme et l'IA, dans ce sens font tous partie de cette tradition[136].

La notion de représentation n'est pas univoque; il y a plusieurs concepts de "représentation", mais en général le terme désigne l'opération de l'esprit d'avoir une image mentale, une idée ou un concept qui correspond à un objet externe à l'esprit lui-même. La notion de représentation est en rapport intrinsèque avec la notion de connaître. Depuis G. d'Ockham (Quodl. , IV, q.3) la représentation est le moyen par lequel on connaît quelque chose[137]. La notion de représentation entendue ici dans le sens général d'Ockam joue un rôle important pour la philosophie et pour la science moderne, principalement en ce qui concerne son expression mathématique et logique qui constitueront l'orientation majeure de la méthode scientifique.

Nous ne pouvons exposer ici toutes les notions de représentation existantes en philosophie ni discuter les problèmes philosophiques liés à la notion de représentation. Nous voulons tout simplement montrer que les rapports entre l'IA et la philosophie tiennent au fait qu'en faisant partie de la tradition scientifique occidentale l'IA adopte la perspective philosophique représentationnaliste lié à des conceptions métaphysiques lointaines.

L'idée principale que l'IA hérite de la tradition philosophique représentationnaliste peut être résumée par la thèse générale selon laquelle notre pensée et nos comportements ont derrière eux une théorie qui peut être exprimée par le moyen de représentations formelles. L'idée exposée ci-haut est sous-jacente aux thèses de l'IA selon lesquelles le comportement de tout système peut être compris et même reproduit par un autre système pourvu qu'il soit bien représenté par des moyens logiques ou mathématiques. Cette conception à caractère mécaniste, a comme antécédents des idées philosophiques chères à plusieurs philosophes et physiciens anciens et modernes qui ont vu dans les

---

136 L'on peut classifier tout les efforts humains pour chercher des représentations non-équivoques capables de fournir des explications sur les choses, comme étant dans le domaine de la tradition représentationnaliste. Ainsi, certains courants philosophiques, les mathématiques modernes, la logique et la géometrie constituent la tradition représentationnaliste en Occident.

137 N. Abbagnano, *Dicionário de Filosofia*, éd. Mestre Jou, São Paulo,1982, p.821.

représentations formelles la clé pour la comprèhension de certaines énigmes de la nature et de la pensée.

Nous allons mentionner, en ce qui concerne la tradition représentationnaliste, quelques philosophes et scientifiques qui ont mis l'accent sur des représentations et qui ont souligné l'importance de l'application de régles comme garantie à l'obtention des connaissances en plusieurs domaines et en particulier pour l'explication de la pensée.

La réussite des sciences physiques à partir du XVIe a été un facteur important pour la création des bases épistémologiques de l'IA. Le développement de la science et des mathématiques modernes a amené l'homme à croire que le calcul pourrait lui fournir des connaissances sur tout le fonctionnement de l'univers, et particulièrement, sur lui-même et son esprit.

Les premières conceptions qui sont à la base de la physique moderne nous donnent une idée de l'influence et de l'importance de la notion de représentation comme moyen de développement de la connaissance. Galilée et Descartes, lorqu'ils donnent complète crédibilité aux calculs mathématiques pour comprendre les phénomènes astronomiques et de la nature, fondent la croyance que la nature avait une structure mathématique analogue à celle du calcul. Galilée disait que le monde était une machine dont le fonctionnement était parfait. Descartes ajoute que ce monde mécanique était aussi peuplé de machines parmi lesquelles quelques-unes (les êtres humains) seraient faites selon un modèle divin et douées d'un esprit. Pour Galilée comme pour Descartes, la connaissance sur la forme de raisonnement ou de calcul est considérée comme la contrepartie mentale (une espèce de portrait exact) de l'objet connu.

Galilée inaugure, selon plusieurs épistémologues, une nouvelle perspective scientifique s'opposant à la crédulité de la Renaissance fondée sur un monde magique et mystique. En permettant une mathématisation de la physique, en réduisant définitivement le réel au géométrique, Galilée occasionnera l'affaiblissement graduel de la force explicative du système d'interprétation aristotélicien de l'univers exposé dans le *De Coelo*. Le système de représentations mathématiques de la physique post-galiléenne représente une rupture avec une image du Cosmos attachée à la connaissance sensible et propose par contre, un schéma d'univers qui, une fois soumis à la discipline rigoureuse du calcul abstrait, peut être objet de la connaissance humaine.

> Philosophy is written in that great book, the universe, which is always open, right before our eyes. But one cannot understand this book without first learning to understand the language and to know the characters in which it is written. It is written in the language of mathematics, and the characters are triangles, circles,

> and other figures. Without these, one cannot understand a single word of it, and just wanders in a dark labyrinth[138].

Galilée donne une crédibilité définitive aux méthodes de représentation mathématique, lesquelles permettront plus tard une axiomatisation croissante des différents savoirs humains. Cette nouvelle compréhension des choses à partir de moyens formels est définitivement adoptée par la tradition scientifique et constitue par conséquent la base théorique représentationnaliste des recherches en IA.

L'IA a des bases épistémologiques qui ont des antécédents dans la philosophie rationaliste ainsi que dans la tradition de l'empirisme. L'empirisme et le rationalisme constituent les deux bases épistémologiques représentationnalistes de l'IA.Thomas Hobbes, par exemple, inspiré de Galilée, a vu dans le calcul un moyen efficace de connaissance et aussi de compréhension de l'esprit. Le thème de l'esprit n'était pas le plus important pour le physicien italien, mais pour Hobbes, il était possible de comprendre mathématiquement la pensée au moyen de représentations. L'idée de Hobbes était que le raisonnement serait une sorte de mécanisme de calcul mental. Cela inaugure non seulement une nouvelle conception de l'esprit, mais aussi la formulation moderne des fondements philosophiques de l'IA.

Hobbes, allant au delà de ce que proposait la physique de Galilée, croyait que l'esprit fonctionnait selon les mêmes principes mécaniques que la nature. Pour Hobbes, la pensée était constituée de symboles physiques qui seraient sujets à la manipulation selon les lois du calcul ou du discours rationnel. Il a créé un modèle mécaniste de la pensée. Il croyait, influencé par les idées empirites de son époque, que l'esprit était un élément naturel sujet à des manipulations physiques réelles de ses composantes, symboles physiques ou parties de pensées. Pour lui, un raisonnement pourrait être compris comme des parties de la pensée dans le corps humain. Ces parties de pensée pourraient être manipulées comme des éléments d'un système physique. Elles étaient, pour Hobbes, comme des particules simples

---

138 Galilée, cf. G. Barbera, éd., *Le Opere de Galileo Galilei*, Nuova Ristampa della Edizione Nazionale, Ministero della Pubblica Istruzione, Rome, 1968 Vol. VI, section 6 p. 232. Cité et traduit de l'italien par Haugeland *op. cit*, p. 19. Selon Haugeland, Galilée ne concevait la géométrie qu'en tant qu'étude des figures et des relations dans l'espace. Intéressé au problème du mouvement, il a utilisé la géométrie euclidienne pour représenter des variables physiques. Par exemple, pour lui, une droite ne représente pas une trajectoire ou une distance, mais un temps et une vitesse. Ainsi, il amplifie les possibilités de la géométrie d'Euclide. La géométrisation du mouvement proposée par Galilée est définie à partir de modèles de représentation qui ne font plus appel à l'expérience sensible; tout ce qui est concret est réduit à des représentations abstraites, des configurations géométriques. En découvrant que le mouvement des planètes pourrait avoir une description formelle précise et cohérente sans faire appel à des causes d'ordre divin, Galilée propose une nouvelle façon de représenter l'univers et une nouvelle vision de monde. Nous allons voir dans le chapitre suivant que pour Dreyfus la réalisation du projet de l'IA aurait besoin d'un "Galilée de l'esprit" capable de fournir des représentations formelles assez puissantes pour permettre de formaliser et reproduire la pensée.

qui se trouvent en mouvement dans la nature: Ces parties de pensée, pouvaient encore être décomposée par des représentations mathématiques.

Hobbes concevait la pensée comme "discours mental"[139]; il est un des premiers philosophes, selon Haugeland, à avoir eu l'intuition que la sémantique pourrait subir une réduction formelle par les moyens d'une méthode de manipulation syntaxique, et ainsi rendre compte des opérations complexes du raisonnement humain:

> The belief that such a total formalization of knowledge must be possible soon came to dominate Western thought. It already expressed a basic moral and intellectual demand, and the success of physical science seemed to imply to sixteenth-century philosophers, as it still seems to suggest to thinkers such as Minsky, that the demand could be satisfied. Hobbes was the first to make explicit the syntactic conception of thought as calculation: "when a man reasons, he does nothing else but conceive a sum total from addition of parcels," he wrote, "for REASON.. . is nothing but reckoning..."[140]..

Une conception de la pensée semblable à celle de Hobbes (où il est possible d'analyser le calcul mental en termes de ses éléments simples) peut être trouvée aussi dans les travaux d'un autre empiriste, D. Hume. Pour ce dernier, la pensée peut être décomposée en atomes d'expérience ou en ensembles d'impressions que nous avons des choses.

Hume croyait qu'il était possible, en appliquant les méthodes des sciences de la nature aux sciences humaines, d'asseoir les bases d'une sorte de "mécanique mentale". Pour lui, la connaissance pourrait être décomposée en éléments plus simples qui seraient soumis à des forces et à des opérations mentales, en particulier des associations d'idées. Il n'y a pas chez Hume de manipulations de symboles par l'esprit dans un sens réel comme dans les théories proposées par T. Hobbes, qui croyait que la pensée fonctionnait selon une mécanique composée de mouvements de matières au sein du cerveau humain.

La notion d'association d'idées chez Hume est basée sur les idées de John Locke ainsi que sur celles de Hobbes. Il s'inspire fortement aussi de la physique de Newton, mais les pensées ou associations d'idées ne sont pas vraiment physiques, car elles sont, selon Hume, *comme* des événements physiques. Il y a selon Dreyfus, une variation contemporaine de l'associationnisme de Hume qui est en rapport avec l'IA:

---

139 Cf. J. Haugeland, *op.cit.* , p.23.
140 H. L. Dreyfus, *op. cit.* (1979), p.69.

> The development of the high-speed digital computer has strengthened a conviction which was first expressed by Lucretious, later developed in different ways by Descartes and Hume, and finally expressed in nineteenth-century associationist or stimulus-response psychology: Thinking must be analyzable into simple determinate operations. The suitably programmed computer can be wiewed as working model of the mechanism presupposed by this theory. Artificial intelligence has in this way made associationism operational and given it a second wind[141]

Hume, affirme Haugeland, a été un des premiers philosophes à voir, dans les méthodes représentationnelles de la physique moderne, un modèle pour la compréhension de la pensée. Il a développé, de façon plus systématique que ses contemporains, une conception mécaniste de la pensée en termes de méthode expérimentale de raisonnement ("expérimental méthode of reasonning"). Dans son *Traité sur la nature humaine*, ("*Treatise of Human Nature*"), il cherche à expliquer les processus de fonctionnement de l'esprit humain à partir de cette nouvelle méthode inspirée de la réussite des sciences physiques, et notamment des travaux de Newton[142].

La théorie de l'esprit de Descartes, plus élaborée que celle de Hobbes, propose clairement que les pensées sont des représentations symboliques. Cette idée est en rapport avec la mathématique de Descartes, laquelle est assise sur une notion de représentation où le symbole est distinct de l'objet symbolisé: l'esprit est distinct du monde, la pensée est une chose complètement distincte de l'objet qu'elle représente. Pour le philosophe français l'idée est le seul objet immédiat de la connaissance.

Ayant unifié l'algèbre et la géométrie, Descartes cherche à faire valoir les résultats de cette unification en proposant que les lois de la physique pourraient facilement être "représentées" sous la forme géométrique. Selon lui, tous les rapports géométriques trouvés à l'intérieur des lois physiques pourraient être exprimés de façon algébrique. Autrement dit, les lois et relations abstraites de la physique pourraient être complètement exprimées en termes d'équations algébriques:

> The new approach to representation has two distinct components: the negative half, rending symbol and symbolized asunder and the positive half, getting them back together again. The negative half is clearly manifested in the mathematical work: the basic point is that algebraic formulations don't intrinsically represent numbers, and Euclidean formulations don't intrinsically represent geometric figures (...) Extend this negative realization to mental representations (thoughts),

---

141 H. L. Dreyfus, "Alchemy and Aritificial Intelligence", *in The RAND Corporation*, Cal. , décembre, 1965, P- 3244, pp.48-49.
142 Cf. J. Haugeland, *op.cit.* , pp. 41-44.

> and you finally conclude the divorce of thought from thing, mind from world, that began so innocently in the old appearance / reality distinction[143].

En dépit de ce dualisme, il y a un point important dans la notion de représentation chez Descartes; pour lui, une notation mathématique (externe) qui soit propre à symboliser ou représenter différentes sortes de sujets doit correspondre à la notation (interne) des symboles dans l'esprit. Le fonctionnalisme, qui est une sorte de philosophie officielle de l'IA et des sciences cognitives, adopte cette idée cartésienne. Les préoccupations de Galilée, de Descartes et autres savants sont analogues, en quelques points à celles de certains chercheurs en IA, dans le sens où ceux-ci comme ceux-là se préoccupent de représenter les choses de telle sorte que cela leur permettrait de résoudre des problèmes dans un domaine donné.

> L'IA est ainsi l'héritière directe d'un courrant représentationnel/calculationnel dans la tradition philosophique et l'on peut considérer qu'elle soulève la question philosophique de savoir si l'esprit est en fait un système formel. Comme Descartes, les spécialistes de l'IA présupposent que tout processus de compréhension consiste à former et manipuler des représentations appropriées, que celles-ci peuvent s'analyser en éléments primaires, et que tous les phénomènes peuvent s'expliquer comme étant des ensembles rélationnels complexes de ces éléments primaires[144].

Selon Haugeland, Descartes pensait que les pensées n'étaient que des symboles dans un système de notation de l'esprit. Cet auteur américain fait quelques commentaires sur les faiblesses du système de représentation cartésien en ce qui concerne son caractère sémantique, mais il montre en même temps que Descartes reconnaissait déjà la plasticité (ou souplesse) de la cognition humaine en proposant l'idée que la pensée était un schéma de représentations basé sur le système de notation de l'esprit.

> Descartes could dream that up because he had a new vision of thoughts as mere symbols in a notational system; and he knew full well that such symbols could equally represent one subject matter or another, or none at all, and the system itself would be no different[145].

Descartes, concevait l'esprit ou les représentations mentales comme une espèce de *système* interne de notation qui réfléchirait parfois d'autres systèmes externes de

---

143 J. Haugeland, *op. cit.*, p. 32.
144 H. L. Dreyfus, "L'Intelligence artificielle (IA): le problème de la représentation du savoir", *in Encyclopædiaphilosophique Universelle*, vol. I, PUF, Paris, 1989, p.973.
145 J. Haugeland, *op. cit.*, p. 33.

représentation. Descartes appelait *représentation* l'image de la chose dans la pensée ou la symbolisation d'un sujet (*Méditations III*). Pour lui, nous avons des représentations lorsqu'une notation est propre à représenter son sujet.

Leibniz explique, lui aussi, la pensée comme un mécanisme de calcul. Pour Leibniz, il fallait créer un système de notation universel valable pour la représentation de tous les objets existants dans le monde. (*Dissertatio de arte combinatoria* - 1666) L'univers est, selon Leibniz, un ensemble ordonné de monades, ces monades étant, selon lui, un miroir du monde, car chaque monade est une substance spirituelle qui compose l'univers, leur place étant ordonnée hiérarchiquement.

Les monades qui occuperaient les places le plus importantes dans l'échelle hiérarchique seraient celles capables de représenter l'univers de façon plus claire et distincte[146]. Dans sa *Monadologie* (1747 §57) la monade est considérée comme un atome spirituel, un composant simple de l'univers. La contrepartie mathématique de ces idées s'exprime dans le passage suivant de l'ouvrage de Dreyfus,*What Computers Can't Do*:

> Leibniz thought he had found a universal and exact system of notation, an algebra, a symbolic language, a "universal characteristic" by means of which "we can assign to every object its determined characteristic number". In this way all concepts could be analyzed into a small number of original and undefined ideas; all knowledge could be expressed and brought together in one deductive system. On the basis of these numbers and the rules for their combination all problems could be solved and all controversies ended: "if someone would doubt my results, Leibniz said, I would say to him: 'Let us calculate, Sir,' and thus by talking pen and ink, we should settle the question. "[147].

Pour Leibniz, par exemple, un concept doit être décomposé en éléments plus simples pour qu'on puisse l'apprendre. Hume, de l'autre côté croit que les impressions qui composent notre expérience sont constituées "d'atomes d'expérience". Ces atomes peuvent être isolés de manière à nous permettre de comprendre la pensée. D'après Dreyfus, Wittgenstein, dans le *Tractatus*, définit aussi le monde comme un système d'atomes de faits qu'on peut exprimer sous forme de propositions logiquement indépendantes.

> (...) in Wittgenstein's *Tractatus*, where the world is defined in terms of a set of atomic facts which can be expressed in logically independent propositions. This is the purest formulation of the ontological assumption, and the necessary precondition of all work in AI as long as researchers continue to suppose that the

---

146 Cf. F.Heinemann, *A filosofia no século XX*, Fundação Calouste Gulbenkian, Lisboa, 1983, p.207, traduction de A. F. Morujão, de l'allemand *Die Philosophie in XX. Jahrhundert*, Zweite, Auflage, E. K. Verlag, Stuttgart, 1963.
147 H. L. Dreyfus, *op. cit.* , p. 69. Les parties entre guillemets sont extraites par l'auteur du texte de Leibnitz, *Zur Allgemeinen Characteristik* qui se trouve dans Wienner, Philip, *Leibniz Selections*, Scribner, NY,1951, pp. 18 et 25.

> world must be represented as a structured set of descriptions which are themselves built up from primitives. Thus both philosophy and technology, in their appeal to primitives continue to posit what Plato sought: a world in which the possibility of clarity, certainty, and control is guaranteed; a world of data structures, decision theory, and automation [148].

Une bonne partie de ceux qui travaillent dans le domaine de l'IA sont en accord avec cette façon de penser du premier Wittgenstein. Pour eux, il est possible de fournir une représentation formelle du monde, basée sur un système de descriptions dans lequel le monde est conçu comme un ensemble d'éléments atomiques plus simples.

La recherche visant la constitution d'un système de descriptions fortement structuré, capable de créer des représentations puissantes à partir d'éléments discrets calculables, définit bien le but de la tradition représentationnaliste. Ce but est aussi à la base des principales motivations de quelques chercherur en l'IA afin de comprendre l'esprit et se donner les moyens de réprésenter quelques de ses caractéristiques sur un programme informatique.

La tradition représentationnaliste fondatrice du modèle scientifique que nous avons actuellement est derrière la notion de représentation courante en IA, laquelle peut être résumée de la façon suivante: Étant donné que (1) notre intelligence resulte du fait que l'esprit peut *représenter* la réalité et que (2) par le moyen de représentations nous avons la chance d' avoir accès à certains éléments formels du fonctionnement de l'esprit, il est possible de rendre compte des comportements intelligents au moyen de modéles de la cognition executés par des programmes d'ordinateur.

1.2- La tradition représentationnaliste et la notion de règle

L'homme a toujours voulu savoir quelles sont les règles qui régissent son esprit ou qui sont derrière sa pensée ou ses raisonnements. La notion de règle est fondamentale pour la tradition réprésentationnaliste, il ne suffit pas de savoir tout simplement comment représenter certaines choses, on a besoin aussi d'un ensemble de règles permettant d'établir une cohérence entre la réprésentation et les choses qui sont réprésentées.

La notion de règle, comme la notion de représentation employée en IA est en rapport avec la tradition philosophique en Occident et constitue un des éléments importants pour la constitution des méthodes scientifiques employées en IA.

---

148 H. L. Dreyfus, *op. cit.* , pp.211-212.

Cette idée est attachée à la conviction que nos raisonnements et nos comportements intelligents, d'une manière générale, peuvent être décrits et représentés à partir d'un ensemble de règles formelles constituant une procédure effective ou programme d'ordinateur. La notion de règle est très générale, mais le sens du terme "règle" tel que nous l'employons ici, sera toujours celui qui se rapproche le plus de la notion de procédure effective.

Dans notre culture, l'homme a toujours manifesté de l'intérêt pour la découverte d'une base rationnelle de règles, logiques mathématiques ou méthodologiques, applicables soit sur le plan intellectuel, soit sur le plan pratique. Cela marque une tendance, en Occident, à créer des connaissances univoques et à éradiquer au maximum l'incertitude dans la constitution des savoirs. Depuis toujours, l'homme essaie de réduire les objets de son intérêt à des éléments discrets, plus simples et pouvant être compris à partir de règles aussi simples que possible.

> As we have seen, the goal of the philosophical tradition embedded in our culture is to eliminate uncertainty: moral, intellectual, and practical. Indeed, the demand that knowledge be expressed in terms of rules or definitions which can be applied without the risk of interpretation is already present in Plato, as is the belief in simple elements to which the rules apply[149].

Pour Kant, certains aspects de la cognition liées à la connaissance seraient régies par des règles; pour lui, la perception et l'expérience, par exemple, seraient le résultat d'un ensemble d'instructions ou règles: "Kant explicitly analyzed all experience, even perception, in terms of rules, and the notion that knowledge involves a set of explicit instructions is even older"[150]. Dreyfus affirme qu'avant Kant, et bien avant les chercheurs en IA, plusieurs philosophes croyaient déjà que la pensée pourrait être saisie par un ensemble de règles:

> Since the Greeks invented logic and geometry, the idea that all reasoning might be reduced to some kind of calculation—so that all arguments could be settled once and for all—has fascinated most of the Western tradition's rigorous thinkers. Socrates was the first to give voice to this vision. The story of artificial intelligence might well begin around 450 B. C. when (according to Plato) Socrates demands of Euthyphro, a fellow Athenian who, in the name of piety, is about to turn in his own father for murder: "I want to know what is characteristic of piety which makes all actions pious... that I may have it to turn to, and to use as a standard whereby to judge your actions and those of other men." Socrates is

149 H. L. Dreyfus, *op. cit.*, p. 211.
150 H. L. Dreyfus, *op. cit.*, p. 176.

> asking Euthyphro for what modern computer theorists would call an "effective procedure," "a set of rules which tells us, from moment to moment, precisely how to behave"[151].

Pour Platon, affirme Dreyfus, l'esprit serait, "pré-programmé" dans une vie antérieure. Cela ne serait pas tout à fait étranger à l'idée contemporaine de nos chercheurs en IA selon laquelle des processus cognitifs peuvent être assimilés à des processus informatiques (*computational processes*):

> This is a decisive issue for the history of our concepts of understanding and explanation. Plato leaves no doubt about his view: any action which is in fact sensible, i.e. , non arbitrary, has a rational structure which can be expressed in terms of some theory and any person taking such action will be following, at least implicitly, this very theory taken as a set of rules. For Plato, these instructions are already in the mind, preprogrammed in a previous life, and can be made explicit by asking the subjects the appropriate questions. Thus, for Plato, a theory of human behavior which allows us to *understand what* a certain segment of that behavior accomplishes is also an *explanation* of *how* that behavior is produced. Given this notion of understanding and this identification of understanding and explanation, one is bound to arrive at the cognitive simulationists with their assumption that it is self-evident that a complete description of behavior is a precise set of instructions for a digital computer, and that these rules can actually be used to program computers to *produce* the behavior in question[152].

Pour résoudre des problèmes en physique, en logique ou en IA, il ne faut pas seulement les représenter de façon très exacte; il est également nécessaire d'organiser systématiquement ces représentations en termes de règles pour que la formulation des problèmes et leurs solutions soient accessibles à tout instant. Des savants comme Descartes et Galilée ont toujours compris qu'il faut avoir une cohérence entre les représentations d'un problème et les règles capables de rendre compte des solutions possibles à ce problème:

> (...) What makes a notation "suitable" for symbolizing some subject matter? It isn't just that convenient symbols can be invented for all the relevant items or variables—that, by itself, is trivial. Galileo didn't just say "the area within the triangle represents the distance traveled" and let it go at that. Nor is analytic geometry merely the clever idea of identifying geometric points with numerical coordinates (that wouldn't have cost Descartes much effort). No, what earned these men their reputations was demonstrating how, if you represented things in certain very specific ways, you could *solve problems*. (...) solving problems obviously involves more than just representing them. There must also be various allowable steps that one can take in getting from (the representation of) the problem to (a representation of) the solutions (for example, in derivations,

151 H. L. Dreyfus, *op. cit.* , p.67.
152 H. L. Dreyfus, *op. cit.* , pp. 176-177.

> proofs, etc.). Hence the integrated system must include not only notational conventions, but also *rules* specifying which steps are allowed and which are not[153].

Les travaux en IA en général considèrent que les règles et les représentations jouent un rôle fondamental dans la production et la compréhension des comportements intelligents.

La notion de règle est liée à la notion de pensée en tant que mécanisme de calcul. En IA, comme dans la conception moderne, les règles relèvent d'une façon organisée de voir le monde et de résoudre des problèmes. Les conduites intelligentes, de ce point de vue, doivent, par hypothèse, être régies par des règles.

> We can begin to reveal the rationalistic tradition by considering the question "what do you do when faced with some problem whose solution you care about?" The rationalistic orientation can be depicted in a series of steps:
> 1. Characterize the situation in terms of identifiable objects with well-defined properties.
> 2. Find general rules that apply to situation in terms of those objects and properties.
> 3. Apply the rules logically to the situation of concern, drawing conclusions about what should be done[154].

Dans les sciences cognitives ou dans une recherche en IA sur la résolution de problèmes, le comportement intelligent est représenté à partir d'un ensemble d'instructions (règles) ou programmes qui permettent une description complète et même l'explication d'un comportement intelligent. Mendonça résume très bien cette perspective:

> (...) le comportement intelligent est le résultat d'un calcul exécuté d'après des règles et des principes rigides. La recherche sur l'intelligence artificielle s'efforce de découvrir ces règles. (Il faut que ce soit des règles *rigides* qui constituent une *procédure effective*, car seulement de telles règles peuvent être « suivies » par une machine.) En adoptant l'hypothèse que les « processus de pensée » sont des processus d'élaboration de l'information ou des procédures informatiques, la *simulation* dans un ordinateur d'une faculté intelligente est en même temps son *explication*. (...) Cette conception aboutit à ce que des processus cognitifs (cela veut dire dans notre cas: comprendre, penser, percevoir etc.) soient conçus comme des opérations formelles (« syntaxiques ») avec des « représentations internes »[155].

Dans la tradition représentationnaliste, les règles, des prescriptions de caractère pratique ou épistémologique, ont été toujours en rapport avec la méthodologie scientifique

---

153 J. Haugeland, *op. cit.* , p. 34-3. Italiques de l'auteur.
154 T. Winograd et F. Flores, *op. cit.* pp. 14-15.
155 W. P. Mendonça, *op cit.* , pp. 7-8.

servant à établir le chemin à prendre pour l'obtention de bons résultats ou pour orienter la réflexion dans un domaine théorique donné. Pensons par exemple à la notion de règle opératoire, qui sont des structures logiques dans une théorie, qui établissent la dérivation de théorèmes à partir d'une série d'axiomes. Des règles, comme les programmes d'ordinateur, possèdent un caractère syntaxique qui établissent les opérations possibles dans un système formel capable de reprèsenter un système physique.

L'importance des règles en IA a été soulignée, par exemple, par Newel et Simon. Newel et Simon ont proposé que tout système permettant de représenter les connaissances, a besoin d'être programmé en termes de règles[156]. Les règles en IA sont en rapport avec la notion de représentation, plus précisément avec la notion de représentation symbolique, procédure méthodologique en l'IA qui consiste à utiliser un ensemble structuré d'éléments symboliques qui sont organisés de façon binaire afin de représenter les caractéristiques des objets.

Dans les programmes informatiques basés sur des règles, celles-ci, sont utilisées pour représenter des relations entre les éléments symboliques du système. Les règles permettent aux systèmes en l'IA de manipuler les représentations symboliques pour, par exemple, faire des inférences, appliquer d'autres règles, inférer d'autres faits à partir de ce qui a été représenté premièrement. Les règles et les représentations symboliques lorsqu'elles sont efficaces, permettent aux systèmes de faire même des prédictions et proposer des solutions sur une situation ou problème donné. Certains de ces tâches sont assimilées, par les chercheurs en IA, aux capacités cognitives humaines.

2- Le calcul et la conception mécanique de l'esprit

L'affirmation qu'il est théoriquement possible d'avoir des machines intelligentes exige une position philosophique sur la question de la nature de la pensée en tant que réprésentation. La recherche sur les machines intelligentes est née à l'intérieur de la tradition rationaliste et empiriste sur laquelle reposent les principales bases de notre pensée scientifique, elle s'est développée à l'intérieur de cette tradition qui est basé sur la capacité formelle des représentations, et est en rapport avec les travaux sur le dévéloppement de la puissance du calcul mécanique et de la mécanisation de la pensée:

---

156 Cf. Newel et Simon "Computer Science as Empirical Enquiry: Symbols and Search", dans *Communications of the ACM*, Vol. 19, nº 3, mars 1976, p. 116. Selon Newel et Simon, interessés à ce moment aux règles heuristiques, les règles appliquées dans les programmes de l'IA devraient être du type si- alors : si une telle situation ou événement se produit alors, telle et telle action est declanchée par le système.

> During the seventeenth century, the idea arose of converting thought into a formal notation and using a calculating device to carry out the reasoning operation. In 1650, the English philosopher Thomas Hobbes proposed the idea that thinking is a rule based computational process, analogous to arithmetic. Gottfried Wilhelm Leibnitz (1646—1716) describes his book *De Arte Combinatoria* (1666) as containing " a general method in which all truths would be reduced to a kind of calculation." Much Later, in 1854, George Boole published *An Investigation of the Laws of Thought on which are Fonded the Mathematical Theories of logic And Probabilites*. (...)
>
> The dream of devising a formal system that could be a basis for all reasoning seemed to be almost at hand with the publication of Russell and Whitehead's *Principia Mathematica* (1910 —1913) The codification of logic and the reduction of significant portions of mathematics to the language of logic appeared to provide the means by which people (or machines could do mathematics without having to understand what was actually hapening; It would be sufficient to manipulate the symbols according to permissible logical transformations[157].

La conception mécanique de l'esprit a comme base le succès de la physique moderne et l'introduction des mathématiques comme moyen de représentation; l'idée d'une pensée mécanique apparaît avec T. Hobbes, passant par les idées de Leibniz et George Boole et elle continue a être développée jusqu'à Turing et au sein de la recherche sur les machines intelligentes:

L'appel à des règles est en rapport avec la conception mécaniste et physique selon laquelle l'esprit est un mécanisme de calcul. Le mécanisme, ou l'idée de la pensée mécanique, est défendu à l'heure actuelle par des thèses beaucoup plus sophistiquées que celles des philosophes modernes Hobbes et Leibniz.

Avec le développement du calcul mécanique, l'hypothèse mécaniste selon laquelle nos raisonnements sont une espèce de calcul suscitera plus tard l'idée que certains calculs produits par l'esprit humain peuvent être reproduit par d'autres systèmes formels mécaniques tels que les ordinateurs[158].

Turing a suggéré qu'une machine peut, dès que convenablement programmée, présenter les mêmes performances que l'esprit. Turing fait partie à la fois de l'histoire du calcul et de l'histoire de l'IA. Ses travaux pendant la Seconde Guerre mondiale ont été d'une certaine façon théoriquement importants pour la recherche sur les premiers ordinateurs

---

157 M. A. Fischler and O. Firschein, *Intelligence: The Eye, the Brain, and the Computer*, Addison-Wesley Publishing Co, Mass. ,1987, p.15.

158 Cf. J. Haugeland, *Artificial Intelligence: The Very Idea*, Bradford Book, MIT Press,1985. Nous nous inspirons du chapitre 1 de cet ouvrage pp. 14-45.

électroniques[159] A l'époque de Turing, il existait une croyance générale que tout type de comportement guidé par des règles pourrait être étudié mathématiquement; cette idée d'inspiration mécaniste était adoptée en Cybernetique et par le Behaviorisme. Le comportement intelligent, conçu comme comportement guidé par des règles devra être compris et formalisé par des procédures effectives calculables ou pouvant être traitées par un ordinateur.

Les discussions mathématiques menées par Turing dans les années 1930 sont importantes pour la prise de conscience du potentiel réel du calcul mécanique. L'évolution de ses réflexions sur ce problème, ajoutée aux perspectives ouvertes par la cybernétique dans les années 1940 et 1950 vont anticiper les premières spéculations, après Descartes, sur les possibilités de l'intelligence des machines.

Nous pouvons dire que les travaux de Turing de 1950 et l'autre "Intelligent Machinery "[160] de 1947, suivis des recherches de McCulloch - W. Pitts et de Frank Rosenblat constituent les premièrs travaux qui portent déjà sur l'Intelligence Artificielle. On voit dans les travaux de ces chercheurs deux orientations de la recherche en IA: les approches ascendante et descendante que nous avons mentionnées.

Turing a démontré plusieurs théorèmes généraux sur les capacités logiques de sa machine. C'est l'une des premières fois que quelqu'un montre, après C. Babbage, et Ada Lovelace, l'importance et les fondements d'une machine programmable. Un autre aspect important des travaux de Turing, c'est son caractère anticipatoire. Dans les années 1940 et 1950 Turing discute, par le biais de discussions logico-mathématiques, des capacités et limitations du calcul mécanique à produire des résultats proches de ceux exigeant de l'intelligence humaine.

---

159 L'idée de programme stocké en mémoire de Von Neumann, est inspirée de la notion de *Machine Universelle de Turing*. Les ordinateurs peuvent être considérés comme étant la réalisation majeur de l'histoire du calcul mécanique. Ils sont des machines universelles de la même façon que les machines universelles de Turing. Cependant ils ne sont pas comme ces dernières des machines abstraites. Le fait que les ordinateurs digitaux représentent la réalisation concrète de ces machines abstraites, nous permet de transformer des théories abstraites sur l'esprit et sur le langage en programmes pour ordinateur et tester ces théories. Les propriétés des machines réelles dérivent en partie des théories abstraites: les machines universelles de Turing. Comme celles là, l'ordinateur digital est capable d'exécuter toute sorte de programmes et prendre pour modèle n'importe quel autre machine à états discrets. En tant qu'équivalents aux machines universelles de Turing les ordinateurs digitaux sont importants pour l'IA aussi bien du point de vue théorique que du point de vue pratique. Il n'est pas nécessaire d'inventer d'autres machines pour exécuter des procédures nouvelles de calcul. Une seule machine Universelle ou ordinateur digital (dès que convenablement programmé) est capable d'exécuter n'importe quelle procédure de calcul ou programme.

160 A. Turing, "Intelligent Machinery", report to the National Physical Laboratory, 1947. Cf. J.G. Ganascia, *op. cit.* , p. 20. Dans ce travail nous trouvons un très bon exposé sur les articles de Turing mentionnés.

2.1- Le calcul mécanique, mécanisme et anti-mécanisme

Avant les discussions suscitées parTuring sur les possibilités du calcul mécanique, des mathématiciens comme Kurt Gödel (1931) et Alonzo Church (1936) avaient déjà publié des articles importants sur les limitations et capacités des systèmes formels. Le premier a établi l'indécidabilité de certaines propositions de la théorie élémentaire de nombres le second a montré l'insolubilité (*insolvability*) de certains problèmes de cette même théorie.

Au début du XXème siècle on croyait que la mécanisation du calcul était possible. Pour le cercle de mathématiciens inspirés des projets de Hilbert, il était possible de démontrer l'absence de contradiction de la mathématique. Selon les écoles mathématiques non intuitionnistes, étant donné un ensemble d'axiomes et les règles de déduction, on pouvait créer de nouvelles mathématiques non contradictoires. Pour arriver à cela on devrait 1) aboutir à une formalisation complète de la mathématique au moyens de symboles[161] (sans contenu), capables d'être manipulés mécaniquement. 2) Démontrer que l'utilisation des règles de déduction ne conduirait jamais à une contradiction formelle de la mathématique[162].

Selon ce courante mécaniste, les mathématiques devraient être consistantes et complètes. Deux théoremes produits à l'intérieur de tels mathématiques ne pourraient jamais se contredire. Toutefois, le mathématicien Kurt Gödel met en question un tel projet mécaniste:

> This expectation was destroyed in 1931 by Kurt Gödel who showed that there are true statements in mathematics that a consistent formal system will not produce, i.e. , that it is impossible to alter the foundations of mathematics to exclude unprovable propositions. Gödel showed how to produce a true statement, S that could not be proved by a consistent system, F, using a set of axioms and a proof procedure. He did this by showing that if S could be proved, then a contradiction would arise. F is therefore "incomplete" since it does not produce all true statements[163].

---

161 Toutes les opérations sur des symboles pouraient être vérifiées mécaniquement sans risque d'erreur, car elles seraien trés precises.

162 Pour une description precise du programme de D. Hilbert pour le fondament definitif de la mathématique et sur Gödel,voir J. Ladrière, *Les limitations internes des formalismes*, Gautier-Villars, Paris, 1957, 715p.

163 M. A. Fischler, et O. Firschein, *Intelligence: The Eye, the Brain, and the Computer*, Mass. , Addison-Wesley Publishing Co, Mass. , 1987, p.43.

Gödel affirme que si l'arithmétique est non contradictoire, elle ne peut être complète. L'incomplétude de l'arithmétique montre, selon Gödel, qu'en tant que système formel, elle a des formules qui ne sont ni démontrables ni réfutables.

Pour Gödel, l'arithmétique est incomplète. Son théorème d'incomplètude dit que pour tout système formel qui contient l'arithmétique, ce système a des propriétés qui ne peuvent être démontrées en n'utilisant que des règles internes au système. Ce système est indécidable, car on ne peut décider ce qui est vrai et ce qui est faux en restant à l'intérieur du système. Les théorèmes du système en question ne peuvent pas être démontrés avec les seuls moyens disponibles dans le système. Voyons comment ce problème se présente de façon intuitive:

> Le théorème de Gödel affirme que dans tout système cohérent assez performant pour produire de l'arithmétique simple, il y a des formules qui ne peuvent pas être prouvées-dans-le système, mais que nous pouvons poser comme vraies. Nous considérons essentiellement la formule qui dit, en effet, «Cette formule est improuvable-dans-le-système». Si cette formule était prouvable-dans-le-système, nous aurions une contradiction: car si on pouvait lui administrer une preuve à l'intérieur du système, alors elle ne serait pas indémonstrable-dans-le-système; ainsi « Cette formule est improuvable-dans-le système » serait faux; de même si elle était prouvable-dans-le-système, alors elle ne serait pas fausse, mais vraie, puisque dans tout système logique rien de faux ne peut être prouvé-dans-le-système, si ce n'est des vérités. Ainsi la formule « Cette formule est improuvable-dans-le-système» n'est pas prouvable-dans-le système, mais improuvable-dans-le-système. De plus, si l'énoncé «Cette formule est improuvable-dans-le-système» est lui-même indémontrable à l'intérieur de ce dernier, alors il est vrai que cette formule est improuvable-dans-le système, c'est-à-dire que l'énoncé précédent est vrai[164].

Ce que dit la citation en haut peut être exemplifié autrement, lors de l'analyse des paradoxes sémantiques, tels que le paradoxe du menteur. Prenons l'énoncé ci-dessous.

Cet énoncé est un mensonge

Appelons l'énoncé ci-dessus de E. On conclut à propos de E que, si E est vrai, E est faux, et que si E est faux, alors E est vrai. Cette aspect indécidable de l'énoncé E est analysée dans la perspective de Gödel de la façon suivante:

E: *Cet énoncé n'est pas démontrable*

Si E est démontré, E n'est pas vrai, le système formel vient de produire une fausseté. Mais si E n'est pas démontré, alors nous avons un énoncé qui est vrai, mais qui n'est pas démontré dans le système et ainsi le système formel devient incomplet.

---

164 A. R. Anderson, ***Pensée et machine***, Ed. du Champ Vallon, 1983, pp. 81-82. Traduit de l'americain par Patrice Blanchard, *Minds and Machines*, Prentice-Hall Inc. ,1964.

L'énoncé ci-dessus permet d'illustrer un aspect important du théorème de Gödel, à savoir que, tout système formel ayant la prétention d'affirmer quelque chose sur le plan logique est confronté fatalement à deux problèmes, l'incomplétude et l'indécidabilité, lesquels peuvent nous amener soit à la démonstration de la validité des choses fausses, soit à l'impossibilité de démontrer certaines choses vraies à l'intérieur d'un système formel.

Les conclusions de Gödel sur l'incomplétude de l'arithmétique eurent une répercussion théorique importante sur les thèses en IA et surtout sur les modèles computationnel de l'esprit.

Les travaux de Godel sur l'incomplétude de l'arithmétique mettent en relief les limites des systèmes formels. Ses résultats rigoureux dans le domaine de la logique mathématique sont utilisés pour montrer que l'esprit humain n'est pas un mécanisme de calcul. Les limitations des systèmes formels découvertes par Gödel inspirent plusieurs auteurs, comme J.R. Lucas qui emploient ses démonstrations mathematiques pour critiquer l'IA par la voie formelle montrant que (1)aucune machine ou système computationnel ne peut servir de modèle adéquat de l'esprit humain et que (2) l'esprit humain est fondamentalement distinct des opérations formelles de tout système déterministe. Les critiques basées sur ces deux arguments ont déclenché plusieurs discussions à l'intérieur de l'IA et de la philosophie de l'esprit[165].

L'indécidabilité de certaines propositions et l'insolubilité de certains problèmes de la théorie élémentaire des nombres analysée par Gödel et Church ont probablement influencé Turing qui a fait lui aussi des analyses sur l'indécidabilité analogues aux travaux de ces deux derniers auteurs en même temps qu'il discutait,à partir de l'idée de "machine de Turing"[166], les capacités du calcul mécanique.

Avec ses discussions sur le calcul mécanique, Turing veut comprendre ce qu'est une procédure mécanique de calcul (*effective procédure)*. [167] et ses limites. Autrement dit, quel

---

165 Cf. J. R. Lucas, "Minds Machines and Gödel", *in Philosophy* nº 36, 1961, 112-127. Ce travail a été publié aussi dans Anderson, A. R. , *Minds and Machines*, Prentice-Hall Inc. ,1964. qui rassemble plusieurs articles importants sur les discussions à propos des possibilités formelles des machines pour la duplications des capacités de l'esprit.

166 La machine de Turing est une machine abstraite à états finis, cest-à-dire, elle est sujette à un nombre fini de règles qui font que à chaque pas d'un calcul cette machine utilise une quantité finie de mémoire. (cependant sa capacité de mémoire est potentiellement infininie) La machine de Turing constitue une base théorique importante pour la compréhension du traitement de l'information par nos ordinateurs actuels. Elle a été très importante pour les premiers pas de l'informatique car elle a permis de faire de nombreuses conjectures sur les limites des procédures effectives de calcul et sur la capacité formelle des machines réelles: "In the Turing machine the reduction of a process to elementary operations is carried to its limit. Even a simple operation such as addition is broken down into a chain of far simpler operations. This increases the number of steps in the computations carried out by the machine, but simplifies the logical structure for theoretical investigations." (M.A. Fischler et O. Firschein,op. cit. , p. 42.).

167 Une procedure effective est un ensemble de règles formelles qui permet à un système formel d' executer pas à pas certaines opérations. Les programmes algorithimiques, par exemple, constituent un bon exemple de ce qu'est une procédure effective.

est le statut d'un algorithme[168] pour le calcul mécanique. Un de ses objectifs est d'établir une notion intuitive de calculabilité[169].

La notion de "machine de Turing" et la notion intuitive de calculabilité[170] constituent des moyens théoriques qui permettent à Turing d'arriver à des résultats intuitifs d'indécidabilité analogues à ceux de Godel. De tels résultats peuvent être exprimés par la thèse selon laquelle nous ne sommes pas en mesure de savoir, parmi les calculs ou algorithmes qui dépassent les capacités finies de toute machine de Turing, quels calculs ou algorithmes sont susceptibles d'amener à un résultat [171]. Selon Turing, aucune procédure effective ne peut permettre de décider si un calcul peut ou non amener à un résultat dans un nombre fini d'étapes. Nous ne pouvons pas décider par des moyens algorithmiques quels calculs peuvent finir par l'obtention d'un résultat et lesquels peuvent continuer indéfiniment sans parvenir à une résolution.

Les analyses de Turing montrent qu'une machine qui opère formellement sur des règles logiques a des limites d'ordre théorique; il y a certaines tâches formelles qu'aucune machine de Turing ou ordinateur ne peut résoudre.

Les résultats négatifs obtenus par Turing à partir de la notion de "machine de Turing," sert à discuter les limitations du calcul mécanique, mais Turing met l'accent plutôt sur les capacités des algorithmes que sur les limitations du calcul mécanique. Si d'un côté Turing emploie les machines de Turing pour discuter les limitations formelles des algorithmes, de l'autre côté, avec le modèle mécaniste de la machine universel de Turing, il montre le pouvoir des programmes à simuler toute autre système formel. Le modèle de la machine de Turing est utilisé par les théories fonctionnalistes qui supposent que l'esprit résulte d'une manipulation complexe de symboles. Ce modèle mécaniste qui associe la pensée à un système formel capable d'opérer des manipulations de symboles est bien exprimé dans le passage qui suit:

> Suppose that the highest level brain processes to which human conscious and unconscious thoughts/symbol manipulations are reducible are algorithmic and

---

168 Un algorithme est par définition une suite finie d'ordre de transformations à éffectuer sur des symboles discrets. Autrement dit, il est une suite finie d'pérations élémentaires effectués par une machine rèel ou abstraite sur des symboles. En tant que shéma de calcul, l'agorithme doit (selon la notion intitive de calculabilité) permetre d'arriver à un résultat.

169 Cette notion est définie en termes de fonction calculable, c'est-à-dire, une fonction qui peut être évaluée au des moyen d'algorithimes finis.

170 Selon la notion intuitive de calculabilité tout algorithme doit amener à un résultat. Le problème de la notion intuitive de calculabilité, comme Turing a bien remarqué, est que nous ne pouvons pas décider pour les algorithmes quei execedent nous ressources de mémoire et de temps lesquels peuvent néamoins mener a une solution en un nombre finnie d'étapes. Nous ne pouvons pas non plus garantir qu'un tel algorithme est precis et complèt.

171 Même si la machine de Turing est définie par un nombre fini de règles et utilise des quantitées de mémoire finites elle a une capacité de mémoire ilimitée et est comme nous avons déjà remarqué potentiellement infinie.

> that these brain processes are really produced by precise, finite, deterministic recipes somehow "wet-vired" into the human brain. Then human cognition is simulatable by Turing machines. Hence, any limitative results about Turing-machine computations apply to humans too, and perfect computer modeling of human cognition is, in principle, possible. This form of mechanism is a principal assumption of modern cognitive science and implies that *artificial* intelligence can in principle, do anything *natural* intelligence can[172].

Pour Turing les capacités des machines à reproduire des comportements intelligents sont dûes au potentiel du calcul mécanique. Il met l'accent (avant même la naissance de l'informatique) sur le logiciel, *software* par rapport au matériel, *hardware*[173], comme la base à la production des sorties (outputs) semblables à celles produites par un cerveau humain.

L'idée de Turing selon laquelle les capacités d'une machine à reproduire les comportements intelligents dépendent de ses propriétés formelles, plutôt que de ses propriétés en tant que système matériel, aura une influence importante sur les théories de l'esprit de caractère fonctionnaliste. Les discussions de Turing sur le calcul mécanique conduisent à la conclusion fonctionnaliste que tout système formel (cerveau ou ordinateur) serait une instantiation d'une machine de Turing. De la même façon que les fonctionnalistes, ce mathématicien comprend que ce n'est pas le caractère fini des machines qui limite leur potentiel ni la quantité de machinerie.

Nous allons voir dans la section qui suit comment les théories sur l'esprit inspirées des théories sur le calcul mécanique de Turing se développent jusqu'à l'idée (quelquesfois fondamentale pour les thèses en IA) selon laquelle les états mentaux sont considérés comme des fonctions abstraites indépendantes des caractéristiques physiques du système qui les réalisent.

3- Les théories sur l'esprit et l'Intelligence Artificielle

Les théories sur l'esprit représentent un autre point important qui nous aide à comprendre les rapports entre les préoccupations des philosophes et de ceux qui travaillent dans le domaine de l'IA. Les discussions sur le rapport entre le corps et l'esprit font partie

---

172 J. Case, *in* S.C. Shapiro et D. Eckroth, *Encyclopedia of arificial Intelligence*, J. Wiley et sons, Inc. USA. 1987. p.1125.

173 Cf. A.Turing, "Computing Machinery and Intelligence" *Mind*, 59, 1959, p.343-469. (Reédition par R. Anderson, *Minds and Machines*, Englewood Cliffs, Prentice Hall N. J. ,1964, pp. 4-30) Dans cet article, Turing renforce l'idée selon laquelle les caractéristiques formelles du système sont fondamentales pour qu'il arriver à avoir un comportement considéré comme intelligent.

des problèmes traditionnels de la philosophie. L'IA fait partie de tradition scientifique et philosophique et elle ne peut pas bâtir son projet de conception de machines intelligentes sans avoir touché à ce problème. Nous allons donner un bref aperçu de la question afin de montrer que certaines théories philosophiques sur l'esprit sont en rapport avec certaines conceptions importantes en IA.

Nous allons parcourir maintenant plusieurs théories sur l'esprit afin de montrer que certaines d'entre elles sont liées à la tradition représentationnaliste et sont à la base de certaines conceptions en IA. Pour montrer cela nous allons voir, dans la tradition philosophique comment des théories sur le rapport entre le corps et l'esprit se développent jusqu'à l'apparition d'une théorie computationnelle de l'esprit, le fonctionalisme qui constitue actuellement le rapport le plus étroit entre la philosophie et l'IA.

Avec Galilée et Descartes s'initie tout un mouvement intellectuel basé sur le calcul qui donnera complète autonomie aux sciences de la nature; l'univers et le mouvement en viendront à être expliqués par des moyens formels à la disposition de l'homme moderne. Il est important de remarquer que ces faits ont eu une influence importante sur les modèles de réprésentation de la nature et de la pensée par la suite. Cela a été fondamental, comme l'affirme Haugeland, pour la conception de "esprit" en Occident[174].

La tradition philosophique représentationnaliste ne fournit pas une conception univoque sur l'esprit; quelques philosophes identifient l'esprit à une substance immatérielle, d'autres préfèrent dire que l'esprit n'est qu'un élément du monde physique comme un livre ou une chaise. Il y a toutefois deux grandes voies théoriques sur l'esprit qui constituent les orientations suivies par plusieurs philosophes. Il s'agit des théories de caractère dualistes et des théories de caractère matérialiste.

3.1- Le Dualisme et l'Intelligence Artificielle

Les théories dualistes

Le dualisme est une philosophie de l'esprit qui comprend celui-ci comme une substance immatérielle. Les théories dualistes dérivent en général de la philosophie de Descartes, laquelle est axée sur l'idée qu'on peut représenter presque tout de façon mathématique. Selon Descartes, en utilisant des moyens formels nous pouvons avoir une

174 J. Haugeland, "The Saga of the Modern Mind", *op.cit.*, pp.15-45.

compréhension de la nature assez objective. Cependant, le philosophe français lie encore la science et la religion, la métaphysique et la physique. Il attribue pour cette raison un statut religieux à l'esprit et considére la matière comme quelque chose de complètement différent de l'âme.

Pour Descartes, l'esprit est par nature constitué de pensées, croyances, désirs, etc. , lesquels seraient soumis aux règles de la raison. L'univers physique, par contre, est constitué de corps, d'objets et de mécanismes physiques régis par des lois de la nature. Le dualisme cartésien a créé une opposition conceptuelle entre l'esprit et le corps. Descartes croit cependant que le corps et l'esprit entretiennent un rapport ontologique, qu'ils sont unis par une interaction..

Pour Descartes les rapports entre le corps et l'esprit sont interactifs à la manière d'un système causal bien réglé où les événements mentaux interagissent avec les événements physiques, et vice versa. Nous constatons immédiatement qu'il est difficile de soutenir un dualisme de ce type et d'admettre en même temps l'existence d'interactions causales entre l'esprit et le corps. La difficulté serait d'expliquer comment le non-étendue peut demeurer dans un corps étendu et y agir de façon causale. Nous ne pouvons pas décrire l'esprit en termes physiques, car il ne possède pas de masse, ni une taille ou une forme, pas plus que nous ne pouvons décrire le corps en termes d'événements mentaux. Enfin, le dualisme ne peut pas rendre compte de l'interaction causale entre l'esprit et le corps, car il n'explique pas de façon plausible comment l'esprit, étant immatériel, peut exercer une causalité physique sur le corps sans violer les lois de conservation de la matière et de la quantité de mouvement.

Le réprésentationnalisme cartésien est à la base de l'IA. Cependant, le dualisme de Descartes est incompatible avec certaines thèses de l'IA. Pour Descartes, les machines étaient dépourvues de raisonnement et d'esprit. Il dit que même si une machine réussissait à dire des mots et à manipuler des symboles répondant à des questions, cela ne prouverait jamais qu'elles est douée d'un esprit car elle n'aurait pas vraiment la maîtrise du langage dans son acception humaine[175]. Le dualisme d'une manière générale ne fournit pas une explication correcte sur les rapports causaux entre l'esprit et la matière. En tant que philosophie de l'esprit, appliquée aux études cognitives et à l'IA, les différents types de dualismes se révèlent inadéquats, car il demeurent incompatibles avec les méthodes

175 Il y a entre la pensée de Descartes et l'IA un point en commun. La pensée pour celui-là n'est pas une simple manipulation de symboles, mais une manipulation rationnelle de symboles matérialisée en forme de langage. Un programme d'ordinateur est une procédure formelle qui permet une manipulation de symboles et il est écrit en langage de programmation (haut niveau) qui devient de plus en plus proche du langage naturel.

scientifiques et les théories employées dans ces domaines. En résumé, l'IA est incompatible avec toute théorie de l'esprit dualiste.

3.2- Les théories matérialistes sur l'esprit et l'Intelligence Artificielle

Le behaviorisme

Le behaviorisme est un courant matérialiste en psychologie, qui a essayé de faire de la psychologie une "science de la nature". Pour J. Watson, (1878-1958) les états mentaux ne constituent pas des données observables objectifs; ils sont considérés plutôt comme une conséquence fortuite ou une espèce du sous-produit de la matière.

Le behaviorisme strict des premiers psychologues du comportement, prend une position nettement radicale sur l'esprit finissant par éliminer des théories psychologiques tout appel à la causalité mentale. D'une manière générale, le défaut de cette perspective béhavioriste radicale consiste à réduire l'esprit, ou mieux, la pensée, à une analyse en termes de conditionnement et d'apprentissage, en essayant de l'expliquer toujours en termes de stimuli de l'environnement et des réponses à de tels stimuli en termes de comportement observable.(S-R):

> Le béhavioriste radical prétend que mieux les psychologues comprendront les relations entre stimulus et réponses, plus vite ils abandonneront toute idée de causalité mentale pour expliquer le comportement[176].

Les expériences en physiologie, en psychiatrie et en psychologie, entre autres, mettent en jeux les positions matérialistes qui nient le pouvoir causal de l'esprit. Les rapports entre le corps et l'esprit en viennent à être reconsidérés d'un point de vue scientifique. Dans plusieurs travaux, les chercheurs montrent que les événements mentaux jouent un rôle important sur l'organisme. Par exemple, on a montré que, en général, des états mentaux comme des peurs, craintes, haines, angoisses ont des effets tels que des ulcères ou d'autres troubles physiques qui affectent notre fonctionnement physiologique.

> En particulier, dans le cas du comportement humain, les théories psychologiques conformes aux principes méthodologiques du béhaviorisme

176 J. Fodor, "Le corps et l'esprit", *Pour la science*, mai, 1981, n° 43, p. 79.

> radical se sont révélées très souvent stériles; on pouvait s'y attendre si les processus mentaux sont véritablement présents et causalement efficaces[177].

Le béhaviorisme strict s'écroule en tant que théorie de l'esprit dans un moment où les discussions sur le langage menées en philosophie et en psycho-linguistique, inspirées des travaux de Chomsky, proposent des nouvelles approches sur la question de l'esprit. Tout cela se passe, dans les milieux universitaires de la fin des années 1950 et 1960, à une époque où les discussions sur le langage et sur la théorie de l'information sont très répandues et coïncident également avec le développement de l'informatique et l'apparition de l'IA.

Des théories béhavioristes assouplies telles que, le béhaviorisme logique ainsi que la théorie de la réduction neurophysiologique appelée "théorie de l'identité" sont alors proposées pour faire face aux difficultés et limitations théoriques du béhaviorisme.

Le behaviorisme logique

Le béhaviorisme logique s'intéresse au caractère sémantique des représentations mentales. Ce courant caractériserait, par exemple, des conceptions sémantiques comme celles de G. Ryle (1900-1976). Ce dernier analyse les événements mentaux en termes d'énoncés d'observation.

Le point de vue philosophique de Ryle sur l'esprit abolit tout appel à des antécédents mentaux comme base de la description du comportement. Pour le béhaviorisme logique, il ne faut pas non plus croire à la causalité du mental. Les propriétés mentales sont définies par des relations logiques interprétées. J. Fodor montre dans le passage qui suit, comment ce courant philosophique analyse la signification des représentations mentales:

> L'idée fondamentale est qu'attribuer un état mental (par exemple la soif) à un organisme revient à dire que «l'organisme est disposé à se comporter d'une certaine façon (par exemple à boire s'il a de l'eau à sa disposition)». De ce point de vue, toute proposition mentale a la même signification qu'un énoncé du type «si... alors...» (que l'on nomme relation conditionnelle du comportement) et qui exprime une disposition à agir. Par exemple, « Durand a soif » est équivalent à l'énoncé conditionnel «s'il avait de l'eau à sa disposition, alors Durand en boirait». Par définition, une relation conditionnelle du comportement n'inclut aucune représentation mentale. La clause «si» de l'énonce de la relation ne se réfère qu'au stimulus et la clause «alors» ne concerne que la réponse comportementale. Stimuli et réponses étant des manifestations physiques(...)[178].

---

177 *Idem.*
178 *Idem*, pp. 79-80.

Pour le béhaviorisme logique les faits mentaux sont réduits à des actions explicitées par des énoncés du langage naturel qui décrivent le corps d'un sujet, ces états physiques, ses dispositions à agir, aussi bien que les événements qui affectent son corps ou ce qui se passe à l'intérieur de celui-ci. On n'a pas besoin, ainsi, des énoncés mentalistes pour parler des événements mentaux. Ryle analyse les énoncés mentalistes en termes dispositionnels.

Il adopte une position proche du béhaviorisme radical, dans le sens où ces deux approches ne confèrent aucune importance à la causalité mentale[179]. Selon Ryle, l'opposition corps-esprit n'est qu'un pseudo-problème dû à un mauvais usage du langage naturel. Nous ne nous apercevons pas que nous décrivons les événements mentaux et les événements physiques en appliquant à ces événements des mots comme s'ils appartenaient aux mêmes catégories logiques. Cela constitue une erreur de catégorie, car les termes appliqués aux faits mentaux sont d'une espèce tout à fait différente de ceux appliqués aux objets physiques[180].

Enfin pour Ryle les événements mentaux n'ont pas d'existence réelle, ils ne sont qu'une façon de parler. Le béhaviorisme logique ne semble pas proposer de solutions sur les rapports entre le corps et l'esprit qui soient compatibles avec les recherches de l'IA.

Le monisme

Il y a, à l'intérieur de la philosophie de l'esprit, des théories qui mettent en doute les distinctions entre le physique et le mental. Ces théories affirment que tout ce qu'il y a est composé d'une seule substance;nous parlons des monismes. Pour le moniste l'opposition corps-esprit n'existe pas, car il n'y a qu'une seule substance. Les monistes idéalistes [181] disent que cette substance est l'esprit ou les idées; pour les monistes matérialistes, la réalité ultime est la matière[182].

---

179 *Idem*, p.81.

180 Cf. G. Ryle, *La notion d'esprit. Pour une critique des concepts mentaux*, (traduit de l'anglais *The concept of Mind* par S. Stern-Gillet) Payot, Paris, 1978, p. 11-24.

181 Le monisme idéaliste a été beaucoup affaibli à partir des découvertes et recherches en chimie, en physiologie et en neurologie du XIXe siècle. À ce moment, les événements mentaux en viennent à être expliqués en termes purement physico-chimiques et neuronaux, et furent considérés comme des phénomènes biologiques comme n'importe quel autre phénomène. Le monisme idéaliste est hors de circulation en philosophie de l'esprit de sorte qu'on associe toujours le monisme au matérialisme.

182 Le problème du rapport entre le corps et l'esprit est mis de côté dans une perspective matérialiste plus radicale. Il n'est pas possible pour les matérialistes, qu'existent des interactions causales entre l'esprit et le corps. Il y a, par contre, des matérialistes comme Searle qui revendiquent la possibilité d'un rapport de causalité entre le corps et l'esprit. Cette interaction reposerait sur une structure de base physique . Searle propose, comme nous allons le voir dans la fin du présent

La théorie de la réduction neurophysiologique ou théorie de l'identité est une sorte de monisme qui considère les événements, les états et les processus mentaux comme identiques à des événements neurophysiologiques. La causalité mentale est réduite à la causalité physique par cette théorie.

La théorie de l'identité, au contraire du behaviorisme logique, n'est pas une théorie sémantique. Elle vise à rendre compte, de façon purement matérialiste, de la causalité des événements mentaux. Dans cette perspective, les événements, états et processus mentaux, comme "avoir peur", "croire aux droits de l'homme", sont des états neurophysiologiques du cerveau et c'est justement cela qui permet aux événements mentaux d'avoir des propriétés causales[183].

Selon la théorie de l'identité il est possible d'analyser objectivement les phénomènes mentaux, car nos représentations mentales se réfèrent à des états physiques. Il est possible alors de comprendre l'esprit à partir de ses manifestations neurophysiologiques. La théorie de l'identité n'a pas un caractère sémantique, dans le sens où elle ne s'intéresse pas aux contenus de ce que les phénomènes mentaux représentent. Pour la théorie d'identité il ne s'agit pas d'expliquer les états intentionnels, mais les phénomènes neurophysiologiques qui correspondent à ces états dans la structure du cerveau.

La théorie de l'identité peut avoir deux aspects théoriques distincts, à savoir: le physicalisme événementiel, et le physicalisme syntaxique.Ces deux perspectives représentent deux points de vue opposés sur l'esprit, lesquels sont en rapport avec l'IA.:"Le physicalisme syntaxique ou structurel refuse aux machines la *possibilité logique* d'avoir des propriétés mentales, tandis que le physicalisme événementiel est favorable à *la possibilité logique* d'une théorisation dans ce sens"[184].

Nous ne pouvons pas discuter ici la portée épistémologique du physicalisme syntaxique ni du physicalisme événementiel; cependant il serait intéressant de les différencier, et de voir dans quel sens ces deux points de vue de la théorie de l'identité pourraient avoir quelque chose à dire sur la recherche sur des machines intelligentes.

Le physicalisme événementiel est la théorie de l'identité qui s'intéresse aux états mentaux individués (par exemple, la peur qu'a X des serpents). Selon cette théorie, les états mentaux se réduisent à des propriétés neurophysiologiques. La notion d'identité du

---

travail, un genre de théorie d'identité qui vise à éviter les problèmes du physicalisme et du mentalisme dans l'analyse des rapports entre le corps et l'esprit.

183 Cette position est proche de la position moniste de J. Searle qui nous allons discuter dans le dernière chapitre de ce travail.

184 Cf. J. Fodor, *op. cit.*, p. 82.

physicalisme événementiel n'est qu'une nécessité logique: être dans un état mental donné est logiquement identique à être dans un état physiologique donné.Le fait que le physicalisme événementiel admet une identité logique entre les états mentaux et les états physiques lui permet aussi d'identifier logiquement les états d'une machine à des propriétés neurophysiologiques du cerveau.

Ainsi le physicalisme événementiel ne rejette pas la possibilité d'existence de machines intelligentes. Cela pour des raisons logiques[185]. L'approche symbolique ascendante de l'IA semble être entièrement compatible avec la théorie événementielle.

Le physicalisme syntaxique ou structurel est la théorie de l'identité qui s'intéresse aux propriétés générales de la vie mentale (par exemple, avoir peur d'animaux). Cette approche de l'esprit affirme la possibilité d'existence des états mentaux *uniquement* en tant qu'états neurophysiologiques.

Le physicalisme syntaxique considère que l'identité entre le physique et le mental ne se restreint pas à une nécessité logique, car l'identité entre événements mentaux et événements physiques devrait avoir la même portée empirique que des réductions physico-chimiques, comme "eau=H2O".

En d'autres termes; "avoir peur" = "il y a quelque état physiologique donné dans le cerveau.". L'identité a une portée empirique. La machine ne peut avoir des états physiques semblables aux états physiques humains, car elle a une structure non-biologique; il n'est pas possible non plus qu'elle aie des événements mentaux. Le physicalisme syntaxique refuse toute possibilité aux machines d'avoir des états mentaux par le fait qu'elles ne possèdent pas de neurones[186].

3.3-.Le fonctionalisme et l'approche computationelle de l'esprit

La présupposition que la pensée est une forme de traitement symbolique des informations reflèterait la défense d'un parallélisme entre la syntaxe et la sémantique. Dans ce cas, les ordinateurs offriraient un modèle informatique de l'intelligence, ou plutôt de la pensée, F. Varela confirme cette idée et montre bien le rapport entre syntaxe et sémantique inspiré du fonctionalisme et defendu par les cognitivistes:

---

185 *Idem* , p. 82.

186 *Idem*. Searle a des arguments qui se rapprochent de cette position naturaliste du rapport entre le corps et l'esprit, mais sa position philosophique , comme nous allons le voir, n'est pas syntaxique, mais sémantique.

> Un ordinateur cependant, ne manipule que la forme physique des symboles. Il n'a aucun accès à leur valeur sémantique. Ses opérations sont néanmoins sémantiquement contraintes car toutes les distinctions sémantiques en jeu dans une computation sont exprimées par le programmeur au moyen de la *syntaxe* du langage utilisé. Car dans un ordinateur, la syntaxe reflète ou est parallèle à la projection sémantique. Le cognitiviste prétend alors que ce parallélisme démontre la réalité physique et mécanique de l'intelligence et de l'intentionnalité (sémantique). L'hypothèse est donc que les ordinateurs offrent un modèle mécanique de la pensée, ou, en d'autres mots, que la pensée s'effectue par une computation physique de symboles[187].

Le fonctionnalisme est une nouvelle théorie de l'esprit basée sur la notion de traitement d'information qui donne priorité à une approche formelle de celui-ci. Il est en rapport avec le développement des recherches en cybernétique en théorie de l'information, en linguistique, en psychologie et en IA. Le fonctionnalisme est l'exemple le plus remarquable du rapport entre l'IA et la philosophie; ce courant de la philosophie de l'esprit a été fortement influencé par la science informatique[188] et par l'IA.

Le fonctionnalisme défend la thése que tout système peut être analysé à partir d'un niveau descriptif intermédiaire entre les descriptions physiques et les descriptions en termes mentaux. Ce niveau intermédiaire privilégie des descriptions de caractère formel. Selon le fonctionalisme il est possible de rendre compte de certaines propriétés (les propriétés mentales) d'un système matériel (le cerveau, par exemple) au moyen des formalisations, lesquelles sont des fonctions abstraites complètement indépendantes de la base matérielle du système. Ainsi, pour le fonctionnaliste l'esprit peut être étudié comme une fonctionnalité, indépendamment de la réalité physique du cerveau.

Le fonctionnalisme propose qu'il est possible d'expliquer les phénomènes mentaux en termes a) d'une individualisation des états mentaux qui composent un événement mental donné b) d'une analyse des relations que les états mentaux entretiennent les uns avec les autres et c) de la spécification fonctionnelle donnée par le rôle causal de ces états.

Autrement dit, nous pouvons décrire la psychologie d'un système soit-il humain ou artificiel sans faire appel aux propriétés matérielles qui le constitue. Pour avoir un esprit, un tel système n'a pas besoin d'être constitué d'un matériel particulier, des neurônes, des *puces* de silicium ou de quoi que ce soit. Il faut simplement que cette matière soit assemblée de

---

187 F. Varela, *op.cit.* , pp.38-39.

188 Pour les fonctionnalistes les théories développées en informatique, en particulier celles dans le domaine de l'IA peuvent lancer une lumière sur plusieurs problèmes en rapport avec la nature de la pensée. Un des exemples le plus mentionés pour montrer que le fonctionnalisme s'inspire de l'appareil conceptuel de l'informatique est celui de la distinction entre software et hardware. Les representations mentales sont comprises pour le fonctionnaliste (par analogie aux rapports entre le materiel et le logiciel) comme des structures de données de caractère symboliques dans le même sens que les structures de données dans un système informatique.

telle et telle façon et ait une structure capable de se traduire en états internes qui permet d'avoir des états mentaux. Pour les fonctionnalistes, donc, il n'est pas impossible que des êtres dont la structure biologique est différente de la nôtre puissent être intelligents.

Les thèses fonctionnalistes de Fodor sur l'esprit et autres présupposent qu'il existe un niveau théorique intermédiaire pour décrire le fonctionnement de l'esprit, et ce niveau de description intermédiaire entre les descriptions des phénomènes mentaux et physique a un caractère formel. Il n'est pas question selon Fodor de décrire les états mentaux comme identiques au cerveau; pour lui, ils sont des propriétés formelles des systèmes physiques ayant un certain type d'organisation. Les états mentaux sont des fonctions abstraites indépendantes de la réalité physique du système qui les réalise.

Le fonctionnalisme défend l'idée que si un événement mental occupe une certaine fonction cela n'implique pas qu'il soit réalisé toujours sur une base physique spécifique, par exemple, le cerveau. Deux systèmes physiquement distincts, mais ayant les mêmes fonctionnalités peuvent avoir des états mentaux identiques. D'ailleurs, une des thèses fondamentales du fonctionnalisme est que les états mentaux peuvent être réalisés dans des structures physiques multiples: il s'agit de la thèse de la réalisabilité multiple des états mentaux[189].

Le fonctionnalisme définit un état mental par ses relations causales avec d'autres états mentaux Pour certains philosophes fonctionnalistes comme J. Fodor, la causalité mentale est une variété de la causalité physique; pour lui, les relations corps-esprit sont causales.

Le fonctionnalisme définit un état mental par ses relations causales avec d'autres états mentaux. Pour certains philosophes fonctionnalistes comme J. Fodor la causalité mentale est une variété de la causalité physique; Le fonctionnalisme défendu par cet auteur n'est pas incompatible avec les thèses d'identité. Cependant, le fonctionnaliste ne s'interresse pas au caractère physique des systèmes mais à leurs états et propriétés causales.

Selon Fodor la structure psychologique d'un état mental individué est déterminée par le rôle causal (fonctionnel) de cet état au sein de l'activité mentale du système. Selon Fodor une fonction permet d'individualiser un état mental car elle représente le mode particulier de sa causalité. Un état mental est toujours lié (correlé) causalement à d'autres états mentaux à l'intérieur du système cognitif et il est défini fonctionnellement par ses relations causales avec d'autres états mentaux.

---

189 Z. Pylyshyn, *Computation and cognition: Toward a Fondation for Cognitive Science*, Cambridge Mass. , 1984.

Il est intéressant de noter que le modèle cognitif du fonctionnalisme est le modèle d'une machine de Turing: Considérons un système cognitif E et ses états corrélés E_1, E_2, E_3,.... , E_n. Ils représentent des dispositions du système à se comporter de telles façon, mais ils ne sont pas des dispositions comportementales. L'état E2, par exemple peut ne pas déclencher un comportement car il dépend des autres états internes du système.

La base de la philosophie de l'esprit fonctionnaliste est l'idée représentationnaliste selon laquelle les états mentaux ont un contenu intentionnel ce qui équivaut à dire qu'ils ont certaines propriétés sémantiques[190].

Pour Fodor les états mentaux individualisés sont analysés comme des symboles et, les processus mentaux sont considérés comme des opérations sur des symboles. Il croît qu'il existe un niveau linguistique, les signes avec lesquels nous exprimons nos états mentaux(le langage), et un niveau symbolique mental, les symboles mentaux de nos pensées. Il aurait une sorte de paralelisme entre les signes du langage et les symboles mentaux. Les symboles mentaux sont des représentations mentales et ont des propriétés sémantiques.

> Supposez qu'il existe des symboles mentaux (réprésentations mentales) et que ceux-ci possèdent des propriétés sémantiques. Croire à quelque chose signifie alors être rattaché à un symbole mental, la croyance héritant ses propriétés sémantiques du symbole qui figure dans la relation. Les processus mentaux(la pensée, la perception, l'apprentissage, etc.) entraînent des intéractions causales parmi des états relationnels tels que le fait de croire à quelque chose. Les propriétés sémantiques des mots et des phrases que nous prononçons sont à leur tour héritées des propriétés sémantiques exprimées par le langage[191].

Retournant à notre exemple de la croyance, pour le fonctionnaliste croire que p, est égal à avoir une certaine configuration symbolique de caractère mental dont les propriétés sémantique sont derivées des relations symboliques entre la pensée et le langage. Les propriétés sémantiques des mots et des phrases sont en rapport avec les propriétés sémantiques du langage, et les propriétés sémantiques des états mentaux sont aussi associées aux propriétés sémantiques des symboles mentaux. Selon Fodor le langage mental permet de rendre les propriétés sémantiques des états mentaux des éléments représentables, capables d'instancier un programme d'ordinateur[192].

---

190 Dans la perspective intentionnaliste, par exemple, le fait que x croit que y est z implique une relation à trois niveaux entre x, la "croyance que y" et une proposition "x est z" qui est le contenu de la croyance de x.
L'état mental de x, selon le point de vue intentionnaliste contient de propriétés sémantiques telles que (a) exprimer une proposition, (b) être vrai ou faux.
191 J. Fodor, *op.cit.* , p. 86.
192 *Idem.*

Selon Fodor le fonctionnalisme est héritier de la tradition représentationnaliste à laquelle il appartient. Fodor nous confirme cet héritage lointain:

> On a cependant abordé l'analyse de l'esprit par le biais des représentations, bien avant la naissance de l'ordinateur. Nous sommes renvoyés à l'épistémologie classique, tradition à laquelle appartiennent des philosophes aussi différents que John Locke, David Hume, George Berkeley, René Descartes, Emmanuel, kant, John Stuart Mill et William James[193].

Fodor mentionne l'associationnisme comme exemple que l'analyse de l'esprit par le moyen des représentations. Le fonctionnalisme, la psychologie cognitive et l'IA n'appliquent pas intégralement la théorie associationniste de Hume mais, en tant qu'approche représentationnaliste ces trois domaines profitent de quelques idées importantes de ce philosophe qui ont quelque fois un caractère associationniste.

Le fonctionnalisme, par exemple s'inspire de la notion huméenne de symbole mental mais, met de côté la notion de ressemblance de Hume comme modèle d'explication de propriétés sémantiques des représentations mentales[194].

Du point de vue du fonctionnalisme, l'analyse d'un état mental, comme par exemple une croyance, implique d'en founnir une explication sur comment cet état mental constitue une représentation et comment cette représentation peut avoir des propriétés sémantiques. Selon Fodor les croyances semblent entretenir des relations avec des représentations mentales sémantiquement interprétées. La difficulté est de fournir une explication acceptable sur l'origine des propriétés sémantiques et des représentations mentales[195]. Le fonctionnaliste cherche à résoudre un tel problème en admettant que les propriétés sémantique d'une représentation mentale peuvent être fournies à partir de l'analyse du rôle fonctionnel de cette représentation, c'est-à-dire, à partir de la spécification de la représentation. Cela est une condition suffisante pour l'explication des propriétés sémantique des représentations.

La façon de spécifier le rôle fonctionnel des représentations mentales est de déterminer ces propriétés sémantiques à partir de leurs rôle causal. Selon Fodor il existe trois types de relations causales à la base de nos représentations mentales: (1) les relations causales entre les états mentaux et les stimuli, (2) les relations causales entre les états mentaux et les réponses, (3) le relations causales entre les différents états mentaux. Selon Fodor les

---

193 *Idem.*
194 *Idem.*
195 *Idem.*

relations causales (1), (2) et (3) possèdent un rapport entre elles lequel est expliqué pour lui de la façon suivante:

> Considerons la croyance « jean est grand ». On suppose que les faits suivants, qui correspondent respectivement aux trois types de relations causales, sont adéquats pour déterminer les propriétés sémantiques des repésentations mentales mises en jeu dans cette croyance. Prémièrement, la croyance est l'effet normal de certaines stimulations, telles que: voir Jean dans des circonstances accusant sa taille. Deuxièmement, la croyance est la cause normale de certains effets comportementaux tels l'énonciation: « jean est grand. » Troisièmement, la croyance est la cause normale et l'effet normal de certaines autres croyances. Par exemple, quiconque croit Jean grand, sera prêt à croire également que quelqu'un est grand. Avoir la première croyance est, normalement une cause suffisante pour avoir la seconde. Et quiconque croit à la fois que toutes les personnes présentes dans une pièce sont grandes et que Jean se trouve dans cette pièce, aura fortement tendance à croire que Jean est grand. La troisième croyance est l'effet normal des deux prémières. En bref, le fonctionnaliste soutient que la proposition exprimée par une repésentation mentale donnée dépend des propriétés causales des états mentaux dans lesquels figure cette représentation[196].

Considérant (a) qu'il n'y a pas d'IA sans l'appel théorique et empirique aux ordinateurs et (b) que les ordinateurs digitaux sont des machines, comme les machines universelles de Turing dont les capacités dépendent du fait qu'elles manipulent des symboles, le fonctionnalisme répond aux besoins théorique de l'IA car elle propose une théorie formelle de l'esprit compatible avec les méthodes et thèses dans le domaine de l'IA.

La notion de causalité des théories fonctionnalistes n'est pas une notion classique (dans le sens où toute relation causale a un caractère purement physique) La notion de causalité est une notion abstraite qui met en rapport des états mentaux (en tant que des représentations) de caractère symbolique selon le modèle formel de la machine de Turing.

L'analyse formelle des propriétés sémantiques des représentations et de la notion de représentation mentale en termes symboliques à la base du fonctionnalisme, font en sorte que ce courrant soit caractérisé comme une sorte de philosophie officielle de l'IA et des sciences cognitives d'une manière générale.

En gros, le fonctionnaliste affirme que nous aurons compris le fonctionnement de l'esprit lorsque nous aurons conçu un programme qui soit l'équivalent fonctionnel de celui-ci. Si nous pouvons, par les moyens formels dont nous disposons, concevoir des

---

196 . Fodor, *op.cit.*, p. 87.

programmes ou des machines capables de réaliser les mêmes processus que ceux produits par l'esprit humain, alors l'intelligence artificielle est possible.

Pour le fonctionalisme, les processus mentaux peuvent être analysés en termes de mecanismes simples de manipulation de symboles. Fodor explique l'argument de base du fonctionnaliste en identifiant: " les processus mentaux, postulés en psychologie, aux opérations d'une classe restreinte d'ordinateurs: les machines de Turing (...) "[197]. Dans cette perspective toute description suffisamment détaillée d'un système psychologique artificiel ou non en psychologie doit être formulée comme un jeu d'instructions (programme) pour des machines de Turing. Tel semble être la position philosophique sur l'esprit, selon Dreyfus, des chercheurs en science cognitive comme Miller, Pribram, Newel et Neisser. Fodor en donne plus de détail à propos du rapport entre la conception fonctionnaliste de l'esprit et les machines de Turing:

> Les états-programmes de la machine de Turing ne sont définis qu'en termes d'entrée-sortie inscrits sur la bande, d'opération élémentaires et autres états du programme. Tous les états-programmes sont ainsi fonctionellement définis par la part qu'il prennent dans l'opération d'ensemble de la machine. Comme le rôle fonctionnel d'un état dépend autant de sa relation avec d'autres états qu'avec les entrées et les sorties, la version «machine de Turing » du fonctionnalisme illustre bien le caractère fonctionnel du mental. (...) Nous nous proposons ici de restreindre la définition fonctionnelle des états psychologiques à ceux qui peuvent être exprimés en termes d'états-programmes d'une machine de Turing. Cette restriction étant observée, nous avons la garantie que les théories psychologiques sont compatibles avec les exigences des mécanismes (...) En conséquence, en formulant une explication psychologique comme un simple programme d'une machine de Turing, le psychologue garantit une explication mécaniste, même si le type de *hardware* qui réalise le mécanisme n'est pas précisé[198].

Le fonctionalisme propose une solution au problème du corps et de l'esprit capable de surmonter les difficultés trouvées par le behaviorisme logique et par les théories de l'identité. Il faut que les descriptions des propriétés mentales soient analysées en termes relationnels, c'est-à-dire, que les théorisations sur l'esprit puissent être faites sans faire appel aux structures physiques en jeu, mais à partir de relations entre les propriétés de l'esprit.

> Le fonctionnaliste peut prétendre aisément, à la fois que les propriétés mentales se définissent spécifiquement par leurs relations, et que les interactions corps-esprit sont typiquement causales, et ce, quel que soit le bien-fondé de la notion

197 J. Fodor, *op.cit.*, p. 84.
198 *Idem.*

> de causalité exigée en psychologie. Le béhavioriste logique ne souscrira qu'à la première affirmation et le partisan du physicalisme structural qu'à la seconde[199].

Selon Fodor, la position de la théorie fonctionnaliste de l'esprit peut combler les faiblesses du behaviorisme logique et de la théorie de l'identité. Ces faiblesses étaient les suivantes: le behaviorisme logique peut rendre compte des interactions causales entre le corps et l'esprit, mais ne peut pas faire la même chose en ce qui concerne le caractère *relationnel* des propriétés mentales. Pour la théorie de l'identité, c'est l'inverse qui se passe. Pour Fodor, seule une position fonctionnaliste peut rendre compte de ce double caractère (formel et causal) du mental.

La théorie fonctionnaliste de l'esprit s'oppose radicalement au behaviorisme,mais elle en concerve quelques traits. Un de ces traits est la thèse que les concepts mentaux du sens commun peuvent être expliqués partiellement en termes du rôle causal des comportements.

Selon le fonctionalisme, les états psychologiques, comme les désirs, les croyances les douleurs sont des types d'états psychologiques qui ont un rôle fonctionnel causal qui interviennent lors de l'interaction de l'organisme ou du système avec l'environnement.

Le fonctionalisme est différent du behaviorisme; premièrement il traite les états mentaux comme des causes authentiques; ces sont des états internes actuels qui jouent un rôle dans la production du comportement. Le fonctionnaliste ne voit pas les états mentaux tout simplement comme une façon abrégés de parler de régularités comportementales. Quelque fonctionalistes ont comme point de départ la théorie "type-identity". Selon les défenseurs de la théorie de l'identité type les événement mentaux sont identifiés à des types d'événements physiques.

Le fonctionalisme propose d'expliquer les rapports entre le corps et l'esprit par intermédiaire d'une Théorie Computationnel de l'Esprit (TCE). La TCE est inspirée des recherches en IA elle est née des analogies entre les programmes informatiques de l'IA et les thèses fonctionnalistes sur la façon dont l'esprit est organisé.

Le fonctionnaliste défend des thèses qui sont compatibles avec le physicalisme événementiel dans le sens il où s'intéresse aux états mentaux individués et aussi par le fait qu'il ne rejette pas la possibilité logique des machines d'avoir des états mentaux si elles sont convenablement programmées. Le fonctionnalisme est aussi compatible avec les thèses de la réduction neurophysiologique, car il défend lui aussi la thèse matérialiste de l'identité entre les états mentaux et les états physiques (de caractère neurophysiologique.)

---

199 J. Fodor, *op. cit.* , p. 83.

L'approche computationnelle de l'esprit, nous le répétons, associe les états mentaux à des états de machine. La notion d'état, dans ce cas, ne concerne pas des états physiques, mais des états abstraits: les états d'une machine abstraite sont reliés à des états mentaux. Autrement dit, le mot "machine" dans le cas de l'approche computationnelle de l'esprit ne designe pas des machines physiques, mais une machine abstraite du type machine de Turing.

Cette approche est défendue par des auteurs comme J. Fodor et Pylyshyn, elle se distingue des thèses monistes qui réduisent le mental au physique. Pour Fodor par exemple, les états mentaux sont similaires à des états abstraits d'une machine[200].

Le fonctionnaliste ne nie pas que lorsque nous avons un état mental nous sommes également dans un état physique donné, comme l'affirment les thèses d'identité. Mais cela ne veut pas dire que le fonctionnement physique du cerveau est identique au fonctionnement abstrait de l'esprit. L'esprit a en général une base physique ; pour avoir un esprit il faut avoir un cerveau. Cependant, le caractère physique sous-jacent à l'esprit est une condition nécessaire, mais non suffisante pour avoir un esprit. Les états mentaux ne dépendent pas du rembourrage physique qui constitue le système; bien au contraire, ils sont indépendants du physique. En résumé: le fonctionnaliste s'intéresse à la structure configurationnelle des états d'un système sans se préoccuper de sa constitution physique.

Les états mentaux, comme les états d'une machine abstraite ont un caractère transitionnel. Pour le fonctionnaliste, les configurations physiques (analysées comme des états intermédiaires de nature abstraite) de l'organisme correspondent à des états ou transitions d'états par lesquels l'esprit passe. La configuration des transitions que des états mentaux ont selon les thèses fonctionnalistes un caractère formel car elles sont régies par des règles. La découverte de ces règles permet de rendre compte du fonctionnement de l'esprit.

Fodor entend que les états mentaux sont comme des états d'une machine de Turing et peuvent être modifiés en fonction de la configuration générale du système dans lequel il se trouvent, et du rapport qu'ils (ces états mentaux) entretiennent avec les autres états. (caractére transitionnel des états mentaux).

L'approche computationnelle de l'esprit adopte la thèse selon laquelle l'esprit en tant que système de traitement de l'information est un processus computationnel. Selon cette approche de caractère fortement syntaxique, la syntaxe du système correspond à sa

---

200 Le fonctionnaliste ne croit pas à des analogie entre le cerveau et la machine, comme dans les thèses physicalistes de l'IA, mais plutôt à une analogie entre l'esprit et la machine.

sémantique. Cependant, les contenus sémantiques des symboles ou représentations sont externes au système: les symboles non-interprétés au niveau syntaxique interne produisent des transformations importantes dans le système, lesquels reçoivent ensuite une interprétation externe.

Un processus computationnel est un ensemble de transformations symboliques non-interprétées régies par un ensemble précis de règles de transition.

Les termes pour décrire les états mentaux correspondent à des notions théoriques, comme par exemple la "douleur", la "crainte", le "plaisir" qui correspondent à des états ou à des transitions d'états de la pensée humaine. L'esprit, selon l'approche computationelle de l'esprit est un ensemble complexe d'états et de transitions d'états mentaux d'un système indépendamment de sa constitution physique sous-jacente.

Dans la perspective de la TCE les systèmes cognitifs sont des systèmes formels qui manipulent des symboles au moyen de règles. Autrement dit, l'esprit, d'après cette conception, fonctionne selon un processus de manipulations symboliques internes, lesquelles ne font pas appel à des manipulations symboliques externes de caractère sémantique. L'esprit fonctionnerait, en résumé, selon des conditions formelles, (*formalities conditions* [201]), de la même façon qu'un système informatique. Pour Fodor, un grand défenseur du courant fonctionnaliste, nous n'avons pas besoin d'un modèle anthropomorphique de cognition (basé sur des analogies réductionnistes entre la machine et le cerveau) pour étudier comment l'esprit fonctionne.

> Dans le cadre de l'analogie avec l'ordinateur, on compare les terminaisons des organes des sens avec des transducteurs par lesquels les événements du « monde extérieur » sont enregistrés pour être traités. Quand on travaille avec une machine de Turing, cela correspond à l'impression de certains signes sur la bande d'entrée. Le système de traitement de l'information lui-même (en l'occurrence: l'esprit) n'a pas le moindre contrôle sur la relation (sémantique) entre les données enregistrées et le monde « au -delà » des transducteurs (ou terminaisons des organes des sens). Son accès au monde s'effectue à travers des données livrées par les transducteurs (...) Dans cette mesure, le système ne peut pas non plus « juger » s'il existe réellement un monde « au-delà » des transducteurs. De toute façon, la question concernant ce « monde extérieur » est sans importance pour la « vie intérieure » du système de traitement de l'information. Pour pouvoir effectuer ses processus celui-ci n'a besoin que de quelques données dans sa « mémoire » (quelques signes dans la bande d'entrée (input-tape)[202].

---

201 Le sens de cette expression est explicité dans J. Fodor,"Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", *in Representations*, The Harvester Press, Brighton, 1981, p.239. Cf. W.P. Mendonça, *op. cit.* , p.9.

202 W. P. Mendonça, "Intelligence Artificielle et signification. À propos des limites et des possibilités des sciences cognitives"*in Philosophique*, vol XVII, Numéro 1, printemps, 1990 p.9 Le terme "transducteurs" semble avoir plus ou moins le même sens que le terme "compilateurs" en informatique, comme nous l'avons mentionné ci-dessous pour plus de

La TCE, considère que l'esprit n'a pas une nature physique, mais une nature formelle ou logique. Selon ce point de vue,les états mentaux ne sont pas considérés comme des propriétés d'un système matériel, mais plutôt comme des fonctions abstraites. Les rapports entre le cerveau et l'esprit ont dans cette perspective un caractère fonctionnel et la notion d'état est dans ce sens privilégié: l'esprit est défini par la TCE comme un ensemble d'états de l'architecture fonctionnelle du cerveau, lesquels correspondent à des états de l'architecture fonctionnelle d'une machine de Turing.

Le fait que la TCE mette l'accent sur la notion d'état lui permet de défendre la thèse selon laquelle tous les comportements intelligents peuvent être analysés et décrits à partir d'un niveau fonctionnel intermédiaire entre les descriptions en termes mentaux et les description en termes physiques.

La notion abstraite de fonction est fondamentale pour le fonctionnalisme qui propose une analyse des structures des représentations mentales à partir des rôles fonctionnels des états mentaux[203].. Pour Fodor les états mentaux sont définis par leurs fonctionnalités ou leurs rapports avec d'autres états[204]. Ils ne sont pas des états physiques, mais des états fonctionnels abstraits d'un système, comme des états d'une machine de Turing[205].

Les états mentaux d'un système sont définis par les rôles fonctionnels qu'il occupent L'esprit possède une architecture fonctionnelle qui met en rapport les fonctions du niveau syntaxique avec les représentations du niveau sémantique.

Si nous pouvons faire des descriptions d'un état mental en termes de fonctions calculables nous pouvons reproduire cet état sur un autre système formel, comme par exemple, l'ordinateur digital.

En tant que fonction abstraite, l'état mental est indépendant de la réalité physique du système qui le réalise, c'est-à-dire, que même si un état mental est causé par une

---

details voir Fodor, J. A, *La modularité de l'espri, Essay sur la psychologie des facultés*, Ed. Minuit, 1986, pp.57-66, traduit de l'american *The Modularity of Mind*, 1983, par Abel Gerchenfeld.

203 L'esprit analysé comme fonction est une propriété abstraite du système nerveux; il n'a, ainsi, rien de physique. Nos états mentaux en tant que fonctionnalités ont une autonomie par rapport au système physique. L'idée de base de la théorie computationnelle de l'esprit est que celui-ci est compris à partir d'un modèle abstrait de machine de Turing. L'esprit est compris selon cette théorie en termes d'états (mentaux) élémentaires, lesquels sont réduits à des fonctions calculables qui peuvent être traitées par n'importe quel autre système formel tel qu'une machine de Turing ou un ordinateur.

204 Le fonctionnalisme est en rapport avec l'holisme du mental. Le fonctionnaliste défend l'idée que les états mentaux seulement produisent le comportement lorqu'ils sont articulés à d'autres états mentaux. Pour le fonctionnaliste l'esprit est défini par les rôles fonctionnels des états mentaux qui sont définis non seulement par les relations des états mentaux avec une entrée(input) et une sortie(output) du système, mais aussi par les relations de tels états avec d'autres états de nature différente qui comptent comme des instances du même type psychologique.

205 Selon Fodor l'esprit ou les processus mentaux représentationnels sont similaires à ceux d'une machine abstraite telle que la machine deTuring. Pour l'auteur, les processus cognitifs résultent des représentations internes lesquelles sont indépendantes des interprétations sémantiques externes au système. (cf J. Fodor, *op. cit.* pp. 84-85.)

configuration physique particulière du corps il est indépendant du corps, car il est d'une autre nature.[206] Ainsi, les états mentaux, par leur indépendance des phénomènes matériels peuvent être réalisés, sans problèmes, dans des systèmes physiquement différents (thèse de la réalisabilité multiple des états mentaux).

Selon la thèse de la réalisibilité multiple des états mentaux s'il y a un programme pour faire des choses comprises comme étant intelligentes (et les fonctionnalistes pensent qu'il y en a) cet algorithme peut être exécuté sur n'importe quel système formel indépendant de sa constituition physique. Cette thèse a une signification important pour l'IA car elle serv d'inspiration à la continuation du projet représentationnaliste de mécanisation de la pensée, c'est-à-dire, de la conception d'une pensée mécanique (mecanical thought).

Conclusion:

Nous avons vu que l'IA est un projet résultant de la tradition représentationnaliste et qu'elle est en rapport avec le modèle de rationalité de la science en Occident selon lequel il est possible d'expliquer la nature et nos pensées au moyen des représentations formelles.

Nous avons mis l'accent sur plusieurs aspects de la tradition philosophique représentationnaliste afin de mieux comprendre la portée des critiques de Dreyfus à l'IA lesquelles ont comme point de départ, nous le verrons, les limitations des modèles représentationnalistes. D'autre part nous avons discuté les rapports entre les thèses fonctionnalistes de l'esprit et l'IA pour montrer que le fonctionnalisme fait partie de la tradition philosophique représentationnaliste. Toute notre exposition sur ce sujet est complémentaire au dernier chapitre de ce mémoire sur les critiques de Searle à l'IA.

Le devéloppement de la science physique et des mathématiques modernes sont en rapport avec l'IA. Les nouvelles conceptions de calcul apparues après Galilée changèrent la conception que l'homme avait de lui-même et de son esprit. L'idée que la pensée peut être mécanisée existe depuis longtemps dans la philosophie de quelques empiristes et dans certaines théories rationalistes pour lesquelles penser c'est calculer. Nous avons vu que les philosophes de cette tradition concevaient la pensée comme des symboles régies par un système bien ordonné (rationnel) de régles de notre esprit. L'IA s'appuie sur cette tradition et/ou sur des outils théoriques en rapport avec le mode de pensée scientifique qui en dérive.

---

206 Cet aspect dualiste de certaines thèses fonctionalistes est parfois condamné par certains auteurs fonctionalistes.

Tout notre effort de parcourir certains éléments importants de la tradition représentationnaliste, allant des idées de Leibniz, Hobbes jusqu'au fonctionnalisme a eu pour fin de montrer que l'IA est profondément ancrée dans un modèle philosophique réprésentationnaliste de la pensée. Ainsi, nous pouvons affirmer que les rapports entre l'IA et la phisophie sont des plus étroits.

Les sciences de la nature ont eu une répercussion énorme sur notre pensée philosophique, en particulier sur notre compréhension de l'esprit. Le succès de la physique à fournir des explications sur le fonctionnement de l'univers a eu une influence importante sur la philosophie en Occident, ce succès a produit plusieurs changements importants dans l'analyse des rapports entre le corps et l'esprit suscitant, par la conception des modèles théoriques, la recherche pour la mécanisation de la pensée. Le developpement des systèmes formels de plus en plus puissants pour la description du mouvement des planétes et l'explication de la nature changerent peu à peu notre façon de comprendre l'esprit. À partir de celà même les pensées en viennent à à être comprise en termes d'éléments plus simples capables d'être représentés et calculés. Cette notion de pensée en tant que mécanisme de calcul en rapport avec des comceptions scientifiques en Occident va reaparraître dans nos théories contemporaines sur l'esprit tel que le fonctionalisme.

Nous avons constaté, à partir de la lecture de plusieurs textes dans le domaine de la philosophie sur l'IA que ce theme agit comme un catalyseur de nouvelles discussions philosophiques importantes. Les domaines philosophiques les plus touchés sont ceux en rapport avec la logique et les théories sur l'esprit et sur le langage. L'IA s'affirme de plus en plus comme un thème du domaine philosophique, permettant toujours de faire ressortir l'actualité des discussions philosophiques comme base au développement critique et épistémologique de la recherche fondamentale en informatique[207]. Cela par la simple raison qu'elle est liée à une tradition philosophique où la représentation formelle de la pensée a toujours été valorisée. Elle est une continuation d'un projet scientifique de pouvoir tout expliquer au moyen de représentations formelles:

> (...) Many contemporary philosophers working in philosophical logic and the semantics of natural language share at least the goal of devising a rigorous logical system in which every statement, every thought, every hunch and wonder can be unequivocally expressed. The idea wasn't reinvented by AI; it was a gift from the philosophers who created modern mathematical logic: George Boole,

---

207 Cf. B. Buchanan, "Artificial Intelligence as an Experimental Science" *in* James H. Fetzer (éd.), ***Aspects of Artificial Intelligence***, Kuwer Academic Publishers, 1988, pp. 209—250. Voir aussi D. C. Dennett, ***Brainstorms, Philosophical Essays on Mind and Psychology***, Bradford Books, Publ. Inc. Montgomery, Vermon, 1978. Voir en particulier le chapitre 2' "Artificial Intelligence as Philosophy and as Psychology" pp.109—126.

Gottlob Frege, Alfred North Whitehead, Bertrand Russell, Alfred Tarski, and Alonzo Church[208].

Nous n'avons pas une théorie générale sur l'esprit comme nous avons des théories générales en physique. Cependant les développements théoriques et pratiques des capacités de manipulation de symboles par les machines digitales ouvrent des possibilités jamais connues à la consécration de la notion de pensée en tant que calcul mécanique faisant de l'IA non seulement un nouveau domaine de recherche scientifique à être développé mais un domaine d'intérêt philosophique.

Nous avons constaté aussi par la lecture de plusieurs travaux et articles sur l'IA qu'un tel thème, suscite trois types d'attitudes de la part des philosophes et scientifiques:

1. Du scepticisme complet, (quelquefois, non justifié) qui est peut-être lié à des préjugés de toute sorte, ou un scepticisme justifié corroboré par des raisons d'ordre moral, religieux et social[209].

2. L'assimilation des travaux sans critique préalable. Cette attitude est liée quelquefois à des intérêts visant des fins pratiques, d'autres fois à la fascination intellectuelle causée par les opportunités que l'IA offre à la discussion de plusieurs thèmes sur la cognition humaine et sur l'esprit.

3. Et finalement, une analyse critique. Cette attitude représente les préoccupations de caractère épistémologique partagées par des scientifiques et philosophes, qui s'intéressent ou qui travaillent directement sur l'IA. Dans ces deux cas leurs approches sont caractérisées soit par une opposition justifiée aux thèses de l'IA, soit à une adhésion à des présuppositions sous-jacentes à l'IA qui peuvent servir de base à leurs recherches[210].

Comme nous l'avons déjà signalé dans l'Introduction de ce travail nous nous intéressons particulièrement au positionnement critique face à l'intelligence artificielle. La

---

208 D.C. Dennett, "When philosophers Encounter Artificial Intelligence", *in* Stephen R. Graubard, éd. , *The Artificial Intelligence Debate, False Starts, Real Fondations*, MIT Press, Mass. ,1989, p. 288.

209 Il convient de mentionner que l'attitude sceptique face à l'IA qui est basée sur des raisons d'ordre moral et religieux ne représente pas toujours des positions dogmatiques sur la question, mais plutôt, une posture qui mériterait d'être classifiée comme un positionnement critique (attitude 3). Il y a plusieurs travaux sur l'IA dans le domaine de la théologie, de la sociologie et de la technologie qui pourraient être considérés comme des travaux critiques. Dans le domaine des préoccupations d'ordre moral, il faut mentionner le travail *Puissance de l'ordinateur et raisonnement humain* de Joseph Weizenbaum (1976) ex-chercheur en IA. Cet auteur signale que l'intelligence artificielle peut représenter un danger pour des raisons morales. Il soutient que les machines ne seront jamais capables de comprendre des choses comme l'amour, la compréhension, le respect mutuel; les machines pourraient intervenir dangereusement et menacer la vie humaine et la société. Ainsi, selon Weizenbaum, les machines intelligentes ne seraient pas au service des hommes, mais contre eux. La croyance à la métaphore des machines intelligentes peut provoquer des changements sur l'image que l'homme a de lui-même. Selon lui l'IA est possible, mais elle est, cependant, indésirable vu que, sur le plan moral, elle peut affecter les humains en les privant de leur dignité.

210 Nous parlerons de telles presupositions dans le prochain chapitre de ce mémoire.

critique à l'IA se présente sous deux aspects différents, à savoir: en tant que critique faite de l'intérieur de la tradition représentationnaliste et en tant que critique faite de l'extérieur de cette tradition.

La critique à l'IA faite à l'intérieur de la tradition représentationnaliste s'appuie notamment sur des arguments de caractère formel: par exemple, les critiques du genre de celles que nous avons mentionnées, faites par J. R. Lucas, qui sont soutenues à l'aide de résultats logico-mathématiques sur l'incomplétude de l'arithmétique de Gödel. Nous inclurons dans cette critique toute critique de caractère épistémologique et méthodologique faite par ceux qui travaillent directement sur le domaine critiqué.

L'autre genre de critique est produite à l'extérieur de la tradition représentationnaliste et peut avoir deux voies: une sémantique ou linguistique un peu plus proche de la tradition représentationnaliste, et l'autre d'inspiration phénoménologique complètement opposée à cette tradition. La voie sémantique est celle suivie par J. Searle, les conclusions de cet auteur sur l'esprit et sur le langage s'écartent de la tradition formelle représentationnaliste qui a tendance à donner priorité à la syntaxe sur la sémantique. En ce qui concerne la voie d'inspiration phénoménologique le meilleur exemple c'est le travail de H. L. Dreyfus.

Nous allons examiner, dans le cours des deux prochains chapitres, quelques critiques qui sont adressées à l'IA dans les travaux *What Computers Can't Do* et *Minds, Brains, and Sciences*, lesquels sont souvent mentionnés dans les ouvrages consacrés à cette recherche.

SECONDE PARTIE

LES CRITIQUES PHILOSOPHIQUES DE H. L. DREYFUS ET DE J. R. SEARLE À L'INTELLIGENCE ARTIFICIELLE

Greats artists have always sensed the truth, stubbornly denied by both philosophers and technologists, that the basis of human intelligence cannot be isolateded and explicitly understood. In *Moby Dick* Melville writes of the tattooed savage, Queequeg, that he had "written out on his body a complete theory of the heavens and the earth, and a mystical treatise on the art of attaining truth; so that Queequeg in his own proper person was a riddle to unfold; a wondrous work in one volume; but whose mysteries not even himself could read... Yeats puts it even more succinctly: I have found what I wanted—to put it in a phrase, I say, 'Man can embody the truth, but he cannot know it.' [211]

Hubert L. Dreyfus

---

[211] H.L., Dreyfus, op. cit pp. 65-66

CHAPITRE III

Les critiques de H. L. Dreyfus à l'Intelligence Artificielle

Présentation:

Dans *What Computers Can't Do,* Dreyfus cherche à mettre en évidence et critiquer les éléments épistémologiques présents pour légitimer les travaux en IA. À partir de la discussion des présuppositions qui orientent l'IA, Dreyfus fait une investigation sur les bases philosophiques représentationnalistes qui fondent cette recherche. Ses discussions sur l'IA dans *What Computers Can't Do* tournent autour de quatre questions que nous avons formulées ainsi:

1) Une analogie cerveau-machine est-elle valable, c'est-à-dire est-ce que les deux fonctionnent de façon binaire, à partir d'un ensemble de règles explicites ?

2) Est-ce qu'il est possible de créer des ordinateurs et des logiciels capables de représenter les processus qui sont propres à l'esprit humain?

3) Est-ce que nos comportements intelligents peuvent être complètement formalisés, c'est à dire représentés sous la forme de programmes informatiques?

4) Est-ce que la conception représentationnaliste du comportement intelligent (ou de l'esprit) comme étant le résultat d'un ensemble d'éléments discrets manipulés par des règles est correcte?

Nous n'avons pas le temps ici de faire une critique minutieuse de la portée des réponses qu'il donne à chacune de ces questions, cependant nous allons nous retrouver devant ces questions tout au long de ce chapitre afin de montrer l'approche de Dreyfus sur la question de la compréhension et de la description des comportements intelligents en IA et les critiques qu'il fait aux présuppositions de caractère philosophique sous-jacentes aux travaux dans ce domaine.

Le livre *What Computers Can't Do* remet en question des modèles traditionnels d'explication de la pensée, établis dans le domaine de l'IA. Un point important à souligner

avant de commencer est que le but de Dreyfus n'est pas de discréditer la recherche en IA, mais de présenter les faiblesses de ses bases ontologiques et épistémologiques.

Nous allons donner ici un aperçu de quelques critiques que Dreyfus adresse à l'IA afin de faire ressortir des éléments importants capables de révéler ce que cet auteur pense des résultats concrets obtenus dans ce domaine de recherche.

Le livre *What Computers Can't Do* vise surtout à accentuer les réelles limites des projets en IA à partir de la distinction entre les processus cognitifs humains et les programmes en Simulation Cognitive et en IA. Dreyfus s'attaque à l'analogie cerveau-machine et aux modèles antropomorphiques qui sont à la base des recherches en IA. Il touche aussi à la question des limitations sémantiques des programmes.

Le livre mentionné est divisé en trois parties,la première partie du livre est consacrée à l'analyse des résultats de recherches en IA obtenus au cours de la période allant de 1957 à 1967. Cette première partie consiste en l'analyse des deux phases de cette recherche.

Dans sa critique de la première phase de l'IA, il vise à souligner les réelles limites des projets en IA à partir de la distinction entre les processus cognitifs humains et les formes de traitement de l'information par des moyens informatisés en simulation cognitive[212]. Dreyfus s'attaque à l'analogie cerveau-machine et aux modèles anthropomorphiques qui sont à la base des recherches en IA.

Dans l'introduction à l'édition révisée de *What Computers Can't Do* de 1979, Dreyfus en profite pour faire une critique à une troisième phase (1967-1972) et une quatrième phase (1972-1977) où il analyse respectivement le problème des programmes basés sur la notion de micro-monde et le thème lié à la représentation des connaissances courantes en IA[213].

La critique de Dreyfus vise le courant *représentationnel-calculationnel*[214], dont le prolongement contemporain est mis en évidence dans les travaux de l'approche descendante. A partir de l'analyse des échecs et limitations théoriques rencontrées par les chercheurs de ce courant, Dreyfus structure toute une critique de ce qu'il appelle la *raison artificielle* [215].

---

212 Dans cette partie l'auteur reprendre plusieurs critique déjà faites dans son rapport *Alchemy and Artificial Intelligence.*
213 Nous allons parcourrir juste quelques parties de cet ouvrage quel nous avons consideré comme les plus importantes pour les fins de ce mémoire.
214 H.L. Dreyfus, (1989),*op.cit.* , p.973.
215 H.L. Dreyfus,*What Computers Can't Do: A Critique of Artificial Reason,* Harper et Row Publishers, N.Y. ,1972.

1- L'intelligence naturelle

Dreyfus exploite l'opposition entre l'intelligence naturelle et l'intelligence artificielle pour montrer les limites de la simulation cognitive par ordinateur. Il veut montrer aussi que les recherches en IA présentent des problèmes d'ordre empirique et épistémologique. Ces problèmes sont mis en évidence par une anlayse de la façon dont les êtres humains et les ordinateurs digitaux "traitent l'information". Dreyfus signale que les processus qui caractérisent nos comportements intelligents dépendent de quatre propriétés qui sont essentiellement humaines:

1) l'activité de la conscience périphérique (Fringe consciousness).
2) la capacité de tolérance à l'ambiguïté.
3) la distinction entre ce qui dans une situation est essentiel ou non.
4) le regroupement par l' intuition des élément pertinents à partir du contexte de l'expérience.

Selon Dreyfus ces quatre modalités de la cognition, que nous allons résumer plus bas, servent à montrer les différences entre la façon dont les ordinateurs traitent l'information et la façon dont l'esprit opère, nous permettant de reconnaître un visage, de résoudre un problème ou de raisonner logiquement. Pour Dreyfus, ces propriétés résistent à la représention complète par des moyens formels. C'est-à-dire, elles sont difficilement formalisables en termes de programmes informatiques.

1.1- L'activité de la conscience périphérique

Les êtres humains sont capables de travailler avec des informations (non explicitées par le contexte) lesquelles sont dérivées d'une sorte de *background* que Dreyfus appelle la *conscience périphérique.*

La *conscience périphérique* est une capacité humaine d'analyse globale des situations ou problèmes. Elle permet de rassembler toutes les informations concernant la périphérie de notre expérience immédiate et fonctionne comme une sorte d'arrière-plan par rapport à l'objet central de notre attention. La *conscience périphérique* est responsable, par exemple,

de la sélection des informations de caractère sémantique qui sont (périphériques) à l'extérieur d'une situation donnée, mais qui contribuent à sa désambiguïsation.

Pour Dreyfus, il est impossible de simuler la *conscience périphérique*. Les programmes créés en IA ne permettent pas une analyse globale des situations ni l'utilisation effective des expériences du passé pour résoudre un problème. Dreyfus signale que les recherches sur l'apprentissage en rapport avec l'activité de la *conscience périphérique* sont les moins développées en IA.

1.2 - Capacité de tolérance à l'ambiguïté

La deuxième modalité concerne la capacité des êtres humains à traiter l'information avec des degrés de tolérance à l'ambiguïté. Nous pouvons analyser toutes sortes de situations ambiguës, sans avoir besoin de procéder à des descriptions précises pour expliciter les éléments qui composent une information reçue ou une situation, ou problème qui nous est présenté. La capacité de tolérance à l'ambiguité met en rapport notre *mémoire* et notre *conscience périphérique*. Pour Dreyfus, la capacité de désambiguïsation et de saisie globale d'un problème ou situation est due à notre *conscience périphérique* et à notre tolérance à des ambiguïtés. Les deux sont des caractéristiques essentielles à l'activité de l'esprit qui n'ont pas été jusqu'à maintenant programmées sur un ordinateur digital.

1.3- Discrimination entre l'essentiel et le non-essentiel

Cette propriété de l'esprit est en rapport avec l'habileté que nous avons de rendre immédiatement claire une situation complexe impliquant des difficultés d'ordre sémantique et perceptive. Une telle habileté est due à la capacité de l'esprit humain de discriminer entre ce qui est essentiel et ce qui ne l'est pas. Nous avons une façon très spécifique d'analyser un problème selon ce qui nous intéresse immédiatement. Par exemple, toutes les situations qui semblent présenter une difficulté pour la prise de décision en IA, sont intelligibles d'un seul coup par les êtres humains.

Selon Dreyfus, pendant la conception d'un programme de résolution de problèmes (demandant de la prise de décisions) les programmeurs essayaient de représenter de la façon la plus claire possible les problèmes et de spécifier leurs solutions possibles. L'échec des règles heuristiques conçues pour résoudre des problèmes complexes requérant cette

capacité de discrimination entre l'essentiel et le non-essentiel ont montré que les méthodes représentationnels basées sur la logique se montrent très limitées[216].

1.4- Regroupement par l' intuition des éléments pertinentes à partir du contexte

La quatrième modalité de traitement de l'information est due au regroupement de données faites par notre *conscience périphérique* ainsi qu'à notre capacité de recourir au contexte. Ces processus mentaux nous permettent également de reconnaître la forme d'un objet, l'aspect global d'un problème ou d'une situation complexe en identifiant, par exemple, les éléments identificateurs de la forme, les choix à faire pour trouver les meilleures solutions à un problème donnée ou la signification d'une situation.

Étant donné l'insuccès des chercheurs en IA à simuler ces quatre habiletés de l'esprit, il est fort improbable, selon Dreyfus, qu'ils obtiendront du succès dans leurs travaux sur la reconnaissance de formes et la résolution effective de problèmes complexes par ordinateur, basées sur les méthodes traditionnelles de simulation cognitive.

2- L'"intelligence artificielle"

Dreyfus analyse ces quatre modalités cognitives typiquement humaines et les compare à quatre formes informatisées de traitement d'information conçues en IA, à savoir:

1) la recherche guidée par des règles heuristiques.
2) la recherche de précision indépendante du contexte.
3) la triage de l'information pertinente par essais et erreurs.
4) Caractérisation et discrimination des situations, problèmes et solutions à partir des listes de données pertinantes emmagasinées au préalable en mémoire.

Selon Dreyfus, ces quatre formes informatisées de traitement d'information sont très rudimentaires; elles montrent par leurs limitations pourquoi il n'y a pas de programmes capables de reproduire même les performances les plus simples demandant les quatre modalités cognitives humaines mentionnées plus haut.

---

216 Selon Dreyfus cette capacité de selection ou discrimination de ce qui est essentiel et ce qui est un donné accéssoire lors de la résolution d'un probleme quelquonque n'a pas été simulé ni programmé de façon satisfaisante.

Pour Dreyfus, la machine affronte des difficultés, jusqu'à maintenant insurmontables, lorsqu'elle est mise devant un problème ou une situation typiquement humaine. Toute information inattendue peut représenter un problème incontournable à la programmation. Les programmes conçus en Simulation Cognitive pour reproduire certaines capacitées de l'esprit procèdent soit par "essai-erreur" ou par une classification hiérarchique de ce qui peut ou ne peut pas être essentiel.

Pour illustrer et faire ressortir les différences entre l'intelligence humaine et les systèmes créés par l'IA, Dreyfus offre comme exemple une activité humaine complexe, il s'agit du jeu d'échecs:

Un joueur humain dans une partie d'échecs considère à peu près 100 positions probables pour le mouvement des pièces en fonction de l'état du jeu. Mais avant cela il a déjà à l'esprit quel est le mouvement le plus prometteur, car il fait une analyse global du jeu et calcule les possibilités de défense et d'attaque ainsi que les ripostes de l'adversaire afin de décider comment procéder.

Un programme d'ordinateur par contre peut trier jusqu'à 26.000 alternatives avant de pouvoir choisir, car les programmes qui jouent aux échecs sont conçus pour traiter chaque jeu isolement; ils ne considèrent pas le jeu dans sont aspect global et n'utilisent pas des expériences des jeu antérieurs. Les programmes pour jouer aux échecs sont conçus pour considérer seulement des informations explicitées. Ils fonctionnent indépendamment du contexte global du jeu. Pour faire un coup, il faut trier les 26.000 alternatives présentées; cela leur demande un calcul assez long afin de permettre au système informatique d'évaluer une liste d'options possibles[217].

Les différences entre les systèmes humains et les systèmes de l'IA, pour la résolution d'un problème, sont très grandes. Le joueur humain peut sélectionner les coups plus prometteurs de façon qualitativement supérieur car il n'a pas besoin d'inventorier les options possibles de coups pour choisir celui qui est le plus prometteur. Les êtres humains peuvent compter sur des processus beaucoup plus élégants pour reconnaître certaines situations ou problèmes et ensuite traiter les données dont ils disposent[218].

Le joueur humain compte sur les quatre modalités cognitives signalées par Dreyfus et, au contraire de l'ordinateur, il associe les situations présentées pendant le jeu d'échecs avec d'autres expériences semblables déjà vécues par lui. Il visualise le jeu globalement, ce qui

---

217 H. L. Dreyfus (1979), *op.cit.*, pp. 101-102.

218 Pour l'être humain, la perception et l'organisation des éléments en jeu, lors de la résolution d'un problème, sont importants pour la discrimination de tout ce que est pertinants ou fondamental à la résolution de ce problème.

lui permet d'établir un rapport entre la situation actuelle (le problème présenté) et les informations provenantes de sa mémoire ou de sa conscience périphérique. Cela fait en sorte que plusieurs associations mentales en rapport avec des connaissances diverses, en relation avec le jeu actuel et des jeux antérieurs, soient utilisées comme éléments de décision, permettant au joueur de formuler des hypothèses sur les mouvements des pièces sur l'échiquier.

Selon Dreyfus, la programmation des ordinateurs pour qu'ils participent à des jeux comme celui des échecs ne rend pas compte des informations non explicitées. Ces informations constituent des données importantes pour la conduite d'un joueur pendant une partie d'échecs.

Les informations non explicitées qui occupent la périphérie de la conscience pendant le déroulement d'une partie d'échecs sont fondamentales pour les êtres humains pour qu'ils analysent le jeu et les coups possibles, sans avoir besoin de procéder à un recensement méthodique et inconscient de toutes les options offertes, comme c'est le cas des programmes informatiques qui jouent aux échecs. À propos de cela Dreyfus fait le commentaire suivant:

> What is needed is a program which selectively carries over from the past just those features which were significant in the light of its present strategy and the strategy attributed to its opponent. But present programs embody no long-range strategy at all.
>
> In general what is needed is an account of the way that the background of past experience and the history of the current game can determine what shows up as a figure and attracts a player's attention. But this gestaltist notion of figure and ground has no place in explicit step-by-step computation[219].

L'utilisation de protocoles[220] visant à décrire les raisonnements du joueur d'échecs pour concevoir des programmes qui jouent aux échecs, révèle selon Dreyfus l'importance de la *conscience périphérique* dans le cas des joueurs humains. Dreyfus cite un passage de l'article " Expérience and Perception of Pattern" de Michael Polanyi pour montrer l'importance de cette capacité de l'esprit dans la résolution de problèmes:

> This power resides in the area which tends to function as a background because it extends indeterminately around the central object of our attention. Seen thus from the corner of our eyes, or remembered at the back of our mind, this area

---

219 H. L. Dreyfus (1979), *op.cit.* , pp.105-106.

220 Un protocole est l'énoncé des règles d'un raisonnement ou des démarches de pensée suivies lors de la résolution d'un problème par un être humain. En IA les protocoles sont établis à partir des rapports verbaux des experts dans un domaine donné (par exemple, le jeu d'échec) sur la façon dont ils resolvent un problème.

> compellingly affects the way we see the object on which we are focusing. We may indeed go so far as to say that we are aware of this subsidiarily noticed area mainly in the appearance of the object to which we are attending[221].

Selon Dreyfus notre attitude naturelle à résoudre des problèmes et la plupart des processus qui donnent naissance a des comportements intelligents sont en rapport avec notre *conscience périphérique,* laquelle nous permet de voir le monde, ou mieux, les situations, comme des structures organisées. Notre *conscience périphérique* nous permet d'avoir ce que les psychologues de la forme ont appelé la *Gestalt* d'une la situation[222].

3- Les présuppositions qui soutiennent l'optimisme en Intelligence Artificielle

Dreyfus explique quelles sont d'après lui les raisons qui font que les recherches en IA continuent à être faites, en dépit de leurs échecs continus. Ces raisons résultent des *présuppositions* (*assomptions*)[223] d'ordre biologique, psychologique, épistémologique et ontologique. La discussion sur les présuppositions sous-jacentes aux travaux sur les machines intelligentes constitue une réflexion philosophique où l'auteur procède à une critique épistémologique des idées élémentaires qui constituent la base de l'IA et de la recherche sur la simulation cognitive.

Selon Dreyfus il est possible d'évaluer la signification des résultats obtenus dans ces deux domaines par l'intermédiaire d'une critique des présuppositions mentionnées. L'intérêt de Dreyfus est de mettre en évidence les limites du projet pour la conception des machines intelligentes. La critique des présuppositions intéresse également les scientifiques et les philosophes.

Les présuppositions de l'IA, discutées par Dreyfus, ont comme point de départ l'idée que les processus cognitifs sont des processus de traitement d'information. Elles sont liées à

---

221 Polanyi, M. , " Experience and Perception" in the modeling of Mind, p.214, cité par H. L. Dreyfus (1979), *op.cit.* , p.103.

222 Le modèle Gestaltiste de la perception est très compatible avec l'analyse d'inspiration phénoménologique de Dreyfus. Le gestaltisme que explique la perception humaine comme étant un phénomène structuré, c'est-à-dire, qui dépend d'un ensemble indissociable de facteurs interliés où chaque élément de la constitution du phénomène de la perception dépend de la structure perceptuelle dans laquelle il se trouve.

223 Nous avons traduit le mot assumption par le terme présupposition (supposition préalable). Notre choix prend en considération l'asseption du mot assumption dans le texte originel. Il faut mentioner, toutefois, que les traducteurs français de *what Computers Can't Do* ont optée plutôt par le terme "postulat "pour traduire le même mot. Il nous semble que le terme postulat choise dans la traduction française est très fort, considérée l'état du dévéllopement dans lequel se situé l'IA et le statut épistémologique conféré par Dreyfus à cette recherche. Em plus, nous avons choisi un terme plus faible (présupposition)pour traduire assumption, étant donné que ceux qui travaillent dans le domaine de l'IA n'affirment pas (ou postulent) tels ou tels principes. Les postulats biologique, psychologique, epistemologique et ontologique ne sont pour Dreyfus que de simples préssupositions (assumptions).

l'analogie cerveau /esprit-machine selon laquelle le cerveau et les ordinateurs sont des systèmes qui traitent l'information selon les principes formels semblables.Cela est corroboré par des thèses qui considèrent l'esprit comme une sorte de système de traitement de l'information.

Dans son analyse des présuppositions *biologiques, psychologiques, épistémologiques et ontologiques*, Dreyfus prend comme cible les recherches de l'approche formelle *descendante* Il est intéressant de mentionner que les présuppositions ont toutes un caractère anthropomorphique, car les chercheurs utilisent comme base théorique de leurs expériences des concepts qui sont en rapport avec des modèles d'explication de l'esprit, du cerveau et enfin de toute la cognition humaine.

3.1- La présupposition biologique

La présupposition biologique consiste à supposer qu'il y a des niveaux d'interaction neurophysiologique où le cerveau fonctionne de façon similaire aux processus digitaux d'un ordinateur. Les neurones seraient l'équivalent biologique des circuits d'un ordinateur et vice versa. Si le cerveau nous permet de penser, alors une autre structure biologique est aussi capable de produire des résultats, sinon semblables, du moins très proches des résultats obtenues par le cerveau.

> Man is an object. The success of modern physical science has assured us that a *complete description* of the behavior of a physical object can be expressed in precise laws, which in turn can serve as instruction to a computer which can then, at least in principle, simulate this behavior. This leads to the idea of a neurophysiological description of human behavior in terms of inputs of energy, physical-chemical transaction in the brain, and outputs in terms of motions of the physical body, all, in principle, simulatable on a digital machine[224].

Le présupposé biologique est basée sur des thèses physicalistes, lesquelles ne sont pas corroborées par les recherches en neurophysiologie. Avoir une présupposition d'ordre biologique, c'est présumer que les caractéristiques logiques du cerveau dépendent des caractéristiques physiques de celui-ci. Selon cette perspective, le cerveau et l'ordinateur fonctionnerait de façon séquentielle[225].

Le présupposé biologique défend un modèle digital du cerveau selon lequel celui-ci fonctionnerait comme un système physique binaire capable de manipuler des symboles. À

224 H. L. Dreyfus (1979), *op.cit.* , p. 177.
225 H. L. Dreyfus (1979), *op.cit.* , p.160.

un niveau quelconque, le cerveau doit traiter de l'information. De cette façon, il y aurait des processus biologiques qui fonctionneraient comme les bascules (*flip-flops*) d'un ordinateur. Le modèle digital du présupposé biologique peut être expliqué ainsi: les commutateurs "on-off" des ordinateurs seraient comparables à des processus physiques qui ont lieu dans le cerveau lors des processus neuronaux.

Selon Dreyfus, le modèle digital du cerveau n'est pas confirmé par les travaux scientifiques: "This model is still uncritically accepted by practically everyone not directly involved with work in neurophysiology, and underlies the naïve assumption that man is a walking example of a successful digital computer program"[226].

Considérant les recherches en neurophysiologie, Dreyfus affirme que le principe de fonctionnement du cerveau est basé sur un parallélisme massif, tandis que l'ordinateur opère, en général, séquentiellement. Il admet qu'il est toujours possible de simuler le parallélisme du cerveau, mais les ordinateurs conventionnels dont l'architecture est séquentielle ne peuvent pas simuler des comportements qui exigent des degrés de parallélisme plus accentués[227].

Les capacités du cerveau dépassent aujourd'hui énormément celles des machines digitales programmables. Selon Dreyfus même la simulation des quelques parties du réseaux neuronal dans un programme d'ordinateur ne constituerait pas une garantie que le cerveau fonctionnerait comme un ordinateur, et vice versa. Nous avons une connaissance encore très sommaire du cerveau humain.

Nous n'avons pas de garantie que les principes de fonctionnement et de stockage d'information du cerveau et des machines digitales soient similaires. En plus, les capacités générales de ces deux systèmes sont très différentes. La présupposition biologique peut être remis en question par quelques résultats empiriques fournis par des études neurophysiologiques récentes.

Les expériences en neurophysiologie, dit Dreyfus, suggèrent que le modèle digital n'est pas valable. Les arguments qu'il présente pour montrer les faiblesses de la présupposition biologique sont les suivants:

Le cerveau ne fonctionne pas selon des principes numériques binaires. Le cerveau ne traite pas non plus des fragments d'information de façon séquentielle au moyen de signaux

---

226 H. L. Dreyfus (1979), *op.cit.* , p.159.

227 Dreyfus ne écarte pas la possibilité théorique que de nouvelles architectures neo-connexinnistes soyent capables de permetre des performances importantes par exemple dans le champ de l'aprendissage. Cela peut être possible, selon lui, dès que ces architectures soient capables de sourmonter les dificultés deja affrontés par les systèmes à programmes algorithimiques et heuristiques. Cf. H. L. Dreyfus (1979), *op.cit.* , p. 326, note 1.

à valeurs discrètes, mais au contraire tout porte à croire que le cerveau fonctionne selon des processus "globaux"[228] opérant par des rafales d'impulsions (*volleys of pulses*) électriques non séquentielles. Dreyfus argumente que la tendance des études sur le fonctionnement neuronal porte à croire croire que le cerveau fonctionnerait de façon analogique. Il fait une distinction entre traitement de l'information numérique et traitement analogique dans les termes suivants:

> The essential difference between digital and analogue information processing is that in digital processing a single element represents a symbol in a descriptive language, that is, carries a specific bit of information; while in a device functioning as an analogue computer, continuous physical variables represent the information being processed[229].

Dreyfus montre que le modèle digital n'est pas valable. Le traitement de l'information sous la forme d'impulsions électriques au niveau neurologique dépend de plusieurs facteurs tels que la variation du diamètre de l'axone qui est en rapport avec (a) le laps de temps depuis la dernière impulsion électrique passée dans le neurones et (b) des effets des axones à la proximité:

> The filter characteristics of the axon would vary with its diameter which in turn might be a function of the recency of signals passing down that axon, or even, perharps, of the activation of immediately environing axons. If such time factors and field interaction play a crucial role, there is no reason to hope that the information processing on the neurophysiological level can be described in a digital formalism or, indeed, in any formalism at all[230].

En resumé pour Dreyfus les processus de traitement d'information dans le cerveau résultent d'un processus global simultané et *interactif*. Les processus du cerveau ne peuvent pas être réduits à des éléments discrets ou *bits* d'information. Pour cette raison, tels processus résistent à toute forme de formalisation visant à les programmer par des moyens heuristiques sur des machines séquentielles.

Le cerveau n'est point, selon l'auteur, un système qui manipule physiquement des symboles. Dreyfus considère la présupposition biologique comme une thèse non

---

228 Dreyfus dit que les recherches des annés 1950 en neurophysiologie montraient déjà le caractère global du fonctionement du cerveau. Il cite le neurophysiologiste Theodore H. Bullock dont le travail sur le potentiel présynaptique et postsynaptique des systèmes neuronaux montre que les impulsions présynaptiques fonctionnent selon des processus analogiques plutôt que selon des processus digitaux ou séquentiels.

229 H. L. Dreyfus (1979), *op.cit.* , p.161.

230 H. L. Dreyfus (1979), *op.cit.* , p. 162.

corroborée par les expériences scientifiques. Nous n'avons pas, jusqu'à maintenant, de bases empiriques cohérentes pour défendre l'idée selon laquelle le cerveau et l'ordinateur digital fonctionnent de façon semblables sur le plan physique.

3.2- La présupposition psychologique

Le présupposé psychologique prend pour point de départ la thèse que l'*esprit* opère en termes binaires, qu'il traite des "bits" d'informations selon des règles formelles pré-établies à la façon d'un ordinateur digital. Cette idée permet à ceux qui travaillent en simulation cognitive[231] (la plupart, des psychologues) de voir l'ordinateur comme un modèle pour l'étude de l'esprit. Sur le plan de leur fonctionnement, l'esprit et l'ordinateur seraient des machines universelles de traitement de symboles. Cela est en rapport avec la tradition représentationnaliste selon laquelle penser équivaut à calculer:

> A psychological assumption that the mind can be viewed as a device operating on bits of information according to formal rules. Thus, in psychology, the computer serves as a model of the mind as conceived of by empiricists such as Hume (with the bits as atomic impressions) and idealists such as Kant (whith the program providing the rules). Both empiricists and idealists have prepared the ground for this model of thinking as data processing—a third-person process in which the involvement of the "processor" plays no essential role [232].

L'expression la plus évidente de la présupposition psychologique peut être trouvée dans les thèses de l'IA qui affirment qu'une bonne partie de ce qui compte comme étant du comportement intelligent s'organise en termes de structure cognitive, laquelle peut être comprise et décrite par des moyens formels[233]. En faisant la description d'un réseau d'entités symboliques qui composent nos pensées, nous pouvons établir des analogies entre les opérations logiques de notre pensée et la manipulation symbolique effectuées par des machines digitales.

Dreyfus affirme que ceux qui travaillent en simulation cognitive ne veulent pas réduire l'esprit, les intentions, les perceptions, les souvenirs, à des descriptions d'ordre

---

231 Dreyfus signale en passant la distinction entre l'IA et le projet de simulation cognitive. Il faut mentionner que la distinction entre ces deux domaines n'est pas claire dans la plupart des travaux de l'IA. Car la simulation cognitive est à peine un secteur de l'IA, elle se distingue de l'IA mais, utilise le même vocabulaire technique et traite des mêmes problèmes. La simulation cognitive se distingue par le fait qu'elle s'intéresse principalement aux processus cognitifs sous-jacents à l'activité de résolution des problèmes. Sont but est de simuler le comportement humain au moyen des techniques informatiques. Elle a un caractère fortement anthropomorphique, au contraire de l'IA.

232 H. L. Dreyfus (1979), *op.cit.* , p. 156.

233 H. L. Dreyfus (1979), *op.cit.* , p. 53.

neurophysiologique, par exemple, à des processus physico-chimiques à l'intérieur du cerveau.

Il rappelle que l'esprit, les intentions, etc., peuvent être considérés aussi à partir d'un autre niveau d'analyse, le niveau phénoménologique:

> There is of course, another level—let us call it phenomenological—on which it does make sense to talk of human agents, acting, perceiving objects, and so forth. On this level what one sees are tables, chairs, and other people, what one hears are sounds and sometimes words and sentences, and what one performs are meaningful action in a context already charged with meaning[234].

Cependant, ceux qui travaillent dans le domaine de l'IA et en Simulation Cognitive ne s'intéressent pas à des analyses de type phénoménologique qui considèrent les choses telles que perçues par les individus, et où il n'est pas question de faire appel à des représentations formelles.

La présupposition psychologique est tributaire des idées liées au développement de la chimie, à savoir, que nous avons la possibilité d'étudier les phénomènes naturels à des niveaux intermédiaires de description.

La présupposition psychologique est sous-jacente aux théories fonctionnalistes selon lesquelles nous pouvons comprendre les comportements intelligents à partir d'une analyse à un niveau intermédiaire, entre les états mentaux et les phénomènes matériels. Dreyfus affirme que les travaux qui défendent un niveau intermédiaire d'analyse, (qu'il a appele de niveau de *traitement de l'information*) sont en rapport avec la théorie de la communication, la cybernétique et l'informatique[235].

Dreyfus s'oppose à la théorie computationnelle de l'esprit, selon lui, la pensée humaine n'opère pas selon un mode binaire et à partir de règles pré-établies semblables à celles qui caractériseraient le fonctionement d'un ordinateur:

> Much of the literature of Cognitve Simulation gains its plausibility by shifting between the ordinary use of the term "information" and the special technical sense the term has recently acquired. Philosophical clarity demands that we do not foreclose the basic question whether human intelligence presupposes rule like operations on discrete elements before we begin our analysis. Thus we must be careful to speak and think of "information processing" in quotation marks when referring to human beings[236].

---

234 H. L. Dreyfus (1979), *op.cit.* , p.177-178.

235 Dreyfus affirme que lorsque le behaviorisme a épuisé ses possibilités théoriques et a été considéré une théorie réductrice et déterministe du comportement humain, le psychologues (fonctionnalistes) furent attirée par les possibilités théoriques d'analyse de l'esprit offertes par les sciences informatiques et théories sur l'information.

236 H. L. Dreyfus (1979), *op.cit.* , p.166.

Dreyfus montre l'inefficacité du modèle computationnel en tenant compte de l'ambiguïté de la notion de traitement de l'information. L'esprit humain ne traite pas l'information dans le sens informatique ou mieux dans le sens de la théorie de la communication[237], car l'esprit possède des propriétés sémantiques lesquelles n'ont rien à voir avec la notion d'information tel qu'employée dans ces domaines.

Les idées défendues en IA et en Simulation cognitive font une utilisation sémantique du terme *information*. Cela n'a rien à voir avec le sens que lui donnent Shannon et Weaver[238]. La *Théorie de l'information* attribue à ce terme un sens formel non sémantique. A ce propos, affirme Dreyfus: "It is a nonsemantic mathematical theory of the capacity of communication channels to transmit data. A bit (binary digit) of information tells the receiver which of two equally probable alternatives has been chosen"[239]. Dreyfus renforce son argument en ajoutant ce passage de W. Weaver: "The word *information* in this theory, is used in a special sense that must not be confused with its ordinary usage. In particular, *information* must not be confused with meaning"[240].

Selon Dreyfus il y a, de la part des ceux qui travaillent en science cognitive, un abus de langage.

Les critiques de la présupposition psychologique sont sans doute dirigées contre la psychologie cognitive d'orientation fonctionnaliste. Ces psychologues cherchaient à exploiter les possibilités que les ordinateurs offraient à la fondation d'une psychologie capable d'être digne du titre de science du comportement. La présupposition psychologique part de l'idée fonctionnaliste et mentaliste selon laquelle le traitement de l'information par l'esprit résulte d'une série d'opérations discrètes.

Dreyfus critique le caractère axiomatique des arguments de cognitivistes tels que Newell, Neisser et Miller. Ces auteurs analysent et défendent un modèle computationnel

---

237 L'informatique et l'IA empruntent le terme "information " à la théorie de la communication dont le but est de définir et d'étudier quantitativement l'information, leur codage, et la capacité de traitement dans les canaux de transmission d'informations. Les théories sur l'information(l'information ici dans le sens que nous venons de mentionner) sont en rapport avec des problèmes d'engéniérie dans le domaine des télécomunications ou de l'informatique. La notion de traitement de l'information issue de la théorie de la communication et en dernière analyse de l'informatique, désigne les moyens formels par lesquels l'information est organisée en système. Comment les signaux transmis, indépendant de leurs contenus, reçoivent une forme.

238 Cf. Shannon C. E. et W. Weaver, *The MathematicalTheory of Communication*, University of Illinois Press, Urbana,1949.

239 H. L. Dreyfus (1979), *op.cit.* , p.165.

240 *Idem*. La citation en question est tiré de W. Weaver,"Recent Contributions to the Mathematical Theory of Communicattion" *in* C. E. Shannon et W. Weaver, 1949 *op.cit.* , p.99. Le terme *information* , duquel nous parle Weaver a une aception essentiellement formel. Il faut mentioner que pour les théoriciens de l'information, par exemple, deux messages ayant deux sens distincts peuvent être formelment équivalents. En effet pour la théorie de l'information, c'est le contenant, c'est à dire le signal transmis, qui compte et non le contenu (le message que le signal transmet).

qui permet de décrire le comportement intelligent en termes des programmes d'ordinateur. S'il est possible de faire des descriptions détaillées et complètes qui caractérisent et expliquent les comportements humains, alors il est possible de composer un ensemble d'instructions bien définies capables de reproduire sur une machine certains de ces comportements.

Selon la méthode fonctionnaliste employée par quelques uns des auteurs ci-dessus, les processus cognitifs sont décrits et expliqués comme un ensemble coordonné de règles visant un but déterminé. D'accord avec Dreyfus, les difficultés théoriques de ce genre de procédure sont omises en faveur de quelques définitions axiomatiques peu défendables, telles que celles qui affirment que les comportements humains peuvent être décrits sous la forme d'un ensemble fini de règles ou jeu d'instructions pouvant d'être traitées par une machine.

Pour Dreyfus, comprendre un comportement donné n'est pas l'analyser par le moyen de protocoles et concevoir des programmes. Les explications en psychologie ne peuvent pas prendre la forme d'un programme ou *jeu d'instructions,* tel que le veulent ceux qui se sont basés sur la présupposition psychologique. Selon lui, ce modèle d'explication appliqué quelquefois préalablement en psychologie cognitive ne rend pas compt des contradictions et ambiguïtés qui apparaissent entre les protocoles d'une description psychologique d'un comportement et le comportement lui-même[241].

La présupposition psychologique est exprimée autrement par l'affirmation que le comportement intelligent peut être complètement expliqué parce que les êtres humains ne font que suivre des règles, lesquelles sont d'une manière ou d'une autre représentées dans leurs têtes.

Selon Dreyfus, les explications psychologiques du comportement humain qui admettent des descriptions exprimées sous la forme de programme d'ordinateur ont un rapport avec des formes de pensée enracinées dans la philosophie occidentale.

Comme nous l'avons déjà mentionné, Dreyfus note que, depuis Platon et Kant, l'expérience, la perception et le comportement rationnel ont été toujours analysés en termes de règles. Cela est lié, rappelons-nous, à la notion d'esprit conçu comme un mécanisme de calcul et à l'idée qu'il est possible de représenter formellement l'ensemble des instructions ou règles de la pensée.

---

241 Cf. H. L. Dreyfus (1979), *op.cit.* , pp.175-176.

> Thus, for Plato, a theory of human behavior which allows us to *understand what* a certain segment of that behavior accomplishes is also an *explanation* of *how* that behavior is produced. Given this notion of understanding and this identification of understanding and explanation, one is bound to arrive at the cognitive simulationists with their assumption that it is self-evident that a complete description of behavior is a precise set of instructions for a digital computer, and that these rules can actually be used to program computers to procuce the behavior in question[242]

Selon Dreyfus les cognitivistes font une confusion entre la compréhension d'un comportement intelligent et la description de ce même comportement. Il y a un glissement de la notion de *compréhension* à l'idée de *description*. Lorsqu'il s'agit de comprendre le comportement humain, nous ne devons pas nous fier totalement à des descriptions. La notion de *description* (ensemble de règles explicites) employée par les chercheurs cognitivistes, est en rapport plutôt avec la notion d'*explication*..

L'analogie fonctionnaliste entre les ordinateurs et l'esprit considère que l'esprit est un dispositif de manipulation de symboles. Par exemple, dans les travaux de J. Fodor (1968) et de Miller, Galanter et Priban (1960) nous trouvons quelques analyses de caractère fonctionnaliste qui suggèrent quelquefois des analogies entre le système cognitif et les systèmes informatiques. Voyons par exemple celle qui suggère que les êtres vivants procèdent de façon séquentielle, "When an *organism executes* a plan *he* proceeds step by step, completing one part and them moving to the next"[243].

Des affirmations comme celle-ci conduisent selon Dreyfus à des incohérences au niveau théorique. Dans le passage mentionnés, il semble que l'organisme fonctionne exactement comme un ordinateur séquentiel. Cependant l'analogie dans ce cas n'a aucun sens car ce sont les machines qui procèdent "étape par étape". Si dans certains cas notre organisme fonctionne de cette façon, cela ne justifie pas la généralisation de certains aspects biologiques de l'organisme à l'ensemble des processus cognitifs ou si on veut à la pensée. Sur ce point, l'analyse de Dreyfus est très fine:

> Here all three levels exist in unstable and ungrammatical suspension. "When an *organism* [biological] *executes* [machine analogy borrowed from human agent] a

---

242H.L. Dreyfus, (1979), *op.cit.* , p.177. L'idée fonctionnaliste qui permet de rendre possible l'*explication* du comportement en termes d'instructions ou de programmes d'ordinateur vise créer un modèle explicatif de l'esprit, qui est différent d'une orientation comme celle proposée par Dreyfus à la fin de *What Computers Can't Do* dont l'objetif est plutôt de comprendre l'esprit. Les mots "comprendre" et "explication" en italiques dans cette citation, veulent à notre avis mettre l'accent sur la distinction entre expliquer et comprendre toujours discutée par les philosophes contemporains. Dreyfus exprime ici la difficulté d'appliquer au comportement et à l'esprit humain des méthodes d'analyse causale inspirées de l'explication de la nature.

243 Cf. M. Galanter et Pribram, *Plan and the Structure of behavior*, Holt, Rinehart et Winston, N. Y. ,1960, p.57. Cité par H. L. Dreyfus (1979), *op.cit.* , p.179.

> Plan *he* [the human agent] ..." Or, one can have it the other way around and instead of the organism being personified, one can find the mind mechanized[244].

Certains auteurs, comme J. Fodor, nettement attachés à la présupposition psychologique, semblent croire que le fonctionnement de l'esprit est lié à des procédures effectives ou algorithmes. Dreyfus note que l'utilisation par quelques auteurs de certaines expressions, telles que "exécution mentale" et "exécution mentale de tâches", "entrée sensorielle" "information-stimulus", etc, suggèrent, que l'esprit humain fonctionne comme une machine digitale.

Toutes ces expressions sont, selon Dreyfus, problématiques car elles sont basées sur un aspect du présupposé psychologique, à savoir, qu'il y a un niveau de description du fonctionnement de l'esprit, ou niveau de traitement de l'information, où l'esprit pourrait être considéré comme un dispositif manipulant des symboles.

Fodor et les chercheurs cognitivistes d'orientation fonctionnaliste comme Neisser, Miller et d'autres, mentionnés par Dreyfus, soutiennent qu'il est possible de rendre compte de la perception à partir des descriptions de niveau intermédiaire. Il est possible selon ces auteurs de formaliser en termes de règles les concepts que nous avons d'un objet.

Fodor et les autres auteurs mentionnés pensent que la perception dépend uniquement de l'intégration des *entrées sensorielles*[245], comme si ce phénomène impliquait tout simplement un traitement de données à un niveau intermédiaire entre le niveau neurologique et un niveau phénoménologique. Dreyfus ne croit pas que cette aproche puisse rendre compte de la perception.

Dreyfus souligne que la notion d'*entrée sensorielle* fonctionnerait comme une sorte d'axiome dans les travaux de Fodor, de Miller, Galenter, Priban et Neisser et d'autres chercheurs cognitivistes. Une telle notion n'est pas entièrement justifiée, mais fonctionne comme un élément important de leurs théories:

> Of course, if we begged the question and assumed that the brain is a digital computer, then sense could be made of the notion that a concept is a formal structure for organizing data. In that case the "sensory input" would be neither a percept nor a pattern of energy, but a series of bits, and the concept would be a set of instructions for relating these bits to other bits already received, and classifying the result. This would amount to an hypothesis that human behavior can be undertood on the model of a digital computer. It would require a theory of

---

244 *Idem.*

245 La notion d'*entrée sensorielle* oscille entre le sens phénoménologique (entrée sensorielle comme un visage , une forme, etc.) et le sens physique (entrée sensorielle comme forme spécifique d'énergie perçue par un organe sensoriel.) Les cognitivistes considerent que l'entrée sensorielle peut être analysée en termes d'éléments purement formels comme des *bits d'information.*

> just what theses bits are and would then have to be evaluated on the basis of empirical evidence.
>
> But for Fodor, as for Miller et al. , the notion of "sensory input" and of a concept as a rule for organizing this input seems to need no justication but rather to be contained in the very notion of a psychological explanation[246].

Selon Dreyfus l'esprit ne fonctionne pas à la manière d'un ordinateur digital, ou selon les principes formels d'une machine de Turing comme semble le prétendre Fodor dans ce passage:

> Les sciences de la connaissance ont fortement tendance à traiter l'esprit essentiellement comme un dispositif manipulant des symbloles. Si l'on peut définir fonctionnellement un processus mental comme une opération sur des symboles, il existe une machine de Turing pouvant effectuer la programmation et un grand nombre de mécanismes pour la réalisation de cette machine de Turing. (...) [247].

Dreyfus s'oppose à tout modèle formel ou si on veut computationnel de compréhension de l'esprit; pour lui les formalismes ne peuvent pas rendre compte de plusieurs aspects importants liés à l'intelligence et à la perception humaine. Il n'y a pas d'équivalence fonctionnel entre l'esprit et l'ordinateur.

Dreyfus s'attaque aussi au modèle fonctionnaliste (appliqué par les psychologues et cognitivistes) de compréhension de l'esprit. Selon lui, il n'est pas possible, par l'emploi de formalismes, d'élaborer des théories psychologiques capables de rendre compte du caractère global du fonctionnement de l'esprit.

Pour lui, l'esprit n'est pas un système de traitement de l'information. L'expression "traitement de l'information" doit être, selon Dreyfus, mise entre guillemets quand il s'agit de l'intelligence naturelle. Celle-ci n'opère pas, fort probablement, sur des éléments atomiques et selon des règles très strictes.

Dreyfus ne croit pas que l'esprit fonctionne à partir de représentations; il insiste dans sa critique de l'IA et de la recherche en simulation cognitive sur des exemples où la perception joue un rôle important.

Il a une perspective phénoménologique de l'esprit. où il n'est pas question pour comprendre l'esprit de l'analyser en termes représentationnels. L'esprit ne résulte pas d'un ensemble de règles lui permettant d'accomplir une suite d'opérations inconscientes à la manière d'un ordinateur. L'erreur majeure des ceux qui travaillent en simulation cognitive

---

246 H. L. Dreyfus (1979), *op.cit.* , p.183.
247 J. Fodor, *op.cit.* , p.84.

est d'affirmer, selon Dreyfus, que les règles utilisées pour la description et la formalisation des comportements intelligents sont les mêmes règles qui produisent ces comportements.

Dreyfus semble vouloir montrer que les processus mentaux humains ne sont pas résultants de la manipulation de symboles qui représentent le monde à la façon des modèles informatiques. Dreyfus comprend l'esprit d'un point de vue phénoménologique. Ainsi, selon lui il n'est pas question de faire appel à des représentations pour les comprendre.

L'analyse phénoménologique de Dreyfus considère l'esprit comme relié fondamentalement à un *background* non représentationnel.

3.3- La présupposition épistémologique

Dans l'analyse de la présupposition psychologique Dreyfus se concentre sur les recherches en Simulation Cognitive. L'analyse de la présupposition épistémologique concerne plutôt le statut des recherches en IA. Dreyfus affirme que ce n'est plus la recherche d'un niveau intermédiaire d'analyse des aptitudes mentales qui entre en ligne de compte comme base de l'optimisme persistant dans le domaine de l'IA. L'optimisme tend maintenant à être soutenu par des arguments d'ordre épistémologiques, basés sur le succès obtenu au niveau des sciences physiques et de la linguistique de N. Chomsky.

Au contraire de la présupposition psychologique qui soutient que le comportement intelligent peut être complètement *expliqué* par des règles heuristiques, la présupposition épistémologique est exprimée par l'affirmation que le comportement intelligent peut être complètement *formalisé* au moyen de règles heuristiques. Les travaux en *Simulation Cognitive* sont soutenus par une présupposition de type psychologique. Dans ces travaux les règles qui servent à exprimer formellement un comportement sont les mêmes règles qui produisent ces comportement. Tandis que les travaux en IA sont soutenus par une présupposition de type épistémologique, dans ces travaux toute conduite délibérée est analysée et exprimée sur la forme des règles formelles et ces règles peuvent servir à reproduire ces mêmes conduites sur un ordinateur digital.

Dreyfus souligne qu'il est difficile de distinguer la présupposition psychologique de la présupposition épistémologique et nous explique que cette différence subtile apparait cependant de façon claire lorsqu'on considère le sens du terme "reproduire", tel qu'il est employé dans l'analyse de Dreyfus.

L'auteur affirme que la présupposition épistémologique est plus modérée et pour cette raison, moins vulnérable que la présupposition psychologique. Pour les chercheurs en IA

(qui défendent la présupposition épistémologique), tout comportement peut être formalisé selon certaines règles, lesquels peuvent être utilisées pour reproduire le comportement humain sur une machine.

Les chercheurs en Simulation cognitive, par contre, (en se basant sur la présupposition psychologique) utiliseraient plutôt un sens fort du terme *reproduire*. Pour eux l'ordinateur (au moyen de règles formelles) serait capable d'expliquer et en même temps reproduire le comportement humain. Autrement dit, c'est la présupposition psychologique qui leurs permet d'affirmer que les règles qui servent à exprimer formellement un comportement sont exactement les mêmes règles qui produisent ce comportement.

La présupposition épistémologique considère possible la formalisation du comportement humain en termes de théorie de la *compétence*. Mais elle n'arrive pas à rendre compte, sur les plan formel, de l'*accomplissement* (*performance*) cognitif humain et dans le domaine de l'IA.

> The epistemological assumption is weaker and thus less vulnerable than the psychological assumption. But it is vulnerable nonetheless. Those who fall back on the epistemological assumption have realized that their formalism, as a theory of competence, need not be a theory of ***human*** performance, but they have not freed temselves sufficently from Plato to see that a theory of competence may not be adequate as a theory of ***machine*** performance either[248].

La présupposition épistémologique est exprimée par deux affirmations:

1. tout comportement A non arbitraire[249] peut être formalisé,
2. la formalisation obtenue permet de reproduire A

Le succès de la physique moderne et celui de la linguistique Contemporaine ont beaucoup influencé les recherches sur l'intelligence artificielle. Dreyfus critique la présupposition épistémologique montrant que, la proposition 1 est une généralisation non justifiée, qui est basée sur les méthodes des sciences physiques. La proposition 2, inspirée des succès de la linguistique contemporaine, dénote une malheureuse incapacité à expliquer, par une théorie de la compétence, la performance cognitive des êtres humains.

Les arguments en faveur de la présupposition psychologique sont basés sur les Sciences physiques.

---

248 H. L. Dreyfus (1979), *op. cit.* , p. 190.
249 C'est à dire tout comportement dirigé à un but precis lorque nous agissons selon des règles.

La présupposition épistémologique ainsi que la présupposition psychologique sont en rapport avec une idée platonicienne chère à la tradition représentationnaliste selon laquelle comprendre une chose, c'est l'exprimer formellement de façon claire et objective. Nous trouvons cette idée dans l'explication du mouvement au sein des sciences physiques: "Galileo was able to found modern physics by abstracting from many of the properties and relations of aristotelian physics and finding that the mathematical relations which remained were sufficient to describe the motion of objects"[250].

En physique on utilise des représentations ou des expressions formelles pour décrire le comportement de certains corps ou le mouvement des planètes. Ces formalisations nous permettent de formuler des lois sur le mouvement. La discussion de Dreyfus sur la présupposition épistémologique vise à discuter les limites du modèle formel de la physique appliqué à l'analyse du comportement humain. Dreyfus se demande s'il est possible de formuler des lois du comportement analogues aux lois de la physique.

La notion de *lois du comportement* est liée a l'idée, commune en science physique, selon laquelle tous les objets du monde matériel obéissent aux lois de la nature. Telles lois peuvent être exprimées par des formalisations. Étant donné que le corps et le cerveau sont des objets matériaux, alors il est possible, par des explications de tout ce qui se passe sur le plan physique, de formuler des lois sur le comportement humain en tant que produit du mouvement du corps et des processus qui ont lieu dans le cerveau.

3.3.1- Les arguments en faveur de la présuposition épistémologique basés sur les sciences physiques

Dreyfus affirme que ce modèle formel des sciences physiques ne s'applique pas à l'IA. Il ne partage pas l'idée que nous pouvons, par des moyens formels, simuler complètement le comportement humain sur des machines. Cette hypothèse, défendue par les chercheurs en IA, est basée sur l'idée qu'en tant que système physique notre comportement est sujet à des règles, car tout comportement non arbitraire est le résultat de règles suivies par un sujet ou un objet. Selon eux les *règles suivies* par un sujet servent à expliquer son comportement.

Selon Dreyfus il y a un équivoque dans cette façon de considérer les choses. Il exprime cela dans le passage suivant:

---

250 H. L. Dreyfus (1979), *op. cit.* , p. 197.

> Consider the planets. They are not solving differential equations as they swing around the sun. They are not *following* any rules at all; but their behavior is nonetheless lawful, and to understand their behavior we find a formalism—in this case differential equations—which expresses their behavior as motion according to a rule. (...)[251].

Nous ne pouvons pas confondre les règles qu'on suit, par exemple quand on conduit une bicyclette, avec les règles qui permettent de décrire l'action que nous accomplissons ou qui décrivent comment les choses se passent quand nous conduisons une bicyclette. Selon Dreyfus *suivre une règle* est différent d'*agir selon de règles*.

> A man riding a bicycle may be keeping his balance just by shifting his weight to compensate for his tendency to fall. The intelligible content of what he is doing, however, might be expressed according to the rule: wind along a series of curves, the curvature of which is inversely proportional to the square of the velocity. The bicycle rider is certainly not following this rule consciously, and there is no reason to suppose he is following it unconsciously.(...) [252].

Selon Dreyfus, les comportements humains ne sont pas complètement formalisables. Il s'attaque d'abord à l'affirmation selon laquelle on peut exprimer formellement le comportement intelligent non arbitraire et ainsi concevoir des programmes informatiques en IA. Pour cerner sa critique du modèle de comportement provenant des sciences physiques, Dreyfus fait ressortir la notion de "machine" telle que conçu par Alan Turing, laquelle est utilisée dans la littérature sur l'IA, Pour Turing, une machine est un système formel abstrait qui opère à partir de règles strictes. Cette définition vaut aussi pour les ordinateurs digitaux:

> A digital computer is a machine which operates according to the sort of criteria Plato once assumed could be used to understand any orderly behavior. This machine, as defined by Minsky, who bases his definition on that of Turing, is a "rule-obeying mechanism" (...)It operates on determinate, unambiguous bits of data, according to strict rules which apply unequivocally to these data. The claim is made that this sort of machine—a turing machine—which expresses the essence of a digital computer can, in principle, do anything that human beings can do—that it has, in principle, only those limitations shared by man[253].

---

251 H. L. Dreyfus (1979), *op. cit* , p.189.
252 H. L. Dreyfus (1979), *op. cit.* , p.190.
253 H. L. Dreyfus (1979), *op. cit.* , p.192

Minsky dit que l'homme est une machine dont le fonctionnement pourrait être comparé à celui d'un ordinateur. Il fonctionnerait selon un modèle formel de machine de Turing[254]. Dreyfus ne partage pas l'idée défendue par M. Minsky et Turing selon laquelle une machine réelle ou abstraite qui suit des règles strictes est capable d'être programmée et de simuler complètement le comportement humain. Cette hypothèse, défendue par une bonne partie des chercheurs en IA, est basée sur l'idée que l'homme et l'ordinateur ne sont que de systèmes physiques capables de travailler à partir d'une base formelle.

Le fait que nous sommes des systèmes physiques (nous avons un cerveau) n'est pas cependant le point de départ de Minsky et Turing, leur point de départ est l'idée que nous sommes des systèmes capables d'être compris par des moyens formels, car nous avons une base mentale où sont opérées des descriptions symboliques, lesquelles sont susceptibles d'être formalisées et transformées en programmes informatiques.

La position de Dreyfus sur ce sujet est nettement anti-formaliste; pour lui, les formalisations qui valent pour les sciences physiques ne servent pas à l'explication des comportements humains et à comprendre l'esprit. Les formalismes en physique visent à formaliser les mouvements des objets matériels; rien ne prouve que les méthodes des sciences naturelles puissent servir à formaliser le comportement humain. Sur l'effort et l'optimisme des chercheurs en IA Dreyfus affirme: "What would be needed to justify the formalists' optimism would be a Galileo of the mind who by making the right abstractions, could find a formalism which would be sufficient to describe human behavior"[255].

Pour Dreyfus les comportements humains ne sont pas régis par des règles strictes capables d'être transformées en programmes informatiques. Nos comportements résistent à la formalisation car nous ne pouvons pas être réduits à des simples systèmes formels.

Turing lorsque il rejette des idées anti-formalistes comme celles de Dreyfus affirme que si d'un côté nous ne pouvons pas produire un ensemble de règles capables de décrire le comportement humain en toutes les circonstances, il est possible au moins de découvrir un ensemble de lois du comportement permettant de prévoir comment un individu donné se comporterait effectivement face à toutes les situations possibles. Il exprime ce point de vue par la distinction entre "règles de conduite" et "lois du comportement":

> By "rules of conduct" I mean precepts such as "Stop if you see red lights," on which one can act, and of which one can be conscious. By "laws of behavior" I mean laws of nature as applied to a man body such as "if you pinch him he will

254 Cf. Minsky, M. "mather, Mind, and Models" In *Semantic InformationProcessing*. Minsky affime dans ce travail qu'il y a une ressemblance entre les processus mentaux et les programmes d'ordinateur.
255 H. L. Dreyfus (1979), *op. cit.* , p. 197.

> squeak." (...) For we believe that it is not only true that being regulated by laws of behavior implies being some sort of machine (though not necessarily a discrete state machine), but that conversely being such a machine implies being regulated by such laws[256].

Dreyfus ne partage pas l'idée selon laquelle il est possible d'établir des "lois du comportement". Pour lui-même si le comportement humain peut obéir à des lois[257], ces lois ne sont pas représentables sous forme de programme d'ordinateur.

Une description formelle du comportement ne constitue pas un élément suffisant pour la représentation du comportement par des programmes informatiques. Le comportement humain ne peut pas être représenté sous forme d'un ensemble de propositions indépendantes décrivant chaque composant du comportement comme dans un système physique.

Dans son argumentation, Dreyfus admet, par exemple, que l'ordinateur peut traiter et même simuler toutes les données concernant les propriétés physico-chimiques d'une douleur ressentie par une personne. Il soutient cependant que, même si les données emmagasinées sur la douleur sont si complexes nous permettant d'affirmer que la machine simule la douleur (car elle traite toutes les informations concernant la douleur) Ce traitement des informations simulant tous les processus physico-chimiques de la douleur ne concerne que l'aspect formel d'un phénomène physique il laisse de côte plusieurs aspects fondamentaux pour la compréhension de la douleur en tant que phénomène mental.

Si, admettant les méthodes des sciences physiques, nous sommes en mesure d'affirmer que les comportements humains obéissent à des lois et que nous pouvons en principe (théoriquement) les simuler, cela n'est pas une garantie que nous sommes capables de le simuler effectivement.

Le traitement des informations par le cerveau est différent du traitement des informations par les machines. On n'est pas en mesure d'affirmer, comme le font les thèses fortes en IA, que n'importe quelle forme d'information peut être traitée par un ordinateur digital. La simulation d'une douleur par ordinateur (si complète soit-elle) est très éloignée de ce que se passe dans le cerveau ou dans l'esprit humain:

> A digital computer solving the equations describing an analogue information-processing device and thus simulating its *function* is not thereby simulating its "information processing". It is not processing the information which is processed

---

256 A. Turing, "Computing Machinery and Intelligence", *in* A.R. Anderson, éd. , *Minds and Machines*, Prentice-Hall Inc. , NJ, 1964, 114p. p. 23.

257 Car il présente parfois certaines régularités capables de nous permettre de formaliser certains de ces aspectes.

> by the simulated analogue, but *entirely different information* concerning the physical or chemical properties of the analogue. Thus the strong claim that *every form of information* can be processed by a digital computer is misleading. One can only show that for any given type of information a digital computer can in principle be programmed to simulate a device which can process that information[258].

Dreyfus affirme que ceux qui assument la thèse forte en IA s'inspirent du succès de la physique pour affirmer que l'expression formelle du mouvement du corps d'un être humain ou d'un objet permet de comprendre les comportements intelligents et le comportement de l'objet. Pour lui, le fait qu'on peut comprendre le mouvement à partir de certaines règles ne nous permet pas d'affirmer la validité de la présupposition épistémologique.

Les formalismes nous permettent de bien contrôler l'étude de certains processus biologiques liés au comportement humain. Cependant quand il s'agit d'étendre le même formalisme à une analyse globale du comportement, c'est-à-dire essayer de rendre compte de l'esprit ou de la pensée, les choses se compliquent[259].

Selon Dreyfus, nous ne suivons pas toujours des règles lorsque nous agissons ou lorsque nous pensons. Ni les règles analogues aux lois de la physique ni les règles du traitement de l'information ne légitiment la présupposition épistémologique. Les comportements humains ne sont pas complètement formalisables. Pour cette raison nous ne pouvons pas employer des règles servant à manipuler des données (représentations des objets du monde) sur le monde pour les reproduire ou les simuler.

Selon Dreyfus le formalisme est limité pour expliquer les comportements humains. Un programme d'ordinateur ne peut pas reproduire le comportement humain sur des bases purement formelles. Il affirme que l'application des méthodes (des sciences naturelles) inspirés des recherches en IA par les sciences du comportement est une tentative condamnée à l'échec.

3.3.2- L'argument en faveur de la présupposition épistémologique basé sur les succès de la linguistique contemporaine

Depuis Austin, l'acte de discours est considéré comme une source de connaissance sur le comportement humain et sa logique. Suivant la même voie, des auteurs comme Searle et

---

258 H. L. Dreyfus (1979), *op. cit.* , p. 195.

259 Dreyfus affirme que le modèle d'analise du comportement humains issu de la physique ne justifie pas l'extrapolation opérée par les chercheurs en IA, I Ils passent d'un niveau physique d'analyse au niveau mental sans faire la distinction entre ce qui peut être formalisé réellement et sans relever les difficultés les réels dificutés de leurs projets.

Daniel Vanderveken cherchent une réponse sur les relations logiques entre notre esprit et de notre langage[260]. D'un autre côté, Chomsky et ses partisans, suivant la voie formelle de la grammaire transformationnelle mettent l'emphase sur les capacités syntaxiques des usagés d'une langue à produire des actes linguistiques.

Searle et Daniel Vanderveken veulent comprendre le langage, à partir de l'emploi de phrases déterminées par rapport à des contextes déterminés. Leur point commun, c'est qu'ils utilisent soit le formalisme, soit des approches analytiques, afin de comprendre les actes linguistiques en tant que manifestation rationnelle de l'intelligence humaine.

Dans *What Computers Can't Do* Dreyfus s'intéresse surtout à l'influence de la linguistique de Chomsky sur les travaux en intelligence artificielle. Selon lui le succès du modèle de Chomskien représente un encouragement précieux pour ceux qui prennent pour base la présupposition épistémologique et qui soutiennent la possibilité de formaliser les comportements sans les réduire au niveau physique.

Dreyfus discute la validité de la base linguistique de la présupposition épistémologique. Il analyse quelles sont les limitations formelles d'une théorie linguistique incorporée comme modèle pour l'IA.

L'avantage que le formalisme de la linguistique offre par rapport aux modèles basés sur les sciences physiques est qu'il permet de décrire les comportements en termes de règles qui peuvent être traitées immédiatement par la machine sans qu'il soit nécessaire de mesurer des quantités physiques ou de décrire des processus physico-chimiques du comportement en question. Chomsky lui-même admet que son modèle pourrait facilement se prêter à la programmation[261].

Chomsky dans son livre, *Language and Mind,* (1968.) affirme que l'étude formelle de la compétence des usagers d'une langue joue un rôle important dans la compréhension de leur performance[262]. Dreyfus affirme que ceux qui travaillent en IA prennent cette idée au pied de la lettre, oubliant les limitations des théories formelles en ce qui concerne les aspects pragmatiques du langage naturel.Les tenants de la présupposition épistémologique veulent employer le formalisme de la grammaire générative comme moyen de réduction des performances linguistiques à des descriptions calculables. Dreyfus observe, à ce propos qu'ils vont au delà de la théorie de la compétence de Chomsky[263]. Dans leurs recherches

---

260 J.R, Searle et D. Vanderveken, *Foundations of Illocutionary Logic*, Cambridge University Press, Cambridge, 1985.
261 Cf. H. L. Dreyfus (1979), *op. cit.* , p.334.
262 Cette idée est exprimée par Chomsky dans *Cartesian Linguistics*, New York, Harper & Row, 1966, p.75, cité par H. L. Dreyfus (1979), *op. cit.* , p.332.
263 Dreyfus mentionne que les idées transformationnelles de chomsky visent à comprendre, par le biais d'une théorie formelle, la compétence linguistique (syntaxique) et non la performance des usagers d'une langue (sémantique et pragmatique). Autrement dit, pour Chomsky, il s'agit plutôt de comprendre ce qui est nécessaire à l'auditeur-locuteur pour qu'il puisse employer un langage plutôt que de fournir un modèle de conduite linguistique lui permettant d'employer ce langage.

sur les langages naturels, par exemple, c'est la performance et non la compétence qu'ils essayent de traduire en termes de règles grammaticales. Selon Dreyfus cette extrapolation pose des problèmes, car la performance présente une résistance à être formalisée.

Selon Dreyfus la création d'une théorie de la performance implique le développement d'une pragmatique formelle que les chercheurs en IA ne sont pas en mesure de développer. D'ailleurs, Dreyfus soutient qu'on n'a pas de bonnes raisons de croire qu'une théorie formelle pragmatique puisse être créée. S'il y a eu succès de la formalisation de la syntaxe, cela n'implique pas qu'il y aura également réussite dans le domaine de la pragmatique. Pour justifier son argument, Dreyfus avance deux raisons:

La première de ces raisons est que l'élaboration d'une pragmatique formelle exigerait une théorie qui comporte toutes les connaissances humaines sur le monde. Selon Dreyfus, cela semble être impossible.

La deuxième raison est la suivante, il est possible par la simple observation des usagers d'un langage naturel de constater que leur usage d'une langue ne suppose pas toujours l'emploi de règles strictes. La syntaxe n'est pas la seule base structurante du langage. Les usagés d'une langue transgressent souvent les règles syntaxiques et cela n'empêche pas, dans la plupart des cas, que les comportements linguistiques puissent être bien compris. Dans sa totalité le comportement linguistique des êtres humains ne semblent pas se conformer à des règles très strictes.

Dreyfus fait plusieurs remarques sur la capacité qu'ont les êtres humains à désambiguïser des expressions dans une langue en fonction de leurs besoins et du contexte. Au contraire d'une machine qui suit des règles strictes, les êtres humains ont une habileté à manier des cas déviants du bon usage linguistique (tels que les infractions aux règles de la syntaxe), à constater et à corriger immédiatement les problèmes de grammaire et de sémantique prévalant dans la plupart des situations conversationnelles critiques comme celle de l'exemple qui suit:

> There are cases in which a native speaker recognizes that a certain linguistic usage is odd and yet is able to understand it—for example, the phrase "The idea is in the pen" is clear in a sitation in which we are discussing promissing authors; but a machine at this point, with rules for what size physical objects can be in pig pens, play pens, and fountain pens, would not be able to go on. Since an idea is not a physical object, the machine could only deny that it could be in the pen or at best make an arbitrary stab at interpretation. The listener's understanding, on

> the other hand, is far from arbitrary. Knowing what he does about the shadow which often falls between human projects and their execution, as well as what one uses to write books, he gets the point, and the speaker will often agree on the basis of the listener's response that the lestener has understood. Does it follow, then, that in understanding or using the odd utterance, the human speakers were acting according to a rule—in this case a rule for how to modify the meaning of "in"? It certainly does not seem so to the speakers who have test recognized the utterance as "odd"[264].

Dreyfus confronte les habiletés linguistiques humaines aux capacités d'un ordinateur en traitant les mêmes genres d'informations de caractère linguistique. Ses conclusions, quant aux défauts et limites de l'ordinateur à traiter le langage naturel, peuvent être résumés par les points suivants:

1. les programmes informatiques présentent de l'intolérance à la plupart des ambiguïtés sémantiques et aux fautes de grammaire ainsi qu'à toute infraction aux règles d'usage et de syntaxe.

2- Ils sont incapables de faire face aux situations linguistiques inattendues. (l'ordinateur applique indûment des règles à des cas nouveaux ou inconnus qui échappent aux règles.) ou doit toujours réviser l'ensemble de règles pour prendre en considération les nouveaux usages linguistiques entrés.

3- Ils n'ont pas la capacité de travailler sélectivement avec le sens des phrases et des mots; pour cette raison ils ne sont pas capables de générer l'application de règles sémantiques qui exigent le recours aux contextes d'énonciation.

Les êtres humains, au contraire des machines numériques, sont capables de percevoir immédiatement les infractions à une règle et de recourir au contexte pour rendre intelligible des expressions problématiques d'une langue. D'ailleurs, la plupart des usagers d'une langue utilisent celle-ci de manière très souple mais sans provoquer des conflits avec les patrons (règles) grammaticaux recommandés. Dans la vie quotidienne, ils comptent beaucoup sur le contexte d'énonciation pour se communiquer. Ils utilisent (ou ignorent) les règles selon les contextes d'énonciation. Dans les situations conversationnelles, les usagers d'une langue recourent à une sorte d'accord tacite entre eux, ce qui leur permet de faire face à des situations linguistiques où les règles sont absentes ou violées. Dreyfus ne croit pas que nous utilisons des méta-règles (comme dans les programmes heuristiques) pour résoudre des situations conversationnelles critiques.

---

264 H. L. Dreyfus (1979), *op. cit.* , p.p.198-199.

Devant un problème difficile qui demande une compréhension globale de la situation, les machines échouent. Elles ne s'impliquent pas comme les êtres humains dans les situations. Toute l'information linguistique est traitée en termes de séquences binaires de données, ayant une valeur mais sans contenu.

> But computers are not involved in a situation. Every bit of data always has the same value. (...) they can apply a rule to a specific case if the specific case is already unambiguosly described in terms of general features mentioned in the rule. They can thus simulate one kind of theoretical understanding. (...) Thus they cannot accept ambiguity and the breaking of rules until the rules for dealing with the deviation have been so completely specified that the ambiguity and the breaking of rules until the rules for dealing with the deviations have been so completely specified that the ambiguity has disappeared[265].

Dreyfus affirme que les recherches sur le langage naturel en IA n'ont pas réussi à rendre compte des situations réelles de communication linguistique. La machine est très loin des formes de vie humaines. L'ensemble des règles qui composent les programmes informatiques est valable pour le traitement des phénomènes qui appartiennent au monde formel et objectif des sciences, mais pas à l'univers de l'homme plein de significations.

Dreyfus ajoute que les contextes linguistiques pragmatiques sont inaccessibles aux machines. Le formalismes conçus pour programmer les machines digitales ne sont pas capables d'aller au delà des formes syntaxiques de traitement de l'information; ils ne sont pas capables de créer "une théorie de la pratique" comme le voulait Leibniz[266]. Le langage humain est marqué par une dépendance du contexte conversationnel, lequel n'est pas complètement soumis à des règles strictes.

Dreyfus affirme que Wittgenstein avait dejà remarqué le problème principal que doit affronter qui veut défendre une présupposition de type épistémologique. Ce problème est celui d'essayer de comprendre le langage au moyen de règles strictes. Selon Wittgenstein dans l'utilisation courante d'une langue nous n'appliquons jamais des règles strictes[267]. Pour cet auteur, nos comportements linguistiques ne peuvent pas être réduits à leur expression formelle. Une telle réduction nous impliquerait dans une régression infinie à des règles exprimées ainsi par Dreyfus:

> To have a complete theory of what speakers are able to do, one must not only have grammatical and semantic rules but further rules which would enable a

265 H. L. Dreyfus (1979), *op. cit.* , p.201.
266 H. L. Dreyfus (1979), *op. cit.* , p.201.
267 Ludwig Wittgenstein, The Blue and Brown Books, Basil Blackwell, Oxford, England, p. 25 1960. Cité par H. L. Dreyfus (1979), *op. cit.* , p.203.

> person or a machine to recognize the context in which the rules must be applied. Thus there must be rules for recognizing the situation, the intentions of the speakers, and so forth. But if the theory then requires further rules in order to explain how these rules are applied, as the pure intelectualist viewpoint would suggest, we are in an infinite regress[268].

La présupposition épistémologique, en tant que recours formel à des règles semblables à celles de la physique ou de la linguistique, est condamnée à l'échec.

Le monde humain n'est pas constitué d'un ensemble de règles et de comportements programmables. L'esprit humain ne traite pas l'information par des méthodes heuristiques. Les démarches formelles, basées sur le succès de la linguistique contemporaine, ne permettent pas de confirmer la présupposition épistémologique Voilà, en résumé, les critiques de Dreyfus formulées à l'égard de la présupposition épistémologique.

3.4- La présupposition ontologique

La présupposition ontologique est définie par Dreyfus comme la thèse selon laquelle les conduites intelligentes peuvent en principe être comprises à partir d'un ensemble d'éléments indépendants et explicites. Les chercheurs en IA ont comme point de départ le présupposé ontologique lorsqu'il considèrent les comportements humains et les connaissances comme étant susceptibles d'être réduits à un ensemble d'éléments atomiques, plus simples, formalisables et capables d'être traités sur des ordinateurs digitaux.

La présupposition ontologique est théoriquement compatible avec la structure de fonctionnement des ordinateurs digitaux, dans le sens où toute démarche de traitement des informations en informatique exige que les données soient des éléments discrets, prédéfinis et indépendants du contexte. L'approche la plus courante en IA selon laquelle l'intelligence résulte d'un traitement d'information a pour base la présupposition ontologique.

> The information-processing approach, however, uses the computer to instantiate symbolic descriptions so that combinations of flip/flops represent discrete facts. If one assumes that these symbolic descriptions are composed of primitives which correspond to isolable features of the world, one makes the ontological assumption[269].

---

268 H. L. Dreyfus (1979), *op. cit.* , p. 203.
269 H. L. Dreyfus (1979), *op.cit.* , pp. 335-336.

Le présupposé ontologique est basé sur l'idée selon laquelle les données concernant um *monde humain* peuvent être comprises entièrement par des moyens formels. La discussion de Dreyfus sur ce problème s'articule autour des thèmes suivants:

a) Le rapport entre la présupposition ontologique et la tradition représentationnaliste en Occident.

b) Le rapport entre la présupposition ontologique et le modèle explicatif des sciences physiques.

c) Le traitement du langage naturel en IA basé sur la présupposition ontologique.

Le premier thème sert à montrer l'importance de la présupposition ontologique en tant que base philosophique pour l'IA. Les deux autres thèmes constituent une critique de l'IA et de ces base ontologiques issues des thèses mécanistes et linguistiques.

3.4.1- Le rapport entre la présupposition ontologique et la tradition représentationnaliste en Occident.

Pour montrer le rôle théorique de la présupposition ontologique pour les recherches en IA, Dreyfus la situe par rapport à la tradition. Le présupposé ontologique est en rapport avec la tradition représentationnaliste, selon lequelle les comportements intelligents et les connaissances humaines peuvent être complètement exprimés au moyen de règles lesquelles une fois appliquées constitueraient l'explication (ou la représentation) de ces comportements et connaissances.

Comme nous l'avons anticipé dans le deuxième chapitre il y a un rapport entre l'IA et la tradition philosophique: ce rapport est exprimé déjà dans la philosophie de Leibniz, Descartes, Hume, Hobbes, pour qui le monde peut être décomposé en éléments plus simples pour qu'on puisse le comprendre. Selon Leibniz, par exemple, notre attitude intelligente exige une recherche en profondeur dans laquelle un concept est décomposé en éléments plus simples, nous permettant de bien les comprendre. Étant donné que ces éléments conceptuels plus simples s'appliquent au monde des choses complexes, ce dernier doit être composé d'éléments de plus en plus simples auxquels correspondent des éléments conceptuels aussi élémentaires.

> Leibniz envisaged "a kind of alphabet of human thoughts" whose "characters must show, when they are used in demonstrations, some kind of connection,

> grouping and order which are also found in the objects." The empiricist tradition, too, is dominated by the idea of discrete elements of knowledge[270].

Les démarches rationalistes et empiristes visent à éviter tout ce qui peut constituer l'imprécision, l'incertitude, sur le plan moral, intellectuel ou pratique. Une fois que nous pouvons décomposer nos connaissances et les objets de la connaissance en élément plus simples on est en mesure de les exprimer clairement sous formes de règles ou de définitions qui peuvent être appliquées sans jugement de valeur, c'est-à-dire sans l'intermédiaire de l'interprétation individuelle et subjective.

Selon Dreyfus, les philosophes des écoles intellectualistes et empiriste, et ceux de l'atomisme logique comme Russell (et le premier Wittgenstein) ont défendu, à leur façon, l'idée selon laquelle on peut tout réduire à des éléments plus simples (des objets isolables) et les comprendre par le moyen des règles. Cette idée, de réduire tout ce qui existe à des *éléments originels* (des parties simples qui composent toute la réalité) remonte à Socrate comme le constate Wittgenstein dans les *Investigations philosophiques:*

> Socrate — dans le *Théétète* «si je ne fais erreur, j'ai entendu dire par quelques-uns: pour ce qui est des *éléments originels* — si je puis ainsi m'exprimer— dont nous sommes composés ainsi que tout le reste — il n'y aurait point d'explication: car tout ce qui existe en soi et pour soi, on ne saurait le désigner que par un nom; (...)

[Wittgenstein ajoute encore:]

> Ces éléments originels étaient aussi les « individuals » de Russell et aussi mes propres « objets » (*Tratactus logico-philosophicus*)[271].

Selon Dreyfus, Wittgenstein (*Tractatus*) entend que le monde pourrait être défini en termes d'un ensemble de faits atomiques pouvant être exprimées par des propositions logiquement indépendantes[272]. La présupposition ontologique des chercheurs en IA consiste à admettre cette façon de penser du premier Wittgenstein. Selon ceux qui travaillent dans ce domaine il est possible de fournir une représentation formelle du monde, basée sur un système de descriptions dans lequel le monde est conçu comme un ensemble composé d'éléments atomiques plus simples organisés sur la forme d'une structure de données.

---

270 H. L. Dreyfus (1979), *op. cit.* , p.211.
271 Cf. L. Wittgenstein, *Investigations philosophiques*, Librairie Gallimard, Paris, 1961 p.136, §46(Traduit de l'allemand par Klossomski, Pierre.) Voir aussi, note19, chap.II plus haut. Il est interessant de noter que, plus bas wittgenstein refuse l'idée exprime dans cette citation (Cf. *Investigations philosophiques*, §47).
272 H. L. Dreyfus (1979), *op. cit.* , pp.211-212.

La présupposition ontologique est selon Dreyfus un élément important pour l'analyse de l'IA. Cette présupposition n'est jamais remise en question. Elle constitue le réflexe de deux mille ans de tradition philosophique[273]. et est sous-jacente à presque tous les travaux théoriques en IA, cependant cette présupposition n'est pas formulée explicitement par ceux qui travaillent dans le domaine de l'IA.

Dreyfus affirme que la présupposition ontologique peut être rencontrée dans des travaux importants de l'IA. il prend par exemple, les travaux de Minsky ou cette présupposition est selon lui, évidente. Minsky prétend que les comportements intelligents peuvent être compris en termes d'éléments indépendants et prédéfinis, afin de les représenter efficacement sur la forme de programme d'ordinateur. Pour cela il faut qu'ils soient décomposés en élément discrets plus simples, lesquels sont traités comme de données informatiques.

Cet auteur reconnaît que nous ne pouvons pas obtenir de l'intelligence du niveau de l'être humain en IA. Pour lui il faut comprendre le comportement intelligent comme le résultat d'un ensemble de processus et de connaissance; car les comportements humains ne peuvent pas être étudiés isolement, mais plutôt comme résultant d'une structure de connaissances.

Selon Minsky nous pouvons analyser et expliciter les comportements humains au moyen de règles heuristiques et nous pouvons comprendre leurs rapports avec toutes sortes de connaissance, modèles et processus en termes de relations entre catégories de faits, catégories d'objets et des faits au sujet d'objets du monde.

D'où vient cette idée selon laquelle le comportement intelligent est attaché à un corps de données bien structurées et qu'il est opéré à partir de règles heuristiques?

Dreyfus répond que cette idée est enracinée dans la tradition philosophique occidentale, depuis Platon. L'idée proposée par Minsky de concevoir le comportement comme résultant des rapports entre faits et objets dans le monde est liée, plus précisément, à la tradition représentationnaliste selon laquelle tout se présente comme des faits déjà prêts à être traités, ou mieux à être représentés par l'esprit.

> In fact, by supposing that the alternatives are either a well-structured body of facts, or some disembodied way of dealing with the facts, Minsky is so traditional that he can't even see the fundamental assumption that he shares with the whole of the philosophical tradition. In assuming that what is given are facts at all, Minsky is simply echoing a view which has been developing since Plato and has now become so ingrained as to *seem* self-evident[274].

273 Cf. H. L. Dreyfus (1979), *op.cit.* p. 207.
274 H. L. Dreyfus (1979), *op. cit.* , p.211.

Dreyfus ne voit pas pourquoi la présupposition ontologique peut servir de base à l'optimisme en IA; pour lui, le savoir humain qui est derrière le comportement intelligent est un genre de savoir pratique, un *savoir-faire* qui ne peut pas être réduit à un ensemble fini de faits ou d'éléments de connaissances décrits en termes de données à être manipulés par des moyens informatiques. Le comportement humain doit être analysée en fonction d'un *monde humain*:

> Even a chair is not understandable in terms of any set of facts or "elements of knowledge." To recognize an object as chair, for example, means to understand its relation to other objects and to human beings. This involves a whole contexte of human activity of which the shape of our body, the institution of furniture, the inevitability of fatigue, constitute only a small part. And these factors in turn are no more isolable than is the chair. They all may get *their* meaning in the context of human activity of which they form a part[275].

Une autre question qui représente de difficultés est la question de la représentation, organisation et formalisation de types de connaissance sur le monde en rapport avec les contextes. Dreyfus se demande si on peut formaliser ces connaissances sur le monde en termes d'enormes *bases de données* isolées. Est-ce que nous pouvons dire, se demande-il, que nos connaissances sur le monde sont en rapport avec la connaissance de milliers de faits et d'autres connaissances isolables traités de façon séquencielle, enmagasinées et repérés par une machine digitale?

Pour Dreyfus les êtres humains ne dépendent pas d'artifices formels comme ceux utilisés en informatique ni de règles pour avoir une connaissance du monde. Nous ne dépendons pas d'une *base de faits* ni d'une *base de connaissances* pour comprendre le monde et pour agir de façon intelligente.

Pour selectionner les stratégies de résolution d'un problème et interpréter certaines données concernant la reconnaissance de formes et pour affronter des situations plus complexes du quotidien des gens il faut que la machine puisse opérer sur des contextes. Pour programmer une machine de manière à lui permettre d'interpréter les données et de les traiter par rapport à des contextes, il faut établir une représentation formelle du contexte où le problème et la solution du problème se situent. Chaque contexte constitue un vaste

---

275 H. L. Dreyfus (1979),*op. cit.* , p.210.

ensemble de données et nous n'avons pas les moyens de savoir comment il se structure et comment il peut être transmi à l'ordinateur.

Nous ne savons pas comment procéder adéquatement, de manière à décomposer le monde en éléments simples formalisables, ni comment transmettre à l'ordinateur le contexte à partir duquel il ira traiter les données sur ce monde. Dreyfus ne croit pas qu'on est en mesure de résoudre ce problème.

Les initiatives en IA pour rendre les contextes exprimables en termes d'une vaste base de données quasi complète ont, selon Dreyfus, un caractère naïf. L'IA n'a, affirme-t-il aucune chance d'aboutir à des résultats pertinents, car notre intelligence ne requiert pas que nous résolvions au préalable le problème du stockage et de l'exploitation de complexes bases de données comme le fait un ordinateur digital.

Les chercheurs en IA savent que les problèmes de la représentations des connaissances, reconnaissance de formes, etc, ne se résument pas au stockage et au repérage dans une base de données. Minsky lui-même reconnait qu'il y a d'autres problèmes plus importants à résoudre que la classification et le repérage des connaissances à partir d'une vaste base de données. Pour Minsky il faut mieux connaître l'agencement et l'organisation des règles heuristiques dont les humains font usage pour exploiter leur structure de connaissance.

> The problem-solving abilities of a highly intelligent person lies partly in his superior heuristics for managing his *knowledge-structure* and partly in the structure itself; these are probably somewhat inseparable. In any case, there is no reason to suppose that you can be intelligent except through the use of an adequate, particular, knowledge or model structure[276].

Dreyfus, n'est pas d'accord avec l'idée de *structure de connaissance*, telle que conçue par les chercheurs en IA. Les processus qui permettent l'intelligence sont, selon lui, beaucoup plus souples et imprévisibles. Le savoir humain et le comportement intelligent ne peuvent pas être compris en termes d'un vaste ensemble de données organisées en une base de connaissance. La façon dont nous traitons les informations n'est pas fixe et n'est pas soumise à des règles formelles opérées en fonction d'input et caractérisant les relations entre des catégories d'objets, de faits ou de sujets qui constituent une "structure de modèle." (*model structure*).

> In general, we have an implicit understanding of the human situation which provides the context in which we encounter specific facts and make them explicit. There is no reason, only an ontological commitment, which makes us

276 Minsky, Semantic Iformation processing, p. 25. Cité par H. L. Dreyfus (1979), *op.cit.* , p.210. Notre italique.

> suppose that all the facts we can make explicit about our situation are already unconsciouly explicit in a "model structure," or that we could ever make our situation completely explicit even if we tried[277].

Selon Dreyfus, notre connaissance sur le monde ne peut pas être décomposée en éléments discrets, ou des *atomes d'expérience* comme voulait Hume[278]. En dépit du fait que le modèle formel représentationnaliste est en rapport avec la recherche de rigueur et de précision formelle qui existe depuis Platon, il ne peut pas constituer une garantie de la réussite des recherches en IA.

3.4.2- Le rapport entre la présupposition ontologique et le modèle explicatif des sciences physiques.

La possibilité de décrire formellement les systèmes physiques stimule des chercheurs en IA, comme J. Mac Carthy, à croire que la plupart des systèmes se prêtent à la formalisation et qu'ils peuvent être simulés par un programme d'ordinateur. Dreyfus affirme que le monde humain n'est pas une structure organisée comme un système physique, lequel peut être expliqué à partir d'un ensemble d'axiomes.

Selon Dreyfus le présupposé ontologique représente une exigence d'objectivité de la part de ceux qui travaillent en IA de la même façon que certains critères concernant l'objectivité (de caractère formel-descriptif) ont été fondamentaux pour le développement de la physique depuis Gallilée. Selon Dreyfus, le modèle tiré des sciences physiques ne constitue pas une garantie à l'explication et à la reproduction des comportements intelligents humains, par des moyens informatiques.

En tant que système physique nous réagissons d'une façon mécanique. Étant des êtres humains, nos réactions sont toujours inattendues. Dreyfus note cependant, que les situations humaines ne peuvent pas être confondues avec des états d'un système physique.

> The ontological assumption that the human world too can be treated in terms of a set of elements gains plausibility when one fails to distinguish between world and universe, or what comes to the same thing, between the human situation and the state of a physical system[279].

---

277 *Idem* .
278 Pour Hume, les impressions qui composent notre expérience sont constituées "d'atomes d'expérience". Ces atomes peuvent être isolés de manière à nous permettre de comprendre la pensée.
279 H.L. Dreyfus, *op.cit.* , p.213.

C'est plutôt le point de vue contraire que John Mac Carthy défend dans "Programs with common sense"[280]:

> One of the basic entities in our theory is the *situation*. Intuitively, a situation is the complete state of affairs at some instant in time. The laws of motion of a system determine all future situation from a given situation. Thus, a situation corresponds to the notion of a point in phase space[281].

Selon Dreyfus, MacCarthy semble ignorer que la situation humaine est plus qu'un état determiné en rapport avec un instant donné. La situation, affirme Dreyfus, est la manifestation du comportement des êtres humains; elle est changeante par le fait qu'il y a autant de situations que d'êtres humains dont les objectifs et les intentions sont différents.

Selon Dreyfus, MacCarthy confond *situation* et état physique de l'Univers. C'est-à-dire, il confond *tokens* d'états physiques (événements physiques particuliers) et *types* d'états physiques. Dreyfus ajoute qu'une *situation token* peut, sans problèmes être identifié à un état physique token (caractérisé comme un point dans un *espace de phases*, dans le cas cité), mais il n'y a pas d'identité entre un *type* de situation et un *type* d'état physique.

Pour expliquer cette distinction *type-type* en rapport avec la notion de *situation* exposée par J. MacCarthy Dreyfus fait appel à l'exemple de ce même auteur écrit dans le langage LISP: "'At (I, home) (s) ' means I am at home in situation s.". Selon MacCarthy chaque situation correspond à un instant donné ou à un état du monde physique. La situation d"être chez soi" correspond à être dans un état physique déterminé.

Dreyfus considère cependant que "être chez soi" n'est pas identique à un type d'état physique. La situation "être chez soi", n'est pas attachée à un type d'état physique, car elle est remplie de significations pour les êtres humains pouvant ainsi correspondre à plusieurs sortes d'états dans le monde physique ou à aucune sorte d'état physique trouvé dans ce monde.

Nous pouvons être dans une situation comme "être chez soi" et néanmoins n'être pas physiquement à l'intérieur de la maison (la nôtre); si nous sommes dans le jardin de la maison par exemple. Si nous sommes dans un bon Hôtel pendant les vacances et tout est à notre goût. L'expression "être chez soi" est encore valable pour exprimer une situation que correspond à un tout autre état de choses physiques. Pour Dreyfus "être chez soi" est une

---

280 J. MacCarthy, *in Semantic processing information*, p.403; cité par Dreyfus, *op.cit.* , p.213.
281 *Idem.* Il vaut mentioner que MacCarthy applique le concept de situation selon un point de vue mécaniste dans le but de construire un programme en langage naturel qui opèere par manipulation de phrases de exprimées sous la forme d'un langage formel de programation du genre Lisp: 'At (I, home) (s)'.

situation humaine qui correspond à un monde humain quelquefois très éloigné de l'univers sans signification de la physique:

> I can also be physically in my house and not be at home; for example, if I own the house but have not yet moved my furniture in. Being at home is a human situation, not in any simple corespondence whith the physical state of a human body in a house. Not to mention the fact that it is a necessary if not sufficient condition for being at home in the sense in question that I own or rent the house, and owing or renting a house is a complicated institutional set of relations not reducible to any set of physical states. Even a physical description of a certain pattern of ink deposited on certain pieces of paper in a specific temporal sequence would not contitute a necessary and sufficient condition for a transfer of ownership. Writing one's name is not always signing, and watching is not always witnessing[282].

Le présupposé ontologique peut servir de base à des recherches sur l'univers physique qui est dénué, selon Dreyfus, de signification, mais ce n'est pas le cas lorsqu'il s'agit de programmer des machines pour qu'elles reproduisent les comportements intelligents humains. Le monde humain est fait des concepts et des comportements qui ont un rapport avec des situations chargées de signification[283].

3.4.3- Le traitement du langage naturel en IA basé sur la présupposition ontologique.

Pour Dreyfus, la présupposition ontologique fait défaut quand il s'agit de reproduire sur un système digital les caractéristiques sémantiques du comportement linguistique humain.

L'échec de la traduction automatique nous permet dit Dreyfus, d'une part, de comprendre la complexité du langage naturel et d'autre part, de constater les faiblesses du formalisme pour rendre compte du caractère sémantique des phrases. La traduction n'est pas une question de manipulation formelle de symboles à la manière de la *cryptographie*. Il faut saisir les contextes linguistiques auxquels se rapportent les phrases et beaucoup d'autres connaissances extra-linguistiques.

Jusqu'à maintenant, aucun programme d'ordinateur ne permet de traduire adéquatement d'une langue à une autre. Il est possible de produire des traductions automatiques pour des domaines restreints, comme les textes scientifiques, résultats météorologiques, etc. Cependant, même dans ces domaines on est encore confronté à plusieurs limites.

---

282 H. L. Dreyfus (1979), p.214.
283 H. L. Dreyfus (1989), p.974.

Il faut beaucoup de connaissances non-linguistiques pour traduire une langue dans une autre. Selon Dreyfus, il est peu probable que nous puissions, par les moyens formels à notre disposition, transmettre aux ordinateurs tout l'univers de connaissance humaine. L'analyse du contexte en termes formels, en attribuant des indices aux différents sens d'un mot selon le contexte et les autres mots voisins, n'est pas du tout suffisante pour régler les problèmes de traduction et du traitement du langage naturel.

Les programmes de traitement du langage naturel ne peuvent pas compter sur des moyens de désambiguïsation assez efficaces car ces moyens sont seulement disponibles chez les êtres humains. Le traitement du langage naturel requiert l'appel aux contextes réels et à des *situations* et non à des faits isolés. La connaissance du contexte nous permet de saisir les aspects sémantiques des phrases traduites et de rendre explicite le texte traduit en fonction de la situation.

Cette distinction entre le recours à des faits et le recours à des situations, comme facteur de désambiguïsation sémantique, a été important pour les recherches d'auteurs, comme Katz et Fodor, et Bar Hillel, qui ont traité des problèmes de sémantique concernant le traitement du langage naturel.

Les auteurs mentionnés conçoivent les modèles sémantiques qui sont basés, selon Dreyfus, sur la présupposition ontologique. Katz et Fodor, par exemple, reconnaissent qu'il est nécessaire, pour choisir le contexte linguistique et désambiguïser les phrases, de disposer de toute une base de connaissance sur le monde. Ils soulignent également que cette connaissance est complètement distincte de nos connaissances linguistiques, et qu'il est pratiquement impossible de réunir toutes ces théories en un seul système.

Dreyfus est d'accord avec l'idée mentionnée plus haut, mais il rejette l'idée selon laquelle le monde peut être analysé sous forme de faits isolés. Il admet l'importance de la connaissance de la situation pour la désambiguïsation en ce que concerne les recherches sur le langage naturel en IA. Pour lui la *situation* joue un rôle efficace pour la désambiguïsation du langage.

La notion de *situation*, pour Dreyfus, a un sens proche de celui défendu par les gestaltistes. Selon lui la situation est une sorte d'arrière-plan qui nous permet d'identifier immédiatement le sens des phrases d'une langue.Il n'accepte pas l'idée que la notion de situation aie une connotation mécaniste. Les chercheurs en IA n'ont pas trouvé un moyen de permettre à l'ordinateur de rendre compte de la situation de manière à lui permettre de désambiguïser le langage ou de surmonter certaines difficultés linguistiques complexes dans la conception des programmes.

En travaillant sur un programme en langage naturel, Joseph Weizenbaum a saisi qu'il fallait rendre compte du *contexte global* de la conversation pour pouvoir bâtir un programme capable de dialoguer en langage naturel. La notion de contexte global que Weizenbaum[284] utilise pour l'examen du comportement linguistique en tant que programmable semble se rapprocher de la notion de *situation* que Dreyfus défend comme étant essentielle pour la compréhension des phrases en langage naturel et d'autres activités exigeant de l'intelligence.

L'appel à des *contextes globaux*, offre, dit Weizenbaum, la possibilité de repérer le sens dans une situation de conversationnelle. Il est lui-même une espèce de marqueur de sens, qui nous permet de comprendre dans une conversation ce qui est dit et de capter d'un seul coup les sous-contextes et sous-sous-contextes qui permettent la communication verbale entre les êtres humains.

Les êtres humains ont cette capacité de faire appel à des *contextes globaux* lorsqu'ils communiquent, dit Weizenbaum. L'auteur constate la difficulté à programmer ce processus dans une machine. Il reconnaît la difficulté de comprendre complètement ces contextes globaux et d'élaborer des programmes capables de reconnaître le langage naturel. Il n'exclut pas,cependant, contrairement à Dreyfus, l'idée que de tels contextes puissent être traités en tant qu'ensemble de faits en forme "d'arbres contextuels" nous permettant de représenter certaines connaissances humaines.

Dreyfus remarque que bien que Weizenbaum admette l'importance de la *situation* (contexte global) dans la compréhension du sens linguistique, son projet est fondée sur la présupposition ontologique, tout comme l'ensemble des recherches en Intelligence Artificielle.

Selon Weizenbaum il est possible de décomposer le contexte global des phrases du langage en traits caractéristiques plus simples comme les "branches" d'une arbre. Il pense également qu'il est possible de représenter tous les ensembles de sous-contextes d'une conversation en termes de procédures pratiques capables de circonscrire le sens des mots et permettre de concevoir des programmes dotés d'une certaine compréhension du langage naturel.

Dreyfus affirme que les recherches en IA sur le langage naturel exigent (en vertu du caractère binaire de l'ordinateur), qu'on sépare le sens en contexte du sens du mot. Cela est complètement en désaccord avec la façon dont les êtres humains comprennent le langage. Il

---

284 Weizenbaum, J. "Contextual Understanding by Computers" in *Recognizing Patterns*. p.181. Cité par H. L. Dreyfus (1979), p. 338.

ne semble pas, pour Dreyfus, que nous hiérarchisons des contextes pour identifier une situation ou contexte global de conversation. On le fait d'un seul coup. Les exigences techniques inhérentes au hardware et au software exigent du programmeur qu'il hiérarchise les sous-contextes linguistiques afin de déterminer le sens des mots.

Le *contexte global* doit être fractionné en éléments plus simples à cause d'une exigence de caractère technique, car les ordinateurs ne peuvent pas capter globalement le langage. Les machines digitales traitent seulement des éléments discrets. Les ordinateurs n'identifient pas vraiment les contextes, ne vivent pas en *situation* affirme Dreyfus. Elles manipulent des fragments de la situation sous la forme d'ensemble de faits.

Le traitement du langage naturel montre, selon Dreyfus, des difficultés théoriques et pratiques suivantes: il est difficile, sinon impossible, d'éliminer les ambiguïtés du langage naturel, car nous ne pouvons pas déterminer au préalable quel ensemble de faits constituent un contexte donné. De plus le nombre de faits à être considérés grandit exponentiellement devenant pratiquement infini par conséquence le nombre de contextes à identifier devient infini.

La programmation des règles permettant de saisir les faits pertinents dans une sitation conversationnelle est insuffisant pour déterminer le contexte dans la mesure où ces règles ne déterminent pas entièrement l'ensemble des traits qui définissent un contexte donné. Le recours à des faits peut causer d'autres ambiguïtés vu qu'ils peuvent définir plusieurs autres contextes distincts avant d'être interprétés. Dans le passage suivant, Dreyfus nous parle de la limitation de la programmation dans l'identification des traits pertinents qui permetent la formalisation d' un contexte de conversation donné et de desambigüiser les phrases du langage naturel:

> Evidently, a broader context will have to be used to determine which of the infinity of features is relevant, and how each is to be undertood. But if, in turn, the program must enable the machine to identify the broader context in terms of *its* relevant features—and this is the only way a computer which operates in terms of discrete elements could proceed—the programmer must either claim that some features are intrinsically relevant and have a fixed meaning regardless of context— a possibility already excluded in the original appeal to context—or the programmer will be faced with an infinite regress of contexts. There seems to be only one way out: rather than work up the tree to ever broader contexts the computer must work down from an ultimate context—what Weizenbaum calls our shared culture[285].

285H. L. Dreyfus (1979), pp.220-221.

Le *contexte ultime* de conversation en rapport avec la culture que nous partageons (shared culture) n'a pas besoin d'être interprété, dit Dreyfus; il est une stratégie utilisée par Weizenbaum afin d'éviter le recours à des contextes toujours plus larges dans la discrimination des faits pertinents qui permettent de désambigüiser les phrases du langage naturel.

Le fait que le recours à des contextes de plus en plus larges représentent une piste pour le repérage du sens et le fait qu'un contexte ultime semble aussi intervenir pour éviter que nous nous perdions dans une régression infinie de sous-sous-contextes résulte, selon Dreyfus, d'une antinomie qui affecte les démarches formelles ayant pour but la conception d'une "intelligence artificielle". Dreyfus expose ainsi cette antinomie:

> It seems that given the artificial intelligence worker's conception of reason as calculation on facts, and his admission that which facts are relevant and significant is not just given but is context determined, his attempt to produce intelligente behavior leads to an antinomy. On the one hand, we have the thesis: there must always be a broader context; other-wise, we have no way to distinguish relevant from irrelevant facts. On the other hand, we have the antithesis: there must be an ultimate context, which requires no interpretation; otherwise, there will be an infinite regress of contexts, and we can never begin our formalization[286].

Selon Dreyfus, il est fort probable que les comportements intelligents des humains soient en rapport avec la culture qu'ils partagent, et que cela soit un facteur important pour la compréhension du sens des mots du langage naturel etc... Mais rien ne peut garantir que nous pouvons rendre compte de façon formelle de cette "culture partagée". Il semble que cet élément de définition et de désambiguïsation qui nous oriente dans le monde humain résiste à être progammée sur un ordinateur digital

Selon Dreyfus, la notion de *culture partagée* (shared culture) rappelle ce que Wittgenstein a nommé *formes de vie*. Cependant les *formes de vie* semblent présupposer une intelligence naturelle capable d'apprendre des règles sans se soumettre à elles, d'identifier les situations en étant déjà en situation, c'est-à-dire, étant directement engagé dans ce vaste "monde de la vie humaine". Dreyfus affirme que nous ne savons pas exactement ce que c'est que ces formes de vie et encore moins comment les formaliser.

---

286 H. L. Dreyfus (1979), p. 222.

Conclusion

Pour Dreyfus, la recherche en IA a des racines profondes, lesquelles constituent comme nous l'avons déjà souligné, la tradition philosophique représentationnaliste de l'Occident depuis Platon.

Dreyfus montre dans *What Computers Can't Do*, que ceux qui travaillent dans le domaine de l'IA ont sous-estimé le problème de la programmation des processus cognitifs sur ordinateur Ils n'ont pas considéré les difficultés théoriques des thèses représentationnaliste selon laquelle le comportement intelligent constitue un mécanisme de traitement d'informations et est facilement décomposable en éléments plus simples capables d'être objets de manipulations formelles.

À partir de l'analyse des quatre présupposés sous-jacents à l'optimisme des recherches en IA et en simulation cognitive Dreyfus conclut qu'en dépit des faiblesses épistémologiques de ces présupposées ils ont été importants car ils ont permis d'éviter que les échecs dans ces domaines causent un pessimisme généralisé et l'abandon de ce genre de recherche. Selon Dreyfus, les présupposés mentionnés posent des problèmes insurmontables.

Dreyfus soutient qu'il est nécessaire de mieux définir ce qu'on entend par explication en IA. Si expliquer un comportement signifie analyser et décrire complètement un comportement par un ensemble d'instructions ou des règles, alors les psychologues inspirés des thèses en IA se trouvent face à des problèmes d'ordre épistémologique: car, nous ne pouvons pas confondre la règle qu'on suit quand on fait quelque chose avec la règle qui sert à décrire ce qu'on fait. Le mouvement planétaire peut bien être représenté en termes d'équations, mais cela ne revient pas à dire que les planètes résolvent des équations quant elles se déplacent.

Les efforts logiques appliqués à la simulation du comportement humain et à l'imitation de notre comprehénsion du langage n'ont pas permis de faire des programmes capables des mêmes performances que les êtres humains.[287] Les expériences en rapport avec la résolution de problèmes simples, recherches faites par Minsky et Simon, ont à peine montré que les êtres humains prennent des raccourcis ou des attitudes cognitives qui résistent à la programmation logique. Les êtres humains semblent obtenir des résultats beaucoup plus élégants en évitant plutôt qu'en essayant de résoudre les difficultés en rapport avec un problème donné ou avec la comprehénsion du langage naturel. Pour

287 Dreyfus affirme que les activités intelligents de caractère non formelle (voir Zone 4 de l'annexe p.) représentent, jusqu'à maintenant, un limite insurmontable pour la programmation en IA. Seulement les activés associationnismes et quelques activités intelligentes de caractère simples et complexes (Voir Zones 1 à 3 de l'annexe mentionné) on été atteintes par les chercheurs dans ce domaine.

résoudre un problème, par exemple, les humains n'opèrent pas formellement à la façon des machines digitales. Toute une gamme de notions et de connaissances entrent en ligne de compte; la situation vécue y est un élément essentiel qui n'a pas encore été compris.

Les conduites intelligentes de l'être humain ne sont, selon Dreyfus, jamais écartées de la *situation* dans laquelle elles se trouvent. Les processus qui donnent naissance à des comportements intelligents ne sont pas discrets et sélectifs à la façon d'une machine digitale. Nos conduites intelligentes sont liées à des faits et situations passés et présents. Elles dépendent du contexte et du rôle que nous jouons dans un contexte déterminé.

Dreyfus affirme que l'esprit humain ne peut pas être compris par le biais des programmes heuristiques comme le soutiennent les chercheurs qui défendent les présuppositions épistémologique et ontologique. Il n'y a pas de preuve empirique ou théorique qui permette de croire correct le présupposé psychologique que le comportement, humain intelligent peut être expliqué par l'intermédiaire d'un ordinateur.

Dreyfus montre que d'un point de vue épistémologique les chercheurs en IA et en science cognitive auraient des difficultés à présenter leurs recherches comme des hypothèses scientifiques à la base des théories. Il ne semblent pas, selon Dreyfus, être capables de soumettre leurs thèses à l'épreuve de la généralisation selon des critères épistémologiques courants, car leurs expériences ont un caractère *ad hoc* exigeant toujours de conditions *caeteris paribus*.

Les présuppositions ontologique et épistémologique ont en commun le fait qu'elles ne permettent pas de résoudre la difficulté de formaliser les processus cognitifs humains dans leur complexité ou leur caractère global. Elles conduisent à l'hypothèse non justifiée selon laquelle il est possible d'analyser la conduite humaine par des règles capables de relier des faits isolés concernant une conduite spécifique et permettant de formaliser ces conduites.

Selon Dreyfus, les présuppositions épistémologique et ontologique ne peuvent pas servir d'appui à une théorie de la pratique. Le comportement humain ne peut pas être compris comme le résultat d'une manipulation formelle de symboles ou de faits, par des règles précises à la manière d'un ordinateur qui fonctione par des opérations binaires et selon des règles strictes.

Pour Dreyfus, il y a cinq caractéristiques communes aux présuppositions mentionnées.

1. On y postule que l'homme fonctionne comme un mécanisme pouvant être appréhendé par des moyens formels.
2. Elles sont toutes admises comme vraies
3. Elles sont considérées comme évidentes et fonctionnent comme des axiomes.

4. Elles ne sont pas tout à fait explicites dans les ouvrages en Intelligence Artificielle parce ce que le chercheur les énoncent rarement en termes clairs.

5 Elles n'ont jamais été remises en question par les chercheurs bien qu'elles soient sous-jacentes à leurs travaux.

La critique philosophique de l'IA faite par Dreyfus montre que cette recherche est en rapport avec la tradition philosophique représentationnaliste et que la discussion de problèmes d'ordre technologique et théorique dans ce domaine peut être exploitée comme thème de discussion philosophique sur les limites de la pensée humaine et de la philosophie traditionnelle représentationnaliste.

Il y a, dans l'ouvrage étudié, une opposition entre deux courants de la philosophie. D'une part, il y a la tradition métaphysique représentationnaliste que nous avons esquissé dans le premier chapitre de ce travail. Cette tradition rationaliste est marquée par la pensée de Platon, de Leibniz, des empiristes anglais, de Russell et du premier Wittgenstein. D'autre part, il y a la phénoménologie qui représente une rupture avec ce rationalisme et une critique à toute formes de réduction formelle de la pensée[228], dans le sens qu'elle ne travaille pas sur une base représentationnaliste.

Selon Dreyfus, nous ne pouvons pas décomposer les comportements intelligents ni le réel en termes d'éléments plus simples formalisables. Les comportements humains sont en rapport avec des situations d'un monde construit par les êtres humains. Le réel est toujours donné par une situation complexe en rapport avec ce monde humain plein de sens et changeant, c'est-à-dire plein d'incertitudes.

Le monde humain résiste, selon Dreyfus à être compris selon un modèle fixe de la logique car il est le fruit d'une interaction culturelle entre individus qui le rend cohérent par le fait qu'ils partagent un langage et des pratiques sociales.

Pour que les machines puissent être intelligentes il faut les programmer avec une connaissance courante afin qu'elles soient capables de percevoir le monde, comprendre le langage naturel et être en situation comme n'importe qui parmi nous:

228Dreyfus nous fourni dans *What Computers Can't Do*, des alternatives non-représentationnalistes à l'étude de la cognition, mais il reconnaît que telles alternatives ne peuvent pas servir comme modèles pour la programmation des machines en IA. Les recherches en IA, basées sur des techniques de programmation traditionnelles et sur des architectures informatiques classiques, exigent des modèles symboliques représentationnalistes. En vertue de leur caractère opérationnel les recherches en IA ne peuvent pas avoir comme point de départ les alternatives non-représentationnalistes offertes par Dreyfus.

> Le problème de la compréhension courante nous apparaîtra clairement si nous réfléchissons qu'un ordinateur est encore plus étranger qu'un Martien lorqu'il pénetre dans notre monde. Il est dépourvu de corps, de besoins ou d'émotions, il n'est pas formé par un langage commun ou autres pratiques sociales. Si vraiment l'ordinateur devait interagir intelligemment avec nous, il faudrait qu'il soit doué de la faculté de comprendre la forme de vie humaine. Tout ce qui fait que nous sommes des humains, par exemple que les insultes nous mettent en colère, qu''il est plus facile de se déplacer vers l'avant que vers l'arrière, que nous pouvons passer devant les choses en nous déplaçant en leur direction puis en nous éloignant d'elles, que le temps passe irrémédiablement et que les événements futurs deviennent des événements passés, tout cela et bien d'autres choses encore doit être programmé dans l'ordinateur comme autant de faits et de règles. Pour reprendre l'expression des chercheurs en IA, il faut donner à l'ordinateur notre système de croyances[289].

La réussite de l'intelligence artificielle serait selon Dreyfus la réussite du projet de la métaphysique représentationnaliste traditionnelle d'analyser le monde et l'esprit comme de choses décomposables en éléments isolables et précis. Selon Dreyfus cette réussite est impossible, car l'IA a comme point de départ un préjugé philosophique de 25 siècles de métaphysique. Ce préjugé est maintenu encore par quelques scientifiques dans le domaine de l'IA[290].

Dreyfus dit que ce préjugé représente les limites de la tradition représentationnaliste lesquelles sont signalées par deux phénoménologues, Merleau-Ponty et Heidegger, qui ont mis en question les thèses représentationalistes. Ces deux philosophes nient qu'on puisse comprendre notre pensée en la décomposant en éléments plus simples (discrets) capables d'être représentés par des moyens formels:

> Merleau-Ponty calls the assumption that all that exists can be treated as determinate objects, the ***préjugé du monde***, "presumption of commonsense." Heidegger calls it ***rechnende Denken***, "calculating thought" and views it as the goal of philosophy, inevitably culminating in technology[291].

Selon Dreyfus, Heidegger et Merleau-Ponty ont compris très bien la crise de la philosophie occidentale et les échecs du rationalisme Ces deux auteurs serviront de base aux thèses de Dreyfus (dont le caractère phénoménologique n'a été pas abordé ici) où il défend que les formes de vie humaines résistent à la programmation.

---

289 H. L. Dreyfus (1989), p.975.

290 Selon Dreyfus, la crise de la philosophie occidentale a été marquée par la crise et les changements trouvés dans la philosophie de Hurssel ainsi que du premier Wittgenstein. Alors, même s'il y a eu un changement important dans la place de la philosophie, les scientifiques adoptent la perspective traditionnelle rationnaliste comme point de départ philosophique à leurs recherches.

291 H. L. Dreyfus (1979), p.212.

Si son analyse n'est pas complètement correcte, si les alternatives philosophiques qu'il propose pour l'étude de la cognition ne sont pas opérationnelles, son travail conduit, au moins, à une approche plus critiques en IA. En plus Dreyfus a le mérite d'être un des premiers, sinon le premier à montrer que l'intelligence artificielle pourrait être l'objet d'une réflexion philosophique sérieuse.

Suppose no one knew how clocks worked. suppose it was frightfully difficult to figure out how they worked, because, though there were plenty around, no one knew how to build one, and efforts to figure out how they worked tended to destroy the clock. Now suppose a group of researchers said, 'We will understand how clocks work if we design a machine that is functionally the equivalent of a clock, that keeps time just as well as a clock.' So they designed an hour glass and claimed: ' Now we understand how clocks work' ,or perhaps:
'If only we could get the hour glass to be just as accurate as a clock we would at last understand how clocks work.' Substitute 'brain' for 'clock' in this parable, and substitute 'digital computer program' for 'hour glass' and the notion of intelligence for the notion of keeping time and you have the contemporary situation in much (not all!) of artificial intelligence and cognitive science.[292]

John R. Searle

292 J. , Searle op. cit p.56.

CHAPITRE IV

Les critiques de J. Searle à l'Intelligence Artificielle et à la science cognitive

Présentation

Comme nous l'avons dit dans le chapitre précédent, la critique de l'IA faite par Dreyfus n'a pas comme point de départ la tradition métaphysique représentationnaliste; ce philosophe prend la voie phénoménologique pour montrer que l'IA ne peut pas, avec les moyens théoriques dont elle dispose, développer des machines vraiment intelligentes. Selon Dreyfus l'IA est dans la continuation du projet scientifique, poursuivi depuis toujours, de pouvoir expliquer complètement toutes les choses, et même l'esprit, par des moyens strictement formels.

En utilisant l'appareil théorique issu de ses thèses sur le langage et sur l'esprit, Searle va dans la même direction que Dreyfus, bien que lié à la philosophie analytique, il est un peu plus proche de la tradition philosophique représentationnaliste, même s'il veut parfois s'écarter de cette tradition[293].

A quel point ne s'agit-il pas d'un préjugé de dire que l'homme est le seul être existant doté d'intelligence? L'intelligence est-elle un critère primordial pour différencier l'homme du reste de l'univers physique?

Allan Turing, (tout en évitant de traiter directement le problème épineux de la pensée des machines) à répondu à cette question à partir de principes logiques et comportementaux[294]. *Dans Minds Brains and Science*[295], J. Searle analyse certains éléments théoriques importants liés au sujet de l'IA et de la science cognitive et donne une réponse à la question à laquelle Turing a choisi de se soustraire, à savoir, "Les ordinateurs digitaux peuvent-ils penser"? Pour le faire, il expose son point de vue sur la question des rapports entre le corps, l'esprit et le langage.

Les notions tirées de la philosophie de l'esprit et du langage de Searle vont appuyer sa critique à l'IA. Les rapports entre le langage et l'esprit sont un aspect important de la

---

293 Searle lui même ne se considère pas comme faisant partie de la tradition philosophique en occident il affirme cela lorsqu'il expose ses notions de représentation , sur les rapports entre l'esprit et le cerveau, et sur la pensée et le langage

294 A. Turing, "Computing Machinery and Intelligence", *in Mind*, vol. LIX, nº 236, G.B. ,1950.

295 J.R. Searle (1984), *op. cit.*

critique. Nous allons voir quelle est l'importance de deux caractéristiques de l'esprit, la conscience et l'Intentionnalité pour les critiques de Searle à l'IA.

Selon Searle la pensée est un phénomène intentionnel, lié à un arrière-fond et à un réseau d'états mentaux. Selon lui, l'IA ne peut pas simuler la conscience et l'Intentionalité, elle ne peut pas non plus représenter formellement l'arrière-fond qui est à la base de nos états mentaux. Pour ces raisons, ceux qui travaillent en IA n'arrivent pas à simuler ou reproduire la pensée humaine sur un programme informatique.

Nous ne savons pas encore très bien ce qui est vraiment l'esprit. Nous parlons d'états mentaux pour décrire toute une série de changements mentaux dont nous témoignons, mais nous ne savons pas du tout comment ces états mentaux se relient pour former nos pensées. Searle commence la discussion sur les difficultés d'explication des phénomènes mentaux par deux questions philosophiques, à savoir:

> Now, how can we square these two conceptions? How, for example, can it be the case that the world contains nothing but unconscious physical particles, and yet that it also contains consciousness? How can a mechanical universe contain intentionalistic human beings —that is, human beings that can represent the world to themselves?[296]

Searle n'accepte pas les thèses radicales proposées par des formes de matérialisme extrême selon lesquelles les énoncés décrivant des événements mentaux n'ont pas de signification ni d'importance, il n'accepte pas non plus les thèses du béhaviorisme logique d'après lequel les énoncés mentalistes sont toujours synonymes d'énoncés physicalistes.

Searle n'est pas d'accord avec les thèses matérialistes qui tout en admettant que les énoncés mentalistes ont leur propre signification les considèrent toujours faux et trompeurs. Il admet qu'il y a une relation causale réciproque entre les événements physiques et les événements mentaux. Pour lui, il est difficile de croire qu'on peut changer sur le plan mental sans changer sur le plan physique.

Mais nous ne pouvons pas dire que pour lui tout ce qui existe, y compris ce qu'on appelle l'esprit peut être réduit à une réalité physique. Il défend une théorie de l'identité entre le mental et le physique, mais il ne tombe pas dans une philosophie de l'esprit réductioniste. Il défend une sorte particulière de monisme.

Searle ne croit pas qu'il existe une opposition si importante entre le physicalisme et le mentalisme (naïfs); à son avis ces deux conceptions peuvent être conciliées. Il affirme que nous pouvons situer sa position sur l'esprit comme une position intermédiaire entre:

---

296 J.R. Searle (1984), *op. cit.* , p.13.

(a) une sorte de physicalisme naïf défini, par lui, comme un modèle physique d'explication de l'univers à partir du langage de la physique.

(b) le mentalisme naïf qu'il définit comme la conception selon laquelle les phénomènes mentaux existent réellement et possèdent une force causale capable de déterminer les événements physiques.

Searle croit à la possibilité de concilier ces deux formes naïves de physicalisme et de mentalisme; pour lui elles ne constituent pas des conceptions qui s'excluent mais elles sont plutôt compatibles, cohérentes et vraies.

Les critiques de Searle envers l'IA ont pour objet principal les rapports entre la syntaxe et la sémantique et les capacités sémantiques des ordinateurs digitaux: selon lui on n'a pas de machines vraiment intelligentes, car on ne peut pas simuler les capacités de l'esprit à l'aide de simples manipulations formelles.

Dans la première partie de ce chapitre nous allons présenter quelques notions de Searle sur l'esprit et sur le langage: nous allons commencer en présentant les caractéristiques des phénomènes mentaux telles qu'exposées par Searle, à savoir: l'intentionnalité, la conscience, la subjectivité. Nous allons discuter des rapports entre l'intentionnalité des phénomènes mentaux et le langage, et la question du rapport entre le cerveau et l'esprit. Ces éléments théoriques sont présentés afin de faire ressortir les critiques searléennes sur les machines intelligentes. Cela nous permettra de comprendre comment la position philosophique de Searle sur les rapports entre l'esprit et le langage est liée à ses critiques à l'IA. Dans la deuxième partie de ce chapitre nous allons présenter les critiques de Searle à l'IA et à la science cognitive.

1- Les rapports entre le cerveau et l'esprit

Searle affirme qu'avec les développements des sciences physiques, les entités mentales sont dévalorisées. Toutes les conceptions modernes sur ces entités ont pour but soit de déprécier leur influence sur le comportement humain et sur le corps, soit de les nier complètement.

> So, most of the recently fashionable materialist conceptions of the mind such as behaviorism, functionalism, and physicalism—end up by denying, implicitly or explicitly, that there are any such things as minds as we ordinarily think of them. That is, they deny that we do really *intrinsically* have subjective conscious mental states and that they are as real and as irreducible as anything else in the universe[297].

D'une part, nous avons une conception fortement mentaliste de l'esprit qui remonte à la tradition cartésienne et, d'autre part, nous avons une conception rationnelle, scientifique de la réalité selon laquelle tous nos comportements et toutes nos pensées sont le résultat des interactions entre des systèmes entièrement physiques.

> Dualists, who correctly perceive the causal role of the mental, think for that very reason they must postulate a separate ontological category. Many physicalists who correctly perceive that all we have in our upper skulls is a brain think that for that reason they must deny the causal efficacy of the mental aspects of the brain or even the existence of such irreducible mental aspects [298].

Comme nous l'avons vu, le matérialisme (dans sa forme extrême) est proposé dans les termes suivants: tout ce qui existe peut être réduit à une réalité physique. Les énoncés qui décrivent des événements mentaux n'ont pas de signification. Searle n'accepte pas ces thèses radicales. Il est aussi contre le behaviorisme logique selon lequel les énoncés mentalistes sont toujours synonymes d'énoncés physicalistes. Searle n'est pas d'accord non plus avec les thèses matérialistes moins radicales, qui tout en admettant que les énoncés mentalistes ont leur propre signification, les considèrent toujours faux et trompeurs.

Selon Searle, nous dénions l'importance de l'esprit lorsque nous pensons que pour parler des phénomènes mentaux nous devons utiliser seulement des énoncés physicalistes. Nous ne pouvons pas non plus réduire notre compréhension des phénomènes mentaux à

---

297 J.R. Searle (1984), *op. cit.* , p. 15

298 J.R. Searle, *Intentionality: an Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge, 1983.

une description entièrement mentaliste[299], c'est-à-dire, dénier complètement la réalité causale des événements mentaux.

Le mental et le physique ne sont pas, pour Searle, deux substances différentes; il rejette toute forme de dualisme, tout en évitant de tomber dans un matérialisme radical. Pour Searle l'esprit et la matière sont une seule chose qui a des manifestations à deux niveaux différents.

1.1- Les caractéristiques des phénomènes mentaux

Pour Searle le cerveau cause l'esprit. Il est important, selon lui de comprendre ce qui se passe lorsque les états mentaux se produisent et quelles relations ils ont avec la matière qui les cause. Toutefois, pour comprendre les relations entre le cerveau et l'esprit il faut rendre compte préalablement des caractéristiques essentielles de l'esprit. Cela a toujours posé des difficultés pour ceux qui ont voulu expliquer les relations entre le cerveau et l'esprit.

Searle affirme, dans *Minds, Brains and Science*, que l'esprit humain a quatre caractéristiques,qui constituent la source de difficultés lors qu'on veut rendre compte des rapports entre le corps et l'esprit, ce sont:

1) l'Intentionnalité, qui est une caractéristique par laquelle nos états mentaux renvoient à, ou portent sur des objets et des états de choses dans le monde (l'intentionnalité concerne aussi d'autres états mentaux différents de l'intention proprement dite)[300];

2) la conscience: les êtres humains sont conscients de la plupart de leurs pensées et de leurs états mentaux;

3) la subjectivité des états mentaux:les êtres humains ne peuvent avoir accès à d'autres esprits pour décrire objectivement leurs phénomènes mentaux;

---

299 Comme nous avons vu dans le chapitres précédents où nous avons mentionné le problème du corps et de l'esprit pour quelques formes de dualisme l'esprit se caractérise, par des choses mentales : états mentaux comme les désirs, les croyances, les intentions, etc. tandis que le corps est une chose matérielle qui est faite de ou est le produit de choses physiques, telles que la voix, les caractères écrits sur cette page etc. Ces deux choses dans la perspective dualiste à laquelle Searle s'oppose necessitéraient d'un vocabulaire distinct pour être décrites.

300 J.R. Searle (1983), *op. cit.* , pp.1-36. Le mot intentionnalité vient du mot latin, *intentio* qui a éte employé par les scolastiques à l'époque medivalle, recuperé contemporainement par F. Brentano. Il est utilisé largement par les philosophes de la phénoménologie et de la philosophie analytique. L'intentionnalité est comprise par Searle, dans *Intentionality* comme renvoi (directedness), elle est une propriété que les états et événements mentaux ont qui leur permet d'être "à propos" (aboutness) de quelque chose du monde. Elle ne caractérise pas tous les états et événements mentaux, mais seulement quelques uns d'entre eux.

4) l a causalité du mental, qui est une caractéristique de l'esprit de pouvoir affecter causalement le corps et le monde.

Afin de mettre en évidence la nullité de la distinction dualiste cerveau-esprit, Searle propose, dans *Minds, Brains and Science* de mettre au clair les principales caractéristiques des phénomènes mentaux mentionnées plus haut. Il fait l'affirmation suivante à ce propos: " They are so embarrassing that they have led many thinkers in philosophy, psychology, and artificial intelligence to say strange and implausible things about the mind"[301].

Dans *Intentionality, Expression and Meaning* et dans *Minds Brains and Science*, Searle aborde plusieurs questions liees à la philosophie de l'esprit afin de rendre compte de certains problèmes concernants le langage, plus précisément sur la sémantique. Pour Searle la philosophie du langage est une branche de la philosophie de l'esprit. Dans sa discussion sur l'IA[302], ses arguments ont comme base ses conceptions sémantiques, les thèses sur l'Intentionnalité et sur les rapports entre le cerveau et l'esprit.

Nous allons passer à la discussion des quatre caractéristiques de l'esprit mentionnées plus haut. Nous commencerons par un aperçu sur l'intentionnalité, afin de montrer comment elle est en relation avec les conceptions linguistiques de Searle. La raison de ce survol sur la notion d'intentionnalité est qu'il permettra de mieux comprendre, lorsque nous traiterons des critiques de Searle à l'IA, comment les thèses intentionnalistes et les conceptions de Searle sur le langage le font critiquer les travaux sur les machines intelligentes.

1.1.1- L'Intentionnalité des phénomènes mentaux et du langage

Certains états mentaux pour Searle ont une caractéristique fondamentale: ils sont Intentionnels. Il croit que les états mentaux ont un contenu et sont souvent dirigés vers des objets. L'étude de l'Intentionnalité peut nous aider, selon lui, à comprendre le fonctionnement de l'esprit en tant que capacité du cerveau de mettre le corps en relation avec le monde et de donner des significations aux choses.

Une des caractéristiques des états mentaux est le renvoi (*directedness*). Selon Searle on peut appliquer un test très simple pour reconnaître le renvoi ou, autrement dit, pour vérifier

---

301 J.R. Searle (1984), *op. cit.* , p.15.
302 J.R. Searle (1984), *op. cit.*

si un état mental est ou non Intentionnel. Selon lui, un état mental S est Intentionnel si on peut répondre à certaines questions sur cet état telles que: S à propos de quoi? S sur quoi ?

Searle explique les thèses sur les états mentaux à partir du modèle de l'acte de discours. Pour lui il y a des affinités et des liens entre les états Intentionnels et les actes de discours. Il entend par là que l'Intentionnalité peut être expliquée en termes linguistiques et utilise sa théorie des différents types d'actes de discours pour expliquer la notion d'Intentionnalité. Les actes de discurs, explique-t-il, sont marqués par des rapports intentionnels.

L'Intentionnalité, selon Searle, se présente sous deux aspects, à savoir:

1) L'*Intentionalité dérivée,* qui caractérise les énonciations. Par exemple, l'utilisation des phrases du langage naturel. Pour Searle l'expression linguistique, c'est-à-dire, les phrases d'une langue (présentées soit sous la forme de sons ou des marques sur du papier) sont des objets du monde comme cette chaise ou cette table, mais ils sont des objets qui servent à représenter le monde et notre pensée. La capacité de représentation de ces objets phonétiques et graphiques du langage n'est pas intrinsèque à eux, mais dérivée de l'Intentionnalité qui est une caractéristique essentielle de l'esprit humain. Le problème de savoir comment l'esprit impose l'Intentionnalité à des entités physiques est appelé, *problème de la dérivation*[303].

2) L'*Intentionalité intrinsèque*, qui caractérise les états mentaux, par exemple, les croyances. Pour Searle, le fait que les états mentaux peuvent être exprimés par le langage ne signifie pas qu'ils sont des objets syntaxiques ou qu'ils ont une Intentionnalité dérivée, car leur Intentionnalité, (et leur capacité représentative) n'est pas dérivée de formes plus primitives d'Intentionnalité, mais elle est intrinsèque à ces états mêmes. Nous ne nous servons pas de nos états mentaux pour agir sur le monde comme nous utilisons le langage. Les états mentaux, on les a. Un point c'est tout.

> The intentionality of mental states, on the other hand, is not derived from some more prior forms of intentionality but is intrinsic to the states themselves. An agent uses a sentence to make a statement or ask a question, but he does not in that way *use* his beliefs and desires, he simply has them[304].

---

303 Searle dit que l'Intentionalité derivée résulte de l'intention avec laquelle l'acte est accompli. Le prblème de la dérivation conserne l'aspect sémantique du langage naturel., pour le résoudre il faut expliquer comment nous donnons une signification aux choses, comment on peut imposer de l'intentionalité à des entités physiques qui ne sont pas intrinsèquement intentionnelles. Autrement dit, il faut comprendre comment ont peut passer de l'énonciation, niveau physique (les sons qui sortent de la bouche du locuteur) au niveau sémantique de l'acte illocutoire.
304 J.R. Searle (1983), *op. cit.* , p. vii.

Searle fait une remarque qui nous aide à bien déterminer sa notion d'Intentionnalité. Selon lui il y a une fausse parenté entre le terme intentionnel avec-un-i minuscule, lié à la notion d'avoir l'intention (*intending*)[305] et Intentionel avec un I majuscule, lié à la notion d'Intentionalité comme renvoi.

> (...) intending and intentions are just one form of Intentionality among others. They have no special status. The obvious pun on "Intentionality" and "intention" suggests that intentions in the ordinary sense have some special role in the theory of "Intentionality"; but on my account intending to do something is just one form of Intentionality along with belief, hope, fear, desire, and lots of others; and I do not mean to suggest that because, for example, beliefs are Intentional they somehow contain the notion of intention or they intend something or someone who has a belief must thereby intend to do something about it[306].

Identifier le concept d'Intentionnel à l'intention avec-un-i- minuscule peut bien nous tromper au point de penser, comme quelques auteurs, que certains états Intentionnels sont des actes mentaux. Pour distinguer ce qui est un acte mental, il nous donne quelques exemples qui servent bien à montrer la différence entre les actes mentaux, les actions et les états intentionnels:

ACTIONS: 1- Écrire des livres,
2- Boire de la bière etc.

ACTES MENTAUX: 1- faire un calcul mental,
2- former des images mentales d'un object.

ÉTATS INTENTIONNELS: 1- Croire à quelque chose,
2- Avoir l'intention de...
3- Craindre, etc.

Dans le passage qui suit, Searle signale bien l'importance de faire la distinction entre Intentionnel et intentionnel:

---

305 Avoir l'intention (intending) n'est qu'une parmi plusieurs formes d'intentionalité. Il vaut remarquer aussi que d'autres philosophes analytiques utilisent la notion d'intentionalité comme base à la discution de problèmes logico-linguistiques, Searle aussi, mais pour lui, les contextes intensionnels ne sont pas de contextes intentionnels, l'intensionalité est un concept metalinguistique qui selon Searle n'a rien avoir avec l'Intentionalité.

306 J.R.Searle (1983), *op. cit.* , p.3. Dans la tradiction philosophique cette caracteristique Intentionnelle de l'esprit de renvoier à des objets et à des états de choses dans le monde est compris de plusieurs façons. Le mot "Intentionalité" a plusieurs emplois, Searle dit que cela fait du terme Intentionalité un terme trompeur tel qu'il est employé par cette tradiction.

> Acts are things one does, but there is no answer to the question, "What are you now doing?" which goes, "I am now believing it will rain", (...)The Intentional states and events we will be considering are precisely that: states and events; they are not mental acts, (...) [Searle reitere cela en disant:]
>
> It is equally confused to think of, for example, beliefs and desires as somehow intending something. Beliefs and desires are Intentional states, but they do not intend anything. On my account "Intentionality" and "Intentional" will occur in these noun and adjective forms, and I will speak of certain mental states and events as having Intentionality or as being Intentional, but there is no sense attaching to any corresponding verb[307].

Étant donné que l'Intentionnalité n'est pas une caractéristique de tous les états et événements mentaux, Searle propose un critère d'identification des états mentaux. Pour identifier un état Intentionnel on doit, selon Searle: (1) Identifier le *mode psychologique* que lui est caractéristique (s'il s'agit d'un désir, d'une croyance, d'une crainte etc.) et sa *direction d'ajustement*[308]. (2) Identifier le *contenu* de cet état Intentionnel..

Les notions de *contenu représentationnel*, *mode psychologique* et de *direction d'ajustement* sont des notions analogues à celles que nous trouvons dans la *théorie des actes de discours* de Searle. Il y a quatre types de rapports ou d'affinités entre les états intentionnels et les actes de discours à savoir:

1) La distinction trouvée dans la théorie des actes de discours entre *Force Illocutoire* et *Contenu Propositionnel* est aussi appliquée aux états Intentionnels.

Comme nous le savons, la forme logique attribuée par Searle aux actes illocutoires est la suivante, F(p), où «F» est la force illocutoire et «p» est le contenu propositionnel de l'acte. Cela donne formellement la fonction F(p) qui permet de représenter une attitude propositionnelle du type, X aime les oranges comme "Aime (oranges)". De la même façon, tout état intentionnel est configuré ainsi: il a un contenu représentationnel (r) sous un certain mode psychologique S. Cela donne formellement la fonction S(r) qui permet de représenter un état intentionnel tel que X aime p.

2) La distinction faite dans la théorie des actes de discours entre les différentes directions d'ajustement est également appliquée à la théorie des états Intentionnels.

---

307 *Idem.*

308 Selon Searle la notion direction d'ajustement (*direction of fit*) designe le rapports entre le langage (ou la pensé) et le monde. Les différentes directions d'ajustement servent à spécifié les conditions de satisfaction d'un acte illocutoire et s'appliquent aussi bien à l'analyse de énonciations que a celle des états mentaux.

Le mode psychologique S d'un état intentionnel détermine la direction d'ajustement de cet état. De la même façon que la Force illocutoire F d'un acte illocutoire détermine la direction de l'acte.

Il y a une similarité entre les différentes directions d'ajustement des actes de discours et celles des états intentionnels. Par exemple: Un acte illocutoire assertif a la direction d'ajustement qui va des mots aux choses, c'est-à-dire, pour être satisfait il doit être conforme aux choses. Un état intentionnel de croyances à son tour, a la direction d'ajustement qui va de l'esprit au monde; pour être satisfait il doit représenter correctement le monde. Un acte illocutoire promissif ou une requête a la direction d'ajustement qui va de choses aux mots de même le désir qui lui correspond a la direction d'ajustement qui va du monde à l'esprit.

3) Il y a un lien nécessaire entre les états Intentionnels et les actes de discours. L'exécution d'un acte de discours exprime nécessairement un état intentionnel correspondant à chaque type d'acte de discours. Par exemple, en accomplissant un acte illocutoire d'un certain contenu propositionnel on exprime généralement un certain état Intentionnel relatif à ce contenu propositionnel que Searle appelle la Condition de Sincérité de ce type d'acte. Ainsi, faire l'affirmation que p, implique un état intentionnel, la croyance que p.

4) La notion de condition de satisfaction trouvée dans la théorie des actes de discours s'applique aussi bien aux actes illocutoires qu'aux états Intentionnels, à la condition qu'ils aient une direction d'ajustement.

Lorsque, dans la théorie des actes du discours on dit qu'un acte de discours est réussi ou qu'il a raté en disant, par exemple, qu'une promesse a été tenue ou enfreinte, qu'une affirmation est vraie ou fausse, on signale au même temps que l'acte illocutoire a réussi ou a échoué à se conformer au réel conformément à la direction d'ajustement de son but illocutoire. Cette même chose se passe avec les états intentionnels, lesquels se conforment ou non au monde selon leur direction d'ajustement.

Comme nous pouvons le remarquer, le schéma explicatif des actes de discours vaut aussi pour les états intentionnels. Ainsi, si pour les actes de discours le contenu propositionnel (p) détermine certaines des conditions pour que l'acte soit satisfait, de la même façon, dans le cas des états intentionnels le contenu représentatif (r) détermine l'ensemble des conditions pour que cet état intentionnel soit satisfait. En faisant

intentionnellement une énonciation nous imposons de l'Intentionnalité à nos énonciations, parce que c'est intentionnellement que nous conférons les conditions de satisfaction des états psychologiques qui concernent l'acte de discours.

> What is crucially important to see is that for every speech act that has a direction of fit the speech act will be satisfied if and only if the expressed psychological state is satisfied, and the conditions of satisfaction of speech acts and expressed psychological state are identical. Thus, for example, my statement will be true if and only if the expressed belief is correct, my order will be obeyed if and only if the expressed wish or desire is fulfilled, and my promise will be kept if and only if my expressed intention is carried out. Furthermore, notice that just as the conditions of satisfaction are internal to the speech act, so the conditions of the satisfaction of the Intentional state are internal to the Intentional state[309].

Les propriétés logiques des états intentionnels résultent de leur nature représentative. Les états intentionnels pourvus d'une direction d'ajustement peuvent être connus sans l'investigation de ces catégories ontologiques; il suffit de savoir quelles sont ses conditions de satisfaction et sous quels aspects ces conditions de satisfaction sont représentées dans le contenu intentionnel, et de savoir aussi quel est le mode psychologique de l'état en question.

Searle affirme que nous ne devons pas nous préoccuper de la catégorie ontologique de l'Intentionnalité, mais plutôt de ses propriétés logiques:

> At this stage the question of how Intentional states are realized in the ontology of world is no more a relevant question for us to answer than it is relevant for us to answer the analogous questions about how a certain linguistic act is realized.(...) The forms of realization of an Intentional state are just as irrelevant to its logical properties as the forms in which a speech act is realized are irrelevant to its logical properties[310].

Selon Searle les états intentionnels sont dépourvus d'individualité et les conditions de satisfaction des états intentionnels ne sont pas déterminées de façon indépendante, mais sont dépendantes des autres états du réseau d'états mentaux. Ces idées dérivent de l'idée qu'un état intentionnel est attaché à tout un réseau d'autres états intentionnels et est basé sur un arrière-plan de capacités mentales non représentatives.

Tous nos états mentaux se retrouvent à l'intérieur d'un réseau composé par d'autres états mentaux. Le réseau s'appuie sur un arrière-plan (*background*). Les relations entre les

---

309 J.R. Searle (1983), *op. cit.* , pp.10-11.
310 J.R. Searle (1983), *op. cit.* , p.15.

états intentionnels et le monde dépendent du Réseau et de l'arrière-plan. Quand on réalise un acte de discours, quand on affirme quelque chose, par exemple, on exprime un état mental, celui-ci est à l'intérieur du réseau d'états mentaux et s'appuie sur une base pré-intentionnelle, l'arrière-plan. Ainsi, toute explication de l'Intentionnalité est circonscrite dans l'ensemble conceptuel intentionnaliste formulé par Searle.

Les états Intentionnels, en faisant partie d'un réseau d'autres états intentionnels, ne possèdent leur condition de satisfaction qu'en fonction de leur intégration au réseau d'états mentaux; ils sont dépourvus d'individualité, dit Searle (le réseau est holiste).L'arrière-plan non représentatif permet le fonctionnement des représentations et la détermination de ses conditions de satisfaction.

La notion d'Intentionnalité est fondamentale pour les arguments sémantiques de Searle, qui seront dirigés contre l'IA. Pour Searle les machines et les programmes conçus jusqu'à présent ne sont pas capables de rendre compte de l'aspect sémantique concernant le langage et la pensée. Elles ne sont pas capables d'être engagées dans des vraies situations conversationnelles comme les locuteurs et les allocutaires humains. Par exemple, selon lui, pour rendre compte théoriquement de la signification du locuteur, il faut être capable de déterminer les relations entre les états mentaux exprimés par le locuteur et les autres états mentaux du réseau aussi bien que d'être capable de déterminer l'arrière-plan, tâche selon lui impossible.

Les liens entre les états intentionnels et les actes de discours montrent bien les rapports entre le langage et la pensée; pour Searle les états intentionnels et les actes de discours sont des formes de représentation analogues. Les actes de discours et les états intentionnels représentent selon Searle les objets et les états de choses du monde, mais ils le font cependant de façon différente et par de moyens différents.

Searle explique l'Intentionnalité par intermédiaire de sa théorie des différents types d'actes de discours; pour lui l'Intentionnalité est à la base de notre capacité de Représentation[311].

> (...)Intentional states represent objects and states of affairs in the same sense of "represent" that speech acts represent objects and states of affairs (event though, (...) speech acts have a derived form of Intentionality and thus represent in a different manner from Intentional states, which have an intrinsic form of Intentionality) (...) We even have something of a theory about these various

311 La notion de réprésentation employé par Searle est bien particulier au *corpus* théorique qu'il a developé à partir de plusieurs travaux sur le langage et sur l'esprit. Cette notion n'a rien à voir avec la plus part des notions actuelles de reprsentation trouvés en Philosophie, I.A. et Psychologie cognitive. La représentation est définie par son contenu et par son mode et non par sa structure formelle comme dans certains thèses fonctionalistes et cognitivistes.

> types of speech acts; and I am going to tap this prior knowledge to try to explain how and in what sense Intentional states are also representations[312].

Même si Searle explique l'Intentionnalité en termes linguistiques, pour lui, l'Intentionnalité n'est pas essentiellement linguistique; il nous rappelle que c'est le langage qui est dérivé de l'Intentionnalité. Il explique le langage en termes intentionnalistes et l'Intentionnalité en termes de langage; son analyse est donc circulaire[313].

La notion d'état intentionnel comme représentation dans les travaux de Searle est en rapport avec les notions de contenu propositionnel, de direction d'ajustement, etc. Lorsque Searle dit que les états mentaux sont représentationnels; il croit que l'intentionnalité comme représentation est analogue au modèle des actes de discours: Il veut dire que tout état Intentionnel a un contenu propositionnel et un mode psychologique. Le contenu propositionnel, comme dans le cas des actes de discours, détermine pour l'état intentionnel l'ensemble des conditions de satisfaction et le mode psychologique détermine la direction d'ajustement du contenu propositionnel.

La notion de représentation n'a pas d'importance capitale pour la théorie de l'Intentionnalité proposée par Searle. Pour lui cette notion est accessoire et pourrait être substituée à une autre notion en rapport avec les aspects théoriques fondamentaux pour la compréhension des états intentionnels et des actes de discours. La notion de représentation de Searle n'est pas celle employée par la tradition philosophique; elle est au contraire, bien particulière au *corpus* théorique qu'il a développé dans ses travaux sur l'esprit et sur les actes de discours. Cette notion n'a rien à voir avec la pluspart des notions de représentation trouvés en Philosophie, I.A. et Psychologie cognitive. Dans *Intentionality* l'usage de la notion de représentation est expliquée de la façon suivante:

> (...) my use of this term differs both from its use in traditional philosophy and from its use in contemporary cognitive psychology and artificial intelligence. When I say, for example, that a belief is a representation I am most emphatically not saying that a belief is kind of picture, nor am I endorsing the *Tractatus* account of meaning, nor am I saying that a belief re-presents something that has been presented before, nor am I saying that a belief has a meaning, nor am I saying that it is a kind of thing from which one reads off its conditions of satisfaction by scrutinizing it. The sense of "representation" in question is meant to be entirely exhausted by the analogy with speech acts: the sense of "represent" in which a belief represents its conditions of satisfaction is the same sense in which a statement represents its conditions of satisfaction[314].

---

312 J.R. Searle (1983), *op.cit.* , pp.4-5.
313 C.f. J.R. Searle (1983), *op.cit.* , p.5.
314 J.R. Searle (1983), *op.cit.* , p.p. 11-12.

Les représentations sont définies, dans ce sens, par leur contenu et leur mode, non par leur structure formelle. L'analyse de l'Intentionnalité n'est pas faite par des moyens formels; Searle n'énonce pas des conditions nécessaires et suffisantes qui caractérisent l'Intentionnalité, en termes de notions plus simples. L'Intentionnalité n'est pas considérée comme un trait logiquement complexe qui résulte de la combinaison d'éléments plus simples tels qu'il est proposé dans les modèles syntaxiques de l'esprit et par certaines thèses fonctionnalistes en IA.

> Furthermore, my use of the notion of representation differs from its use in contemporary artificial intelligence and cognitive psychology. For me a representation is defined by its content and its mode, not by its formal structure. Indeed, I have never seen any clear sense to the view that every mental representation must have a formal structure in the sense, for example, in which sentences have a formal syntactic structure[315].

Pour Searle l'Intentionnalité est un aspect important des phénomènes mentaux qui peut être analysé de façon objective, les autres caractéristiques des phénomènes mentaux mentionnées, telles que la conscience et la subjectivité et la double causalité qui relie les événements mentaux aux événements physiques, ne constituent pas, non plus un obstacle à l'explication objective des phénomènes mentaux. Pour montrer cela nous passons à l'examen des autres caractéristiques des états mentaux lesquelles seront présentées à l'intérieur des sections suivants ou nous traiterons:

1) de l'importance de la conscience et comment elle peut, selon Searle avoir une existence ;

2) de la subjectivité des états mentaux: nous verrons comment, selon Searle, il est possible de concilier la subjectivité des états mentaux avec une conception objective du monde réel;

3) des rapports causaux entre le cerveau et l'esprit et le pouvoir causal de l'esprit: nous exposerons les explications de Searle sur la façon dont le cerveau peut, tout en étant physique, engendrer des états mentaux et d'autre part, comment les événements mentaux peuvent provoquer des événements physiques.

---

315 J.R. Searle (1983), *op.cit.* , p. 12.

1.1.2- L'importance de la conscience

Pour Searle, la conscience est une des caractéristiques les plus importantes des phénomènes mentaux; il affirme qu'un univers dépourvu de la conscience serait dépourvu aussi de toute signification. La conscience joue un rôle fondamental, c'est elle qui permet que le monde soit plein de signification pour nous.

Searle esquisse sa notion de conscience par rapport à sa notion d'Intentionnalité; pour lui, tout état mental est nécessairement un état qui est potentiellement ou actuellement conscient. Il y a une liaison entre l'Intentionnalité et la conscience qui est donnée par la configuration même (*aspectual shape*) de l'état intentionnel, c'est-à-dire, ce que lui permet d'être intrinsèquement intentionnel: d'être pensable ou dont on peut faire l'expérience de quelque façon. Cela est expliqué de la façon suivant par Searle:

> The argument for this thesis is a bit complex, but the central idea behind it can be given a simple formulation: the concept of an intrinsic intentional mental state is the concept of something that has an aspectual shape. All representation is under aspects. You can see this, if it is not obvious on its face, by reminding yourself that mental contents are possible or actual contents of *thoughts* or *experiences*. What you can believe, you have to be able to think; and what you can perceive, you have to be able to experience perceptually. But the notions of thinking and experiencing are notions which imply the presence of aspectual shapes and that in turn implies accessibility to consciousness. The link, then, between intentionality and consciousness lies into the notion of an aspectual shape. To be intentional, a state or process must be thinkable or experienceable, and to be thinkable or experienciable, it must have an aspectual shape under which it is at least in principle, consciously thinkable or experienciable. It must be the sort of thing that could be the content of a conscious thought or experience[316].

Searle admet qu'il y a un rapport important entre les notions d'Intentionnalité et de conscience. Cependant il n'est pas d'accord avec la thèse selon laquelle il y a une identité entre conscience et Intentionnalité. Cette thèse est exprimée ainsi: Si toute Intentionnalité est considérée comme renvoi à quelque chose ou à des états de choses et si toute conscience est conscience de quelque chose, alors il y a une identité entre les états intentionnels et les états conscients. Prenons, par exemple, l'état intentionnel B (identifié par le fait qu'avoir peur c'est avoir peur de quelque chose) et l'état conscient C, (être conscient de quelque chose). Selon Searle certains auteurs disent qu'il y a une identité entre

---

316 J.R. Searle, "Consciousness, unconciousness, and Intentionality", *Philosophical topics*, volume XVII, nº1, spring 1989, p.198.

les états intentionnels et les états conscients exprimés par le terme «de». Pour eux le «de» exprime l'identité B=C parce que B et C renvoient à quelque chose ou à des états de choses.

Pour Searle l'identité en question pose des problèmes, car le «de» de l'Intentionnalité (peur de) est carrément différent du «de» d'une expérience consciente (conscience de). Supposons, dit Searle, qu'on fait l'expérience consciente de l'inquiétude, on est à ce moment-là conscient de quelque chose: l'inquiétude. L'expérience consciente de l'inquiétude et ce à quoi cette expérience renvoie (l'inquiétude) sont identiques. De l'autre coté, pour montrer que B et C ne peuvent être identiques, Searle examine le «de» intentionnel: un état intentionnel exprimé par l'affirmation " jai peur des serpents" n'est pas identique aux serpents. Il y a dans ce cas une distinction entre l'état intentionnel et la chose sur quoi cet état porte.

Cette "conscience de", est basée sur une identité entre l'état conscient et ce à quoi cet état renvoie, tandis que le "de" Intentionnel dans "Peur des serpents" renvoi à d'autre chose qui n'est pas l'état intentionnel lui-même. Selon Searle, il ya une distinction entre l'état Intentionnel et l'object intentionnel.

Selon Searle il existe des états Intentionnels qui n'exigent pas l'existence de l'objet ou de l'état de choses auquel ces états intentionnels renvoient. Ainsi le «de» intentionnel et le «de» conscience ne sont pas identique, par exemples:

1) Je peux espérer qu'il pleuve même s'il ne pleut pas,

2) Je peux croire que le roi de France est chauve même si il n'existe plus aucun roi de France.

D'un autre côté, on peut avoir des états conscients non-Intentionnels, par exemple, un sentiment soudain d'exaltation.

On peut avoir des états Intentionnels non conscients: par exemple, nous avons une quantité énorme de croyances auxquelles nous n'avons jamais pensé.

Searle affirme que la notion de conscience n'est pas un sujet facile à concilier avec notre conception scientifique du monde selon laquelle tout phénomène fondamental pour notre vie mentale est d'ordre entièrement physique. Ainsi plusieurs théoriciens sérieux s'intéressant aux phénomènes mentaux ne trouvent pas que la conscience constitue un sujet scientifique. La difficulté et l'hésitation à considérer la conscience comme un élément objectif pour l'étude des phénomènes mentaux sont exposées par Searle de la façon suivante:

> It is just a plain fact about the world that it contains such conscious mental states and events, but it is hard to see how mere physical systems could have consciousness. How could such a thing occur? How, for example, could this grey and white gook inside my skull be conscious?[317]

Selon Searle, ce problème peut être surmonté parce qu'il est possible de démontrer l'existence de la conscience. Pour lui, la conscience est quelque chose de réel qui se manifeste physiquement comme n'importe quelles autres caractéristiques humaines, la digestion par exemple. Comme la digestion, la conscience a une explication biologique objective: nous pouvons parler objectivement des processus électro-chimiques qui se produisent au niveau neuronal et qui sont derrière la conscience.

> It should seem no more mysterious, in principle, that this hunk of matter, this grey and whit oatmeal-textured substance of the brain, should be conscious than it seems mysterious that this other hunk of matter, this collection of nucleo-protein molecules stuck onto a calcium frame, should be alive. The way, in short, to dispel the mystery is to understand the processes. We do not yet fully understand the processes, but we understand their general *character*, we understand that there are certain specific electro-chemical activities going on among neurons or neuron-modules and perhaps other features of the brain and these processes cause consciousness[318].

Searle défend que, même si nous comprenons à peine le caractère général des processus neuro-physiologiques qui donnent naissance à la conscience, nous savons toujours quelque chose d'objectif permettant de caractériser les états dits conscients.

1.1.3- Sur la subjectivité des états mentaux

Searle affirme que c'est une erreur de penser que la subjectivité de notre vie mentale est quelque chose qui ne concerne pas la science. Pour lui le fait d'avoir un esprit n'est pas contradictoire avec le fait d'avoir un cerveau. Il n'y a pas d'incompatibilité entre les connaissances philosophiques provenant du sens commun, que nous avons sur l'esprit et les connaissances scientifiques que nous avons du cerveau. Pour Searle, la subjectivité des états mentaux peut être expliquée de façon tout à fait objective. Elle ne doit pas, selon lui, être considérée comme un obstacle à la compréhension des rapports entre le cerveau et l'esprit.

---

317 J.R. Searle (1984), *op. cit.* , p.15.
318 J.R. Searle (1984), *op. cit.* , pp. 23-24.

L'analyse objective d'un objet n'exclut pas sa subjectivité. La subjectivité des états mentaux est un élément de la description de tels états, lequel est un fait scientifique objectif comme n'importe quel autre. Pour Searle, comme pour la plupart des scientifiques, n'importe quel type de faits concernant la réalité peuvent constituer l'objet d'une investigation systématique. La subjectivité est un fait concernant la réalité des états mentaux. Si la subjectivité des états mentaux constitue un problème pour l'explication scientifique des rapports entre le cerveau et l'esprit c'est parce qu'on a une certaine conception de la "science" que nous devons, selon Searle, mettre en question:

> Thus the existence of subjectivity is an objective fact of biology. It is a persistent mistake to try to define 'science' in terms of certain features of existing scientific theories. But once this provincialism is perceived to be the prejudice it is, then any domain of facts whatever is a subject of systematic investigation. So, for example, if God existed, then that fact would be a fact like any other. I do not know whether God exists, but I have no doubt at all that subjective mental states exist, because I am now in one and so are you. If the fact of subjectivity runs counter to a certain definition of 'science', then it is the definition and not the fact which we will have to abandon [319]..

Selon Searle, une conciliation de la subjectivité des états mentaux avec une conception objective du monde réel ne peut être considérée comme un problème que si nous considérons la subjectivité du mental comme quelque quelque chose qui ne doit pas être considérée objectivement. Pour lui, au contraire, la subjectivité des états mentaux est un fait objectif; elle constitue, en plus, un élément important pour la description complète des rapports entre le cerveau humain et l'esprit.

Searle affirme que les processus neuronaux causent notre vie mentale. Selon lui, il n'est pas si difficile d'expliquer, provisoirement, comment les processus neuronaux peuvent donner naissance à des événements mentaux. Il est possible, pour lui, de concilier la connaissance, du sens commun que nous avons de l'esprit avec la connaissance scientifique que nous avons du cerveau.

Tous nos états intentionnels sont causés par des processus cérébraux. Les phénomènes d'intention sont tout simplement matérialisés dans la structure du cerveau humain et se comportent comme n'importe quel autre phénomène naturel qui se prête à une analyse scientifique. Pour expliquer la nature de nos états intentionnels, Searle donne l'exemple de la soif:

319*Idem.* , p.25.

> As far as we know anything about it, at least certain kinds of thirst are caused in the hypothalamus by sequences of nerve firings. These firings are in turn caused by the action of angiostensine in the hypothalamus, and angiostensine, in turn, is synthesized by reins, which is secreted by the kidneys. Thirst, at least of these kinds, is caused by a series of events in the central nervous system, principally the hypothalamus, and it is realized in the hypothalamus[320].

Pour Searle, l'intentionnalité ainsi que la conscience ne représentent pas un «mystère» si nous décrivons de façon minutieuse comment elles sont liées aux processus biologiques qui les provoquent. Certaines catégories de soifs, comme dans l'exemple présentée plus haut, sont causées par des événements physiques cérébraux: la soif, dans ce cas, n'est rien de plus (sur le plan mental) que d'avoir envie de boire. L'état intentionnel "avoir envie de boire" est un état mental qui est à la fois causé par certains processus cérébraux et matérialisé dans la structure du cerveau. Tel état intentionnel a un contenu spécifique lequel détermine les conditions selon lesquelles l'état intentionnel sera satisfait. La soif est un exemple d'état intentionnel qui nous permet de constater aisément comment le cerveau, étant physique, peut engendrer de phénomènes mentaux.

Selon Searle, le pouvoir causal des états mentaux représente une difficulté pour ceux qui veulent expliquer les rapports entre le cerveau et l'esprit. Pour lui, cette difficulté est à peine apparente. Il propose une solution, basée sur une analogie avec la physique, pour expliquer les rapports de causalité entre l'esprit et le cerveau. Une telle explication permet de comprendre aussi comment des événements mentaux peuvent provoquer des événements physiques, c'est-à-dire comment nos pensées peuvent engendrer des actions. Passons tout d'abord aux rapports causaux entre l'esprit et le cerveau.

1.1.4- Les rapports de causalité entre l'esprit et le cerveau

Pour Searle, les états mentaux ainsi que la conscience sont des données objectives parce qu'ils sont une caractéristique évidente de notre vie mentale, laquelle résulte ou est causée par ce qui se passe à l'intérieur de notre cerveau. Selon Searle, l'esprit et le corps agissent l'un sur l'autre, mais ils ne constituent pas deux substances distinctes.

Pour mieux clarifier cette idée Searle prend le cas de la douleur qui est un phénomène mental causé par des processus neuro-physiologiques à l'intérieur du cerveau; il en donne quelques exemples en rapport avec la douleur: 1)certains amputés, par exemple, ressentent des douleurs dans le membre qu'ils ont perdu. Il se produit ainsi dans le cerveau quelque

---

320 *Idem*, p.24.

chose qui provoque la douleur. Celle-ci se produit donc indépendamment des stimuli extérieurs. 2)La stimulation artificielle de certaines régions cérébrales sert à montrer que la douleur s'y localise, peut se répercuter et être sentie dans d'autres parties du corps reliées au champ cérébral affecté. Searle conclut que ce qui est vrai pour la douleur est aussi vrai pour les phénomènes mentaux en général: Les deux exemples mentionnés par Searle pour montrer que les phénomènes mentaux proviennent physiquement du cerveau viennent renforcer les deux propositions suivantes:

> (...)all mental phenomena whether conscious or unconscious, visual or auditory, pains, tickles, itches, thoughts, indeed, all of our mental life, are caused by processes going on in the brain. (...)Pains and other mental phenomena just are features of the brain (and perhaps the rest of the central nervous system)[321]..

En résumé, les deux énoncés affirment que les états mentaux sont causés par le cerveau, mais en même temps ils sont de vraies propriétés de celui-ci. Il explique qu'à première vue la première et la seconde propositions semblent contradictoires parce qu'il paraît en effet peu probable que l'esprit puisse être un trait caractéristique du cerveau et, à la fois, que le cerveau puisse causer l'esprit: si, par exemple, les phénomènes physiques dans le cerveau causent la douleur—événement mental— comment les douleurs peuvent-elles être au même temps une caractéristique du cerveau?

Pour Searle, les deux propositions présentées peuvent être simultanément vraies. Il suffit de sauvegarder la relation causale entre les phénomènes physiques et mentaux tout en conservant l'idée que l'esprit est un trait caractéristique du cerveau. L'auteur affirme que si nous trouvons ces propositions contradictoires c'est parce que nous faisons certaines associations d'idées qui ne correspondent pas à la nature des phénomènes mentaux.

Selon Searle, l'apparente contradiction entre les deux propositions dérive d'une compréhension inexacte de ce qu'elles énoncent et aussi d'une mauvaise compréhension de la notion de causalité. Selon lui nous nous trompons lorsque nous comprenons les relations causales entre le cerveau et l'esprit par le biais d'une notion d'interaction.

> If mental and physical phenomena have cause and effect relationships, how can one be a feature of the other ? Wouldn't that imply that the mind caused itself-the dreaded doctrine of *causa sui* ? But at he bottom of our puzzlement is misunderstanding of causation. It is tempting to think that whenever A causes B there must be two discrete events, one identified as the cause, the other identified

---

321 *Idem.* , pp.18-19.

> as the effect; that all causation functions in the same way as billiard balls hitting each other[322].

Searle affirme que pour comprendre que les deux énoncés cités plus haut ne sont pas opposés, il faut une notion plus subtile de causalité. Une telle notion emprunte à la physique la distinction, entre les micro et de macro propriétés des systèmes physiques. Selon Searle, il y a deux niveaux réels de causalité pour décrire les rapports entre le cerveau et l'esprit: le cerveau et l'esprit constituent un seul système qui à un niveau plus élevé —niveau des états mentaux— peut être décrit en termes de macro-propriétés et à un niveau plus bas — niveaux des états neuronaux— peut être décrit en termes de micro-propriétés par analogie avec les systèmes physiques.

L'analogie avec la physique est expliquée par l'exemple suivant: de la même façon que l'état solide d'une table constitue une macro- propriété de cet objet qui est définie à partir de ses micro-propriétés, c'est-à-dire, les caractéristiques moléculaires et atomiques de la table, nos états mentaux sont des macro-propriétés définies par des micro-propriétés du cerveau. En physique, plusieurs propriétés d'un système peuvent être expliquées causalement par des micro-propriétés: la table est solide parce que (au niveau des micro-propriétés) leurs molécules sont organisées en treillis.

Ces deux niveaux réels de causalité, expliqués en termes de micro et macro propriétés du cerveau, font comprendre qu'il existe une relation réciproque entre les phénomènes mentaux et les phénomènes physiques. C'est-à-dire: nous avons une bonne explication de la façon dont l'esprit et le cerveau agissent l'un sur l'autre, sans être deux choses distinctes, et comment l'esprit peut être à la fois, une propriété du cerveau et causé par lui.

> I want to suggest that this provides a perfectly ordinary model for explaining the puzzling relationships between the mind and the brain. In the case of liquidity, solidity, and transparency, we have no difficulty at all in supposing that the surface features are *caused by* the behavior of elements at the micro-level, and at the same time we accept that the surface phenomena *just are* features of the very systems in question. I think the clearest way of stating this point is to say that the surface feature is both *caused by* the behavior of micro-elements, and at the same time is *realized in* the system that is made up of the micro-elements. There is a cause and effect relationship, but at the same time the surface features are just higher level features of the very system whose behavior at the micro-level causes those features[323].

---

322 J.R. Searle (1984), *op. cit.* , p. 20.
323 *Idem.* , p.21.

Comme nous l'avons vu, le pouvoir causal du système esprit-cerveau peut avoir deux niveaux de descriptions, un niveau mental et un niveau physique; tous les deux ont une causalité réelle. Tout ce qui se passe au niveau de proprietés des processus mentaux est à la fois causé par et matérialisé dans la structure des processus neuronaux ou micro-propriétés. Les états mentaux sont des propriétés réelles du cerveau qui peuvent causer des événements physiques.

> My conscious attempt to perform an action such as raising my arm causes the movement of the arm. At the higher level of description, the intention to raise my arm causes the movement of the arm. But at the lower level of description, a series of neuron firings starts a chain of events that results in the contraction of the muscles[324].

Le modèle de description des rapports causaux entre le cerveau et l'esprit à deux niveaux, est selon Searle adéquat pour expliquer comment un événement mental, étant une propriété réelle du cerveau et causée par lui, peut être aussi la cause de certains événements physiques. Pour Searle, les événements mentaux causent les événements physiques à cause du double caractère de la causalité qui relie les événements mentaux aux événement physiques.

Les caractéristiques de l'esprit telles que l'Intentionnalité, la conscience, la subjectivité, le double rapport de causalité entre le cerveau et l'esprit sont toutes en jeu dans le fonctionnement de notre pensée et lorsque nous utilisons le langage naturel. Un autre aspect important lié à la théorie de l'esprit et du langage de Searle est la capacité représentationnelle de l'esprit humain. Searle affirme que les techniques de programmation actuellement connues, ne sont pas capables de représenter de la même façon que l'esprit des éléments très simples du monde réel. La capacité de représentation demanderait une duplication de l'esprit, alors que les ordinateurs digitaux ne font qu'imiter certains aspects très élémentaires des nos processus mentaux.

2- Les critiques de J. Searle à l'Intelligence Artificielle

Les discussions de Searle sur les limitations de l'IA et de la science cognitive reposent sur des réfutations où l'argument d'ordre sémantique et les caractéristiques de l'esprit humain jouent un rôle important. La duplication de l'esprit par des moyens informatiques

---

324 *Idem.* , p.26.

est un mythe[325], selon Searle. L'IA ne pourrait jamais se réaliser étant donné les conceptions informatiques dont on dispose actuellement.

Comme nous l'avons mentionné, Searle essaie de répondre dans *Minds Brains and Science* à la question (posée par Descartes, par La Mettrie et après par A.Turing): "Les machines peuvent-elles penser?"[326]. Cependant leur façon de poser la même question n'est pas triviale. Les arguments de Searle ne sont pas d'ordre technologique; il ne vise pas les limitations techniques mais les limitations théoriques de l'IA. Lorsqu'il se demande si une machine digitale peut penser, il ne veut pas mettre en question le fait que les machines sont différentes des êtres humains, ou qu'elles sont des simples artefacts[327]. Mais le fait qu'en tant que machines à programme (comme la machine de Turing) définies de façon purement syntaxique, elles sont incapables de comprendre le langage et le monde.

Searle a comme cible, lorsqu'il critique la recherche en IA, une sorte de courant qu'il appelle "l'IA forte" par opposition à un autre courant dit "l'IA faible". La distinction entre les deux courants est la suivante: Pour les chercheurs de l'IA faible l'ordinateur digital est un instrument important capable d'aider dans la compréhension de l'esprit. Il permet de formuler et tester des hypothèses sur son fonctionnement. Les défenseurs de l'IA forte entendent que l'ordinateur n'est pas un simple instrument pour l'explication de la pensée. Pour eux toute machine adéquatement programmée peut littéralement avoir des états

---

325 Cf. "The Myth of computers", *New York Review of Books*, 3-6, avril, 29, 1982.

326 Turing lui même trouvait que cette question pose de contraintes linguistique car il considère difficile de définir précisement la signification des termes "machine" et "penser" (Cf. A. R. Anderson, (1964) *op. cit.* , pp.4-5). Ainsi il substitue à la question "les machines peuvent-elles penser" un jeu qu'il appelle de jeu d'imitation. Ce jeux suggère que l'ordinateur pourrait présenter des sorties semblables à des comportements humains dans une situation de conversation simulée. Dans ce test, un joueur doit essayer de reconnaître s'il parle avec un homme, A ou avec B, une femme (ou une machine) . Le seul moyen dont il dispose est la conversation avec eux par l'intermédiaire d'un clavier et d'un écran d'ordinateur. Il essaye d'identifier les deux joueurs A et B à partir de l'analyse de l'échange des questions et réponses qu'il a avec eux. Selon Turing, si le joueur n'est pas capable de distinguer qui est qui, il ne peut pas être certain que les machines ne pensent pas. A l'époque du test de Turing il existait une croyance générale que tout type de comportement soumis à des règles pourrait être étudié mathématiquement; celá était une idée d'inspiration cybernétique et béhavioriste. Les comportements intelligents, étaient considérés comme des processus qui pourraient être compris et formalisé par de procédures effectives calculables pouvant être traitées par un ordinateur digital.
Turing a suggéré à titre d'expérience que la question « Les machines peuvent-elles penser ? » devrait être remplacée par «Peut-on imaginer des ordinateurs digitaux qui fassent bonne figure dans le jeu de l'imitation ? » Si nous le souhaitons, nous pouvons rendre cette question superficiellement plus générale et demander : « Y a-t-il des machines à états discrets qui puissent y faire bonne figure?» Mais eu égard à la propriété d'universalité, nous voyons que chacune de ces deux questions c'est équivalente à celle-ci : « Fixons notre attention sur un ordinateur digital particulier C. Est-il vrai qu'en modifiant cet ordinateur pour avoir une capacité de mémoire adéquate, en accroissant de manière satisfaisante sa vitesse de travail, et en lui fournissant un programme approprié, on peut faire jouer à C le rôle de A dans le jeu de l'imitation, le rôle de B étant tenu par un homme? (Turing, A. , in A.R. , Anderson, éd, *Pensée et machine*, Ed. du Champ Vallon,1983, Traduit de l'american *Minds and Machines* par Patrice Blanchard *Minds and Machines*, Champ Vallon,. p.48.

327 Searle dit que nous pouvons donner une reponse trivial à la question également triviale sur la pensée de machines. Cette réponse est oui, les machines peuvent penser car nous sommes tous des machines. Nous pouvons aussi affirmer que oui les machines peuvent penser car des qu'elles ont le même pouvoir causal qu'un cerveau humain. Cependant, si on demande "Est-ce qu'un ordinateur digital adequatement programmé (instantiating or implementing the right computer program) peut être intelligent?" La question n'est plus une question triviale et elle exige une réponse non trivial: qui est Non. (cf. J.R. Searle (1984), *op. cit.* , p.36).

cognitifs et, par exemple, comprendre le langage naturel. Les programmes informatiques constituent, en eux-mêmes, des explications de la cognition humaine[328].

Les critiques de Searle à l'IA prennent en considération sa thèse sur l'esprit et ses conceptions sur la sémantique que nous venons d'exposer dans la partie précédente de ce travail. Nous allons maintenant exposer les critiques de Searle à l'IA en prenant en considération quelques éléments importants liés aux rapports entre l'esprit et le cerveau et entre le langage et la notion d'Intentionnalité.

2.1- L'expérience de pensée de la chambre chinoise

Dans son livre, *Minds Brains and Science*, Searle présente encore une fois[329] son expérience de pensée appelée "expérience de la chambre chinoise". Ce "Gedankenexperiment" nous permet de mieux comprendre ses arguments.

Dans son expérience de pensée de la chambre chinoise, Searle nous invite à penser à la situation suivante: imaginons qu'une équipe d'informaticiens a créé, avec succès, un programme informatique capable de permettre à l'ordinateur de "comprendre" le Chinois et de répondre à une série de questions posées dans cette langue.

Sommes-nous en mesure d'affirmer, en considérant les procédures formelles sur lesquelles est basé ce programme, que l'ordinateur peut, littéralement, comprendre le Chinois? Pouvons nous dire qu'un tel programme servirait à expliquer certains processus cognitifs humains en rapport avec la compréhension du langage naturel?

Évitons de répondre immédiatement à ces questions. Imaginons encore, suivant l'expérience de Searle quelqu'un qui joue exactement le rôle mécanique de l'ordinateur à manipuler des symboles. Supposons que nous plaçons une personne ne comprenant pas un seul mot du Chinois dans une pièce isolée, et que cette personne peut répondre à des questions en chinois qui lui sont posées (sous la forme de suites de symboles du Chinois) à l'aide d'un manuel contenant un ensemble de règles syntaxiques du chinois qui lui

---

328 Pour Searle l'IA faible est liées aux recherches appliqués des ingénieurs et techiniciens, tandis que l'IA-forte serait celle de ceux qui font de l'IA une sorte de recherche fondamental.e L'IA faible considère que l'ordinateur est tout simplement un outil pour l'étude de l'esprit, tandis que pour l'IA-forte l'ordinateur, dés que convenablement programmé peut être doué d'un esprit, c'est à dire être dotés d'une intentionnalité et avoir des états mentaux comme les êtres humains. Dans le sens fort l'IA est en rapport avec certaines courants fonctionnaliste en philosophie.

329 Cette expérience est exposée pour la première fois en dans "Minds, Brains and Programs", *The Behavioral and Brain Sciences*, n°3, pp.417-424, Cambridge University Press, 1980, . Ensuite elle est présenté dans l'article"The Myth of the Computer" *New York Review of Books*, 3-6, avril, 29, 1982. et enfin dans *Minds Brains and Science*, chapitre 2, pp.28-41.

permettent de manipuler des symboles chinois gardés dans un panier et de donner des réponses correctes aux questions posées.

Pouvons-nous dire que du fait que la personne dans la chambre chinoise soit capable de répondre avec succès à des questions en Chinois qu'elle comprend (dans le sens littéral de "comprendre"), cette langue? Sommes-nous en mesure d'affirmer, en considérant les moyens sur lesquels la personne s'oriente pour répondre à des questions en chinois, que les procédures de manipulation des symboles du Chinois servent à expliquer l'habileté humaine à comprendre une langue naturelle?

La réponse de Searle à ces quatre questions est "Non". Selon lui ni l'ordinateur adéquatement programmé ni la personne renfermée dans la chambre chinoise ne comprennent un seul mot du Chinois. Ils se comportent, plutôt, *comme si* ils comprenaient le Chinois. Nous ne sommes pas autorisés à dire que, les ordinateurs comprennent *littéralement* la langue chinoise, car les êtres humains, disposant des mêmes moyens formels qu,eux ne la comprennent pas. Autrement dit, Searle veut montrer avec l'expérience de la chambre chinoise que l'homme peut suivre les mêmes moyens formels qu'une machine réelle ou abstraite et ne comprendre rien d'une langue naturelle en manipulant uniquement des symboles sans contenu.

> Now the point of the story is simply this: by virtue of implementing a formal computer program from the point of view of an outside observer, you behave exactly as if you understood Chinese. But if going through the appropriate computer program for understanding chinese is not enough to give you an understanding of chinese, then it is not enough to give any other digital computer and understanding of chinese. And again, the reason for this can be stated quite simply. If you don't understand Chinese, then no other computer could understand Chinese because no digital computer, just by virtue of running a program, has anything that you don't have. All that the computer has, as you have, is a formal program for manipulating uninterpreted Chinese symbols[330].

Searle ne voit pas pourquoi une machine manipulant des symboles dénués de signification puisse faire mieux que nous. Les ordinateurs peuvent exécuter n'importe quel programme, cela ne serait jamais une condition suffisante pour qu'ils comprennent le langage naturel.

Les arguments de Searle sur l'expérience de la chambre chinoise sont opposés à l'argument de Turing sur le "jeu d'imitation". Selon Searle, on ne peut pas assimiler l"intelligence artificielle" à l'intelligence humaine en fonction des sorties (outputs) ou comportements observables de l'homme et d'une machine digitale. Pour J. Searle, même si

---

330 J.R. Searle (1984), *op. cit.* , pp.32-33.

l'homme et la machine répondent de manière identique à des sollicitations identiques ils ne fonctionnent pas nécessairement de la même manière.

Searle defend que l'ordinateur est incapable de dupliquer les capacités cognitives liées à l'habilité linguistique des êtres humains. Dans le cas de la chambre chinoise le manipulateur de symboles comme l'ordinateur fonctionnerait toujours selon une syntaxe.

Searle affirme que la syntaxe ne suffit pas à produire la sémantique. Le langage naturel est une activité qui présuppose syntaxe, sémantique et pragmatique et il est lié à des facultés de l'esprit, lesquelles ne se résument pas non plus à de simples manipulations symboliques sans signification.

Répondre que «oui», (à la questions posée précédemment,) les machines et les hommes comprennent le chinois, serait réduire radicalement, affirme Searle, la sémantique à la syntaxe. Searle remarque que si l'information que l'homme et la machine traitent ne possède qu'un caractère formel il n'y a aucune différence entre la capacité de manipulation des symboles par l'homme et par la machine; l'un et l'autre ne feraient qu'opérer selon des règles syntaxiques, même s'ils faisaient "comme si" ils comprenaient le chinois. La compréhension d'une langue, ne se réduit pas à manipuler formellement des symboles.

Dans *Minds Brains and Science*, Searle mentionne deux objections,faites à ses arguments sur la chambre chinoise: La première objection est que soit dans le cas de la personne dans la chambre chinoise ou soit dans le cas de l'ordinateur programmé pour comprendre le Chinois, le manipulateur de symboles et l'unité centrale de traitement de l'ordinateur ne sont que des parties d'un système plus complexe, lequel dans son ensemble serait capable de comprendre le Chinois,

La deuxième objection est la suivante: Si le système programmé pour comprendre le Chinois était placé dans un robot capable d'intervenir dans le monde de façon causale, on serait en mesure d'affirmer que le système tout entier comprend le chinois. Cette objection s'appuie sur les conceptions linguistiques de Searle selon lequel, pour comprendre une langue, il faut une interaction entre l'individu et le monde[331].

Searle répond à ces objections de la manière suivante: pour lui, les arguments qu'on lui oppose ne constituent pas de vraies objections. À la première objection, Searle répond par l'affirmation que ni même le système tout entier (soit-il composé de la chambre, du panier de symbole, du manipulateur ou d'un micro-processeur, d'un programme, etc.) n'est pas capable de passer de la syntaxe à la sémantique. Si par exemple, une partie du système, le

331 Selon ce dernier le langage est une activité socialle qui exige une interaction entre l'individu et sont environement. Comprendre une langue est comprendre comment le actes linguistiques interviennent dans le monde et cela exige que celui qui comprend soit en contact avec ce monde.

programme ou le micro processeur n'est pas capable d'interpréter les symboles manipulés, le système tout entier ne peut pas faire mieux.

En ce qui concerne la deuxième objection, Searle répond que le robot se comporterait à peine *comme s'*il comprenait le Chinois. Le fait que le robot est en interaction physique avec le monde n'a pas d'importance, car toutes ces actions résultent d'un système de règles avec lesquelles il a été programmé. L'unité responsable des mouvements du corps du robot et de la manipulation des symboles dans le robot ne serait que le résultat de l'exécution d'un programme informatique. L'exécution d'un programme informatique ne suffit pas pour que les symboles manipulés par le robot soient significatifs.

L'interprétation des symboles par le robot demanderait de lui une capacité (intentionnelle) de bâtir des représentations. Cependant les interactions causales du robot avec le reste du monde ne lui permettent pas de représenter ce monde. Le fait d'avoir un corps ne suffit pas, dans ce cas, pour avoir un esprit à moins qu'on considère que l'esprit est un ensemble d'opérations formelles et syntaxiques. Ce avec quoi Searle n'est absolument pas d'accord.

2.2- Les réfutations serleénnes aux deux thèses générales de l'IA

L'argument de la chambre chinoise s'oppose à deux thèses principales de l'IA forte 1º) un ordinateur digital peut *comprendre* le langage naturel ou avoir des *états mentaux* dès qu'il est adéquatement programmé pour faire cela 2º) les programmes informatiques constituent des *explications* plausibles sur les habiletés cognitives humaines. Voyons comment Searle s'oppose à de telles idées:

a)Réfutation de la première thèse: pourquoi les ordinateur digitaux ne peuvent pas comprendre une langue naturelle ou avoir des états mentaux.

Selon Searle, les ordinateurs digitaux sont basés sur une structure formelle et syntaxique.Les sous-programmes de manipulation de langage dans un programme de compréhension du langage naturel comme ELIZA de Weizenbaum, ou le SHRDLU de Winograd ou celui de Shank[332] ne permettent pas aux machines de comprendre le langage naturel. Searle défend que les programmes informatiques basés sur des critères formels d'une machine de Turing ne sont que des systèmes de règles entièrement syntaxiques de manipulation de symboles:

---

332 Cf. R. C.Shank et Abelson, *Scripts Plans Goal and Undertanding*, Lawrence Erbaum Associates, Hilldale, N. J., 1977.

> But this feature of programs, that they are defined purely formally or syntactically, is fatal to the view that mental processes and program processes are identical. And the reason can be stated quite simply. There is more to having a mind than having formal or syntactical processes. Our internal mental states, by definition, have certain sorts of contents[333].

Derrière l'IA forte nous trouvons l'analogie faite par certains philosophes, psychologues et chercheurs en I.A. entre le fonctionnement des ordinateurs digitaux et le fonctionnement du cerveau humain. Selon l'analogie cerveau-machine le cerveau fonctionnerait comme un ordinateur dont le programme correspondrait à notre esprit. " (...) the mind is to the brain, as the program is to the computer hardware"[334]. Sur cette analogie reposent, selon Searle, les thèses fortes de l'IA..

Tout ce que nous programmons sur les machines doit avoir une compatibilité avec la structure binaire du matériel informatique; alors tout ce que nous voulons des machines doit, dans un dernier niveau (ou niveau plus bas de programmation), être programmé selon certaines règles et codé dans une combinaison logique précise par des séquences binaires.(0 et 1) en accord avec le processus «tout ou rien» des circuits électroniques. Programmer c'est *grosso modo* exprimer les étapes d'opération d'une façon logique et encoder toutes ces procédures par une suite de 0 et1 inscrites sur une bande magnétique ou d'autres supports.

Selon Searle les séquences binaires et la syntaxe des langages de programmation informatique qui sont à la base formelle de toutes les procédures informatiques n'ont pas de valeur sémantique. La machine ne traite pas des informations selon nos critères sémantiques. Une machine ne traite pas également des connaissances; elle peut avoir, bien sûr, une base de connaissances mais cela n'implique pas qu'elle va traiter l'information de caractère sémantique. Un programme d'ordinateur n'accomplit que des opérations formellement explicitées; il possède seulement une syntaxe, car il opère des manipulations de symboles purement formelles à la façon d'une machine de Turing:

> A typical computer 'rule' will determine that when a machine is in a certain state and it has a certain symbol on its tape, then it will perform a certain operation such as erasing the symbol or printing another symbol and then enter another state such as moving the tape one square to the left. But the symbols have no meaning, they have no semantic content; they are not about anything. They have to be specified purely in terms of their formal or syntactical structure. The zeroes and ones, for example, are just numerals; they don't even stand for numbers. Indeed, it is this feature of digital computers that makes them so powerful. One and the same type of hardware, if it is appropriately designed, can

333 J.R. Searle (1984), *op. cit.* , p.31.
334 *Idem*, p. 28.

> be used to run an indefinite range of different programs. And one and the same program can be run on an indefinite range of different types of hardwares[335].

Retournant au problème de la chambre chinoise, si l'ordinateur et l'homme en tant que manipulateurs de symboles se comportent comme s'ils comprenaient le chinois cela ne veut pas dire qu'ils comprennent réellement cette langue. Il s'agit d'un processus comme si, dit Searle, et comme tel, n'est qu'une simulation, et ne permet pas une duplication des caractéristiques sémantiques du cerveau humain:

Simuler un processus à partir d'une description formelle de celui-ci n'est pas suffisant pour avoir le processus en question: La simulation d'une explosion nucléaire ou d'un accident automobile ou de la pensée humaine par ordinateur n'arrive pas à répendre de l'irradiation ni causer des dégâts en quoi que ce soit ni permetre non plus d'avoir des états mentaux artificiels. Pour Searle aucune simulation ne peut constituer une duplication des processus mentaux.

> It doesn't matter how good the technology is, or how rapid the calculations made by the computer are. If it really is a computer, its operations have to be defined syntactically, whereas consciousness, thoughts, feelings, emotions, and all the rest of it involve more than a syntax. Those features, by definition, the computer is unable to *duplicate* however powerful may be its ability to *simulate*. The key distinction here is between duplication and simulation. And no simulation by itself ever constitutes a duplication[336].

Le fait que le système se comporte comme s'il y avait une explosion nucléaire ou un choc entre deux automobiles ou comme s'il pensait ne veut pas dire que ces choses simulées se passent réellement à l'intérieur de la machine. Le fait qu'une machine simule l'intelligence humaine ou sa capacité de compréhension du langage naturel ne permet pas d'affirmer qu'elle est intelligente ou qu'elle comprend une langue.

2.2.1- Les limitations sémantiques des ordinateurs

Lorsque, dans son expérience de la chambre chinoise, Searle compare l'homme à l'ordinateur, il veut montrer qu'il n'y a pas de "compréhension" dans le sens littéral, car, dans ces cas, ni l'homme ni la machine n'auraient aucune idée du contenu qui est derrière la suite de symboles qu'ils manipulent. L'homme et la machine exécutent des opérations

---

335 *Idem.* , p. p. 30-31
336 *Idem.* , p. 37.

syntaxiques sur des symboles chinois sans les interpréter; ils sont loin de rendre compte de l'aspect sémantique de tels symboles.

La simple différence entre la syntaxe et la sémantique nous aide à comprendre la thèse anti-fonctionnaliste de Searle selon laquelle les états mentaux ne peuvent pas être réalisés dans n'importe quelle structure physique et avoir un programme n'est ni équivalent ni suffisant pour avoir un esprit:

La syntaxe du langage naturel étant liée seulement aux propriétés formelles de la phrase, ses catégories syntaxiques, le nombre de mots, leur longueur, leur place, etc, ne peut pas permettre, selon Searle, de rendre compte de la sémantique, laquelle est liée à la capacité des phrases d'évoquer des concepts,de nous permettre d'établir des rapports de signification.

Comme nous l'avons vu, les propriétés sémantiques de l'esprit sont en rapport avec l'Intentionnalité. C'est l'Intentionnalité en tant que propriété du cerveau, qui permet que nous donnions des significations aux choses.

Si la manipulation de symboles permet de simuler certains aspects formels de l'intelligence ou de la compréhension d'une langue cela ne veut pas dire que la machine est intelligente ou qu'elle comprend le chinois ou quoi que ce soit. L'intelligence et la pensée, pour Searle, ont un trait fondamental: elles portent sur quelque chose, elles sont basées sur des significations, tandis que l'ordinateur ne manipule que des symboles de façon syntaxique.

Les ordinateurs digitaux sont par définition des machines basées sur des opérations formelles et syntaxiques qui manipulent des symboles sans contenu selon des régles strictes.Ce sont des machines à programme qui se comportent de façon purement formelle (syntaxique). La syntaxe de ses programmes en elle-même n'est pas suffisante pour produire des opérations sémantiques; une machine n'a pas encore la capacité de manipuler des significations, mais seulement des unités syntaxiques (symboles): (...) the computer program is defined purely syntactically. But thinking is more than just a matter of manipulating meaningless symbols, it involves meaningful semantic contents. These semantic contents are what we mean by 'meaning'[337].

Selon Searle, quand nous pensons, nous faisons plus qu'une exécution d'opérations formelles. L'esprit n'est pas un mécanisme syntaxique; au contraire, il se caractérise par sa capacité de donner signification et d'articuler intentionnellement syntaxe et sémantique. Pour cette raison, selon Searle, les ordinateurs ne peuvent pas penser. L'activité de l'esprit

---

337 J.R. Searle (1984), *op. cit.* , *p.* 36.

exige des contenus représentationels; la pensée ne résulte pas de manipulations formelles dénues de signification.

Les arguments de Searle sur la chambre chinoise sont basés sur ses thèses sur l'esprit, à savoir 1) Il y a une relation causale entre les processus mentaux et les processus cérébraux; 2) l'Intentionnalité des états mentaux résulte de certains processus qui ont lieu dans notre cerveau; elle est une des principales caractéristiques de l'esprit nous permettant de penser, d'agir, de percevoir, enfin d'avoir un comportement intelligent; 3) l'Intentionnalité est une donnée objective que nous permet d'expliquer le comportement intelligent.

En résumé les arguments de Searle contre L'IA sont basées sur les points suivants:

(a)La simulation en IA n'est qu'un processus *comme si.* Par exemple, le fait qu'un système informatique fasse comme s'il comprennait une langue naturelle, ne signifie pas qu'il la comprend réellement, dans le sens que les êtres humains le font. Il est impossible, selon Searle, de dupliquer les propriétés de l'esprit.

(b).La syntaxe n'est pas suffisante à l'explication de nos capacités intelligentes lesquelles sont en rapport avec la capacité de l'esprit de donner des significations au monde. Selon Searle les machines digitales ne sont pas capables d'articuler la syntaxe et la sémantique; pour faire cela il leurs faut avoir un esprit. Les programmes de l'IA ne suffisent pas à rendre les machines intelligentes. Cela est une tâche impraticable, car les machines n'ont pas de capacités sémantiques, lesquelles sont le fruit de l'Intentionnalité des états mentaux.

(c) Pour avoir un esprit il faut avoir le même pouvoir causal que le cerveau:

Comme nous l'avons vu dans la partie précédente de ce chapitre, Searle affirme que le cerveau cause l'esprit; pour qu'une machine puisse penser, il faut qu'elle possède le même pouvoir causal que le cerveau. Pour Searle nous ne pouvons pas avoir de machines vraiment intelligentes, car nous n'avons pas les moyens de reproduire artificiellement la causalité du cerveau humain à produire certaines caractéristiques fondamentales de l'esprit telles que l'Intentionnalité et la conscience.

Le fait que le monde a une signification pour les êtres humais dépend de cette caractéristique de l'esprit qui est l'Intentionnalité. L'Intentionnalité, en tant que l'une des caractéristiques fondamentales de l'esprit, n'est pas une propriété formelle du cerveau.

La pensée humaine dont la base est physique est totalement différente des systèmes informatiques, car elle est un phénomène d'Intentionnalité et ne peut pas être définie de

façon purement syntaxique. Lorsque nous pensons ou que nous comprenons une langue, nous faisons plus que des opérations formelles. Nos pensées ont des contenus sémantiques qui nous permettent de donner de la signification aux symboles (les phrases du langage naturel).

Selon Searle, il y a un lien causal empirique entre le cerveau et l'esprit; l'Intentionnalité est le fruit de ce lien: certains processus cérébraux sont des conditions suffisantes de l'Intentionnalité. Les ordinateurs digitaux ne peuvent pas avoir des états intentionnels car ils n'ont pas les mêmes pouvoirs causaux qu'un cerveau biologique. Searle affirme que tout mécanisme capable de produire l'intentionnalité doit avoir un pouvoir causal au moins égal à celui du cerveau.

Étant donné que l'Intentionnalité est causée par le cerveau, elle ne peut jamais être le résultat de l'exécution d'un programme informatique. L'IA ne peut permettre, selon Searle de dupliquer l'intelligence humaine à l'aide de manipulations formelles (syntaxiques) de symboles. Les états computationnels de la machine ne peuvent, selon Searle, être comparés aux états intentionnels humains; l'exécution d'un programme informatique n'est pas une condition suffisante de l'Intentionnalité.

2.2.2- Les limites des programmes informatiques en tant que explications plausibles sur le fonctionnement de la pensée

b)Réfutation de la seconde thèse:

Pour réfuter la thèse de l'IA forte selon laquelle les programmes d'ordinateur de simulation sont en eux-mêmes des explications de l'esprit, Il faut que nous retracions toute la structure logique de l'argumentation de Searle dans *Minds Brains and Science*. Nous allons faire cela, maintenant, en guise de conclusion de cette partie:

A partir des propositions suivantes, mentionnées plus haut, (1) le cerveau cause l'esprit, (2) la syntaxe ne suffit pas à la sémantique, (3) les programmes d'ordinateurs ne sont définis que par leur structure formelle et syntaxique, (4) l'esprit possède un contenu mental, c'est-à-dire qu'il a un contenu sémantique; Searle arrive aux conclusions suivantes:

(A) Aucun programme informatique ne constitue, par soi même, une condition suffisante pour (avoir de l'Intentionnalité) donner un esprit à une machine digitale: Les

programmes informatiques de l'IA ne constituent pas des pensées et ne donnent pas en eux-mêmes la garantie que les ordinateur digitaux peuvent avoir des états mentaux.

(B) L'explication du fonctionnement d'un programme de l'IA ne constitue pas une l'explication de la manière dont le cerveau cause l'esprit (c'est à dire dont il produit de l'Intentionnalité)

(C) Tout mécanisme ou système capable de causer un esprit (produire de l'Intentionnalité) doit absolument avoir des pouvoirs causaux aussi importants que ceux d'un cerveau humain.

(D) L'exécution d'un programme d'ordinateur ne serait pas une condition suffisante pour l'obtention des systèmes capables d'avoir des états mentaux équivalents à ceux produits par un cerveau humain. Pour produire de la pensées artificielles, il faut concevoir des artefacts capables d'avoir un pouvoir causal comme celui du cerveau. De tels artefacts n'ont jamais été créés[338].

La conclusion A résulte des propositions 2, 3 et 4 et elle permet à Searle d'affirmer que le projet de l'IA forte est un mythe.

La conclusion B vient de la première proposition et de la conclusion A et permet à Searle d'affirmer que le cerveau, ne doit pas être expliqué uniquement par ses propriétés formelles comme s'il était un ordinateur. Selon Searle c'est le caractère biologique du cerveau qui constitue un élément important à l'explication des phénomènes mentaux et non le fait qu'ils ont certaines propriétés formelles semblables à un ordinateur digital.

La conclusion C est une conséquence triviale de la proposition 1 et permet à Searle d'admettre partiellement la thèse de la *réalisabilité* multiple des processus mentaux: d'autres systèmes, avec des caractéristiques chimiques et biochimiques différentes de celles du cerveau seraient capables d'avoir un esprit, à la condition, bien sûr qu'ils aient des pouvoirs causaux égales à ceux d'un cerveau humain.

La conclusion D est dérivée des conclusions A et C, elle est liée aux conceptions searlennes selon lesquelles les états mentaux sont des phénomènes biologiques qui ont une existence réelle de la même façon que n'importe quel autre phénomène biologique, par exemple la digestion ou la photosynthèse. De la même façon que ces derniers phénomènes biologiques les états mentaux sont causés par d'autres phénomènes biologiques, qui ont lieu à l'intérieur de notre cerveau.

---

338 Searle signale que les programmes d'ordinateur produits en IA ne sont pas cappable de dupliquer les pouvoirs qui permettent au cerveau de causer l'esprit. Il ne peuvent par conséquent produire l'intentionnalité qui est une condition nécessaire pour pouvoir penser. Les machines digitales , dans ce sens, ne peuvent pas penser et leurs programmes ne peuvent, non plus, constituer par eux-mêmes des explications de la pensée.

3- Les critiques de Searle à la science cognitive

Le thème de l'I.A. se situe, particulièrement aujourd'hui, à un carrefour théorique dont certaines voies semblent attirer plus d'attention que d'autres. Voilà l'impression de Jean-François Le Ny qui indique une de ces voies, la science cognitive:

> L'idée de science cognitive implique au contraire que l'on peut donner une autre sorte de description des événements qui ont lieu et des structures qui sont inscrites dans le cerveau. Cette description se situe à un niveau particulier, que l'on qualifie parfois de «niveau symbolique», ou de façon équivalente de «niveau des connaissances» (Newell, 1982). Elle met l'accent sur les fonctionnalités, les relations et les structures ainsi mises en oeuvre dans le cerveau, plutôt que sur la façon dont celles-ci se réalisent physiquement[339].

Pour Searle, au contraire de ce qui est exposé ci-haut, l'esprit n'est pas le résultat de fonctions abstraites, il est causé par des événements mentaux. Dans cette partie nous allons voir les critiques de Searle de la science cognitive. La raison pour laquelle nous voulons traiter de ce sujet est que la science cognitive et l'IA ont des liens étroits. Ces deux domaines de recherche se ressemblent par le fait que pour les deux la notion de programme informatique est essentielle; de plus les deux courants considèrent que l'ordinateur digital est l'image artificielle de l'esprit.

Nous commencerons cette partie en faisant une distinction entre l'IA et la science cognitive. Pour les chercheurs en IA, les programmes informatiques et les conceptions du matériel (hardware) reposent sur des notions distinctes, mais elles sont indissociées en tant que partie du système informatique. Pour les chercheurs de la science cognitive (ou des Sciences cognitives), par contre, la notion de programme est privilégiée; pour cette raison, pour eux, les notions de règles et de traitement d'information sont si importantes.

Au contraire des chercheurs de l'approche ascendante en IA, les cognitivistes (ceux qui se consacrent à la science cognitive) n'ont pas comme point de départ des modèles neuronaux; ils s'intéressent surtout aux capacités de manipulation symbolique des ordinateurs comme moyen de comprendre l'esprit.

Les cognitivistes s'intéressent aux machines intelligentes non pas par le fait qu'ils croient qu'elles peuvent penser, c'est le cas des chercheurs en IA, mais par le fait qu'en tant que machines à programmes et capables de manipuler toute sorte de symboles, les ordinateurs sont des outils efficaces pour la compréhension de l'esprit.

---

339 J. F. Le NY, , *Science cognitive et compréhension du langage*, Presses Universitaires de France, Paris, 1989 p.9.

La science cognitive privilégie un niveau intermédiaire d'analyse de la pensée (cognition); ce niveau se situerait entre le niveau physique des processus neuronaux et le niveau psychologique de descriptions des événements mentaux.

Ceux quei travaillent en science cognitive proposent que si un jour nous pouvions construire une machine dotée d'un programme équivalent à l'esprit, alors nous comprendrions mieux l'esprit, et ainsi nous donnerions à la psychologie une base scientifique plus rigoureuse.

Pour Searle, la psychologie cognitive et les sciences cognitives en général subiront les mêmes déceptions que le béhaviorisme: elles échoueront dans leur tentative de fournir une explication plausible du fonctionnement de l'esprit. Nous essayerons de montrer maintenant quels sont les arguments qui lui permettent de faire une telle prévision.

Les principaux points que Searle touche dans sa critique de la science cognitive dans *Minds, Brains and Science* sont à propos des notions de *règle* et de *traitement d'information*. Il s'oppose à l'idée cognitiviste selon laquelle l'ordinateur peut servir de modèle aux études sur la cognition humaine. Cette idée fut déjà mentionnée sur la croyance à des analogies entre le cerveau et les machines. Ces analogies ont deux points d'appui: les notions de traitement de l'information et de suivre des règles lesquelles s'opposent aux conceptions sémantiques de Searle.

Les cognitivistes de l'approche descendante emploient la notion d'«information» telle qu'elle est proposée dans la *Théorie de la communication*. Pour eux, le cerveau se comporte comme n'importe quelle machine; il est possible de décrire les caractéristiques du cerveau sans tenir compte de son caractère biologique. Cette idée prend ses origines dans les perspectives fonctionnalistes de la philosophie. Pour étudier les processus cognitifs il suffit de comprendre le fonctionnement formel du cerveau en tant que système similaire aux systèmes informatiques; ainsi nous n'avons pas besoin de chercher les aspects biologiques et psychologiques des états mentaux conscients.

Les notions de traitement d'information et de suivre des règles sont à la base des affirmations cognitivistes. Searle se demande si les ordinateurs suivent des règles et traitent de l'information de la même façon que les être humains. Selon lui, nous avons toujours essayé d'utiliser les modèles technologiques pour mieux comprendre les processus du cerveau, et le fonctionnement de l'esprit: on a déjà affirmé auparavant que le cerveau ressemblait à un standard téléphonique, à un télégraphe, à des systèmes hydrauliques et électromagnétiques. Aujourd'hui, les cognitivistes travaillent, souligne Searle, avec le modèle de l'ordinateur digital. Tous ces modèles d'après lui ne sont que des métaphores

parfaitement compréhensibles même si elles sont intégrées parfois au vocabulaire scientifique. Le problème pour lui est que les gens oublient souvent le sens métaphorique et parlent littéralement lorsqu'ils établissent des analogies entre le cerveau et l'ordinateur.

3.1- La critique de la notion des règles

Il est possible d'identifier, dans la tradition représentationnaliste occidentale, l'origine lointaine de l'idée cognitiviste selon laquelle il y a une analogie fonctionnelle entre les systèmes cognitifs (la pensée) et les ordinateurs.

La tradition représentationnaliste depuis Platon et Leibniz, basée sur la confiance dans le calcul comme garantie de l'obtention des connaissances nous amène à croire que les conduites intelligentes et même l'esprit peuvent être analysés en termes de fonctionnement par des règles. C'est-à-dire que les processus mentaux sont réalisés formellement et ont des causes théoriques.

Les thèses de caractère représentationnaliste peuvent être identifiées par l'axiome selon lequel: la syntaxe ou l'ensemble des règles formelles d'un système lui confèrent sa signification. Pour tout système, s'il donne un résultat porteur de signification, ce système contient en lui-même une théorie qui constitue l'explication de la capacité du système de fournir des résultats porteurs de signification. En ce qui concerne la cognition, l'interprétation de l'axiome représentationnaliste est la suivante: tout système capable d'agir de façon intelligente ou d'avoir une pensée est régie par des règles internes, lesquelles une fois suivies par lui sont en elles mêmes l'explication du fonctionnement de la pensée ou de l'intelligence du système.

Pour montrer les faiblesses de la science cognitive à fournir un modèle adéquat du fonctionnement du cerveau et de l'esprit, Searle mentionne tout d'abord les preuves psychologiques de la validité du cognitivisme, à savoir: 1) les preuves basées sur les expériences de temps de réaction et 2) les preuves basées sur les théories linguistiques de caractère syntaxique. Ces preuves ont deux caractéristiques distinctes:

> The first comes from reaction-time experiments, that is, experiments which show that different intellectual tasks take different amounts of time for people to perform. The idea here is that if the differences in the amount of time that people take are parallel to the differences in the time a computer would take, then that is at least evidence that the human system is working on the same principles as a computer. The second sort of evidence comes from linguistics, especially from the work of Chomsky and his colleagues on generative grammar. The idea here

> is that the formal rules of grammar which people follow when they speak a language are like the formal rules which a computer follows[340].

Cependant, ces deux preuves ont en commun deux idées de base, à savoir: 1) le fait que nous pouvons concevoir des machines qui suivent des règles (lorsqu'elles résolvent des problèmes) est analogue au fait que les êtres humains eux aussi suivent des règles lorsqu'ils agissent de façon intelligente. 2) Il est possible de prouver par des moyens empiriques (première preuve) ou par des moyens formels (seconde preuve) que le cerveau et l'ordinateur suivent des règles lorsqu'ils traitent de l'information.

Pour Searle, la première preuve, basées sur les expériences de temps-réaction n'a absolument pas d'importance. Il se concentre plutôt sur la preuve linguistique fournie par les travaux de N. Chomsky selon lesquels le langage peut être compris en termes purement syntaxiques: les règles grammaticales peuvent être comparées à des machines de Turing; elles permettent des changements formels à l'intérieur du langage fournissant des explications sur la façon dont le langage est engendré dans notre esprit[341].

Searle discute d'abord la notion de règle à suivre. Nous savons ainsi que les êtres humains suivent des règles, les ordinateurs digitaux eux aussi en suivent. L'auteur remarque qu'il faut se demander quelle est la différence entre les êtres humains et les machines quand les deux suivent des règles:

> In the case of human beings, whenever we follow a rule, we are being guided by the actual content or the meaning of the rule. In the case of human rule-following, meanings cause behavior. Now of course, they don't cause the behavior all by themselves, but they certainly play a causal role in the production of the behavior [342].

Les comportements humains sont régies quelquefois par des règles, mais pas toujours. Lorsque nous intériorisons les règles, nous ne les suivons plus. Les propriétés formelles qui nous permettent de décrire ou de déterminer un comportement dans une situation donnée ne sont pas suffisantes pour démontrer qu'une règle est suivie.

---

340 J.R. Searle (1984), *op. cit.* ,p.p .44-45.

341 Les théories linguistiques de Chomsky inspirent certains auteurs en IA et en science cognitive à affirmer que nous pouvons d'identifier certains processus intelligents à des systèmes formelles dotés des règles de déduction et d'inférence spécifiées exclusivement en termes syntaxiques. En philosophie nous avons des théories syntaxiques de l'esprit que dans le même chemin des chercheur de cognitivistes et de l'IA, se préoccupent de expliquer les processus formels que sont derrière le fonctionnement de l'esprit.

342 *Idem.* , p. 46.

Selon Searle il y a une distinction entre les règles qu'un agent suit et les hypothèses ou généralisations qui servent à décrire correctement son comportement. Selon Searle cela peut être constaté par une autre distinction importante entre suivre des règles et agir selon des règles (*in accordance with*). Lorsque les êtres humains suivent des règles, ce qui compte n'est pas la règle en soi (comme un ensemble de formalités sans signification à accomplir), au contraire suivre une règle est un acte guidé par la signification ou contenu que telle règle renferme.

Une règle est suivie lorsqu'elle a une signification pour quelqu'un. La signification joue un rôle causal pour la production des comportements intelligents.

Nous pouvons nous demander si les ordinateurs procèdent de la même façon. S'ils suivent réellement des règles, est-ce qu'en analysant les résultats des procédés algorithmiques nous pouvons affirmer que les programmes informatiques suivent des règles?

> (...) *in that sense computers don't follow rules at all. They only act in accord with certain formal procedures*. The program of the computer determines the various steps that the machinery will go through; it determines how one state will be transformed into a subsequent state[343].

Selon Searle il y a une différence importante entre la notion de suivre des règles appliquée aux êtres humains et lorsque nous l'appliquons pour décrire ce que fait un ordinateur. Le sens humain de suivre des règles (où la signification des règles joue un rôle causal dans le comportement.) ne s'applique que métaphoriquement aux ordinateurs. Nous ne pouvons pas dire qu'un ordinateur suit littéralement des règles. Lorsqu'un ordinateur fait quelque chose en fonction de certaines procédures pré-programées il ne suit aucune règle. Il accomplit certains processus formels, il fait comme si il suivait des règles ou mieux, il fonctionne *selon des règles*. Il exécute de procédures formels et dans ce cas, la sémantique ne joue aucun rôle,.

Nous ne devons pas prendre le sens métaphorique de suivre des règles comme étant littéral. Nous pouvons dire que les ordinateurs suivent des règles. Cela peut être utile aux sciences informatiques et en IA. Au sens littéral seuls les êtres humains suivent des règles (et selon Searle ils ne le font pas souvent). Nous devons remarquer que Searle ne dit pas que les règles ne jouent pas un rôle dans les processus mentaux et par conséquent dans

---

343 *Idem.* , p. 47.

notre comportement; ce qu'il veut dire est que les hypothèses ou généralisations qui permettent de décrire un comportement ne constituent pas une garantie que les règles sont souvent suivies par un agent ou qu'il suit toujours de règles.

Cette analyse de la notion métaphorique et littérale de règle permet à Searle d'expliquer pourquoi la preuve linguistique du cognitivisme ne fonctionne pas. Pour approfondir sa critique, il ajoute l'idée attachée à la notion métaphorique de règle, selon laquelle tout comportement pourvu de sens a une théorie interne qui le soutient.

La critique à la notion de règle telle qu'elle est utilisée en IA vaut aussi dans le contexte de l'analyse linguistique de Chomsky:

> So for example, Chomsky's search for a universal grammar is based on the assumption that if there are certain features common to all languages and if these features are constrained by common features of the human brain then there must be an entire complex set of rules of universal grammar in the brain[344].

Selon l'interprétation fonctionnaliste des recherches linguistiques de Chomsky, nous devons, pour comprendre la pensée, privilégier les caractéristiques formelles de l'esprit (les règles qui régissent son fonctionnement) indépendamment du cerveaux biologique qui les engendre. Les fonctionnalistes prennent l'hypothèse de la grammaire universelle exposée ci-haut comme si elle était l'explication définitive de l'esprit et du langage. Pour eux, il existe une sorte de grammaire universelle qui explique comment l'esprit fonctionne; il existe un ensemble de règles complexes qui sont à la base de nos pensées.

Selon Searle nous n'avons pas besoin de règles formelles pour expliquer comment l'esprit fonctionne; il suffit de comprendre comment les structures neurophysiologiques causent les processus mentaux.

Pour Searle, les explications neurophysiologiques peuvent très bien être comprises selon un modèle explicatif formel de la façon dont le cerveau engendre des grammaires possibles. Cependant, le cerveau lui-même n'a pas besoin d'une grammaire universelle fondée sur un niveau intermédiaire[345] de règles et de théories. Searle explique bien ce qu'il veut dire avec l'exemple suivant:

> It is a simple fact about human vision that we can't see infra-red or ultra-violet. Now is that because we have a universal rule of visual grammar that says: 'Don't see infra-red or ultra-violet.'? No, it is obviously because our visual apparatus simply is not sensitive to these two ends of the spectrum[346].

---

344 *Idem.* , p. 51.
345 Entre les descriptions en termes mentaux et une description en termes physiologiques)
346 *Idem.* , p. 51.

La preuve linguistique du cognitivisme est mise en doute par Searle: pour lui le cerveau humain n'est pas structuré sur un ensemble complet de règles complexes et universelles imprimées dans notre code génétique. Les règles ne servent pas à expliquer nos comportements; elles ne constituent pas non plus une base sûre à une théorie sur l'esprit:

> (...) If we tried to do a theoretical analysis of the human ability to stay in balance while walking, it might look as if there were some more or less complex mental processes going on, as if taking in cues of various kinds we solved a series of quadratic equations, unconsciously of course, and these enable us to walk without falling over. But we actually know that this sort of mental theory is not necessary to account for the achievement of walking without falling over. In fact, it is done in a very large part by fluids in the inner ear that simply do no calculating at all.(...) We have no good reason for supposing that in addition to the level of our mental states and the level of our neurophysiology there is some unconscious calculating going on[347].

Il croit que même si le langage est régi en partie par des règles syntaxiques, le cerveau et/ou l'esprit fonctionnent de façon complètement distincte des ordinateurs, ou d'une machine abstraite comme la machine de Turing. Le cerveau n'est pas un simple mécanisme physique de traitement de l'information ou de manipulation de symboles. Il est le siège de notre pensée laquelle ne peut pas être comprise complètement par le moyen de programmes d'ordinateur.

3 2- La critique à la notion de traitement de l'information

Searle doute de la notion de traitement d'information telle qu'elle est employée par les cognitivistes pour expliquer comment les êtres humains pensent. Comme la notion de «suivre des règles»; la notion de traitement de l'information a aussi un sens littéral et un sens métaphorique qui doivent être distingués.

La notion de traitement de l'information est une notion qui prend en considération seulement la forme (le contenant) et pas le contenu de ce qui est traité.

Selon Searle, nous ne pouvons pas confondre la façon par laquelle les machines «manipulent des informations» avec la façon humaine de le faire. Lorsque nous calculons, par exemple, nous procédons par un effort mental, une machine procède formellement sans

---

347 J.R. Searle (1984), *op. cit.* , pp. 51-52.

recourir à des processus mentaux; alors, nous ne pouvons pas dire, que les machines traitent les informations dans le même sens que les humains le font.

Pour cet auteur, traiter des informations dans le sens humain du terme n'est nullement manipuler des symboles. Quand nous traitons des informations nous avons des engagements et nous avons conscience du contenu de l'information traitée. Nous avons d'abord et avant tout des processus mentaux. Une machine ne présente pas ces mêmes caractéristiques, donc elle ne traite pas d'informations de la même façon que nous, elle simule tout simplement, dans certains cas les caractéristiques formelles des processus mentaux exigés pour que les êtres humains traitent de l'information:" (...) even if the steps that the calculator goes through are formally the same as the steps that I go through, it would not show that the machine does anything at all like what I do(...)"[348].

Searle distingue deux sens dans la notion de «traitement d'informations»:

1)le traitement humain (psychologique) de l'information qui est essentiellement humain et implique, des processus mentaux.

2)le traitement de l'information du type "comme si" qui ne dépend d'aucun état mental. Il s'agit selon Searle d'un processus simulé qui est entièrement physique.

Searle suggère que les cognitivistes confondent fréquemment le sens 1 et 2 parce qu'ils mettent en rapport les caractéristiques formelles de la pensée humaine avec les caractéristiques formelles des programmes informatiques.

> From the fact that I do information-processing when I think, and the fact that the computer does information-processing — even information-processing which may simulate the formal features of my thinking — it simply doesn't follow that there is anything psychologically relevant about the computer program. In order to show psychological relevance, there would have to be some independent argument that the 'as if' computational information-processing is psychologically relevant.The notion of information-Processing is being used to mask this confusion, because one expression is being used to cover two quite distinct phenomena [349].

Les processus de traitement de l'information et les règles computationnelles sont des mécanismes formels du type "comme si", lesquels sont psychologiquement neutres, c'est-à-dire ils peuvent permettre de simuler les caractéristiques formelles de la pensée, mais pas

---

348 *Idem.* , p.48.
349*Idem*, p.49.

l'aspect Intentionnel qui est à la base des états mentaux. Par conséquent, la notion de traitement de l'information dans le sens employé par la science cognitive n'est pas psychologiquement valable.

> However, there is a deeper and more subtle confusion involved in the notion of information-processing. Notice that in the 'as if' sense of information processing, any system whatever can be described as if it were doing information-processing, and indeed, we might even use it for gathering information. So, it isn't just a matter of using calculators and computers. Consider, for example, water running downhill. Now, we can describe the water as if it were doing information-processing. And we might even use it to get information. We might use it, for example, to get information about the line of least resistance in the contours of the hill. But it doesn't follow from that there is anything of psychological relevance about water running downhill. There's no psychology at all to the action of gravity on water[350].

Notre pensée peut fonctionner, bien sûr, sur une base formelle au niveau des micro-structures neuronales. Mais selon Searle, cela ne constitue pas complètement la base de notre vie mentale, de nos pensées.

> When we step in wet sand and make a footprint, neither our feet nor the sand does any computing. But if we were going to design a program that would calculate the topology of a footprint from information about differential pressures on the sand, it would be a fairly complex computational task. The fact that a computational simulation of a natural phenomenon involves complex information-processing does not show that the phenomenon itself involves such processing [351].

3.3- Le caractère anti-fonctionnaliste des critiques de Searle à l'IA

Le débat ouvert par Searle sur l'I.A et la science cognitive peut être compris dans le fond comme une critique du fonctionnalisme.

Searle montre que la science cognitive a comme base l'idée fonctionnaliste selon laquelle l'esprit humain ne dépend d'aucune structure biologique *spécifique*, c'est-à-dire celle d'un cerveau humain. Le cerveau humain pour le fonctionnaliste n'est qu'une sorte d'ordinateur à programmes parmi d'autres. Pour eux l'esprit (ou le fait d'avoir des états mentaux) ne dépend pas du matériel (*hardware*) constituant les systèmes mais du logiciel (*software*). Il est théoriquement possible que les machines puissent avoir des états mentaux

350 J.R. Searle (1984), *op. cit.* , pp.49-50.
351 *Idem.* , p.52.

semblables à ceux de l'être humain; pour cela, il suffit que les machines soient programmées adéquatement.

L'analogie entre l'esprit et les programmes d'ordinateur proposée par les théories fonctionnalistes est importante aussi bien pour l'IA que pour la science cognitive. Cette analogie est basée sur l'idée selon laquelle l'esprit résulte d'un traitement formel de symboles à la façon d'un ordinateur digital ou machine de Turing.

Le fonctionnaliste dit que le modèle computationnel sert à expliquer les événements mentaux et peut servir de base à la psychologie. Searle, par contre, ne croit pas qu'il soit possible de donner un nouveau statut à la psychologie, au moyen des théories de type fonctionnaliste.

Il ne croit pas qu'un niveau intermédiaire d'analyse des états mentaux, (entre la neurophysiologie du cerveau et l'intentionnalité de l'esprit) soit efficace pour l'explication des phénomènes mentaux.

Les thèses fonctionnalistes esquissées plus haut sont admises par plusieurs chercheurs en IA, tels qu'Herbert Simon, Alan Newell, Marvin Minsky et John McCarthy.

Lorsque Searle critique l'IA et la science cognitive, il veut également prouver que les thèses fonctionnalistes sont insuffisantes pour comprendre la pensée et la dupliquer. Pour Searle, nous l'avons vu, l'esprit ne doit pas être comparé à un programme d'ordinateur, les définitions d'ordinateur digital et de programme informatique ne rendent pas compte du fait que nos états mentaux internes ont un contenu.

Dire que la pensée résulte des fonctions abstraites indépendantes de la réalité physique du cerveau, comme l'affirment les fonctionnalistes équivaut à expliquer les états mentaux par des caractères uniquement formels et syntaxiques exactement comme nous le faisons pour caractériser les programmes informatiques.

Searle n'est pas d'accord avec les thèses physicalistes ni avec les thèses de caractère béhavioriste (Turing) et fonctionnaliste (Fodor et autres) sur l'esprit. Un programme d'ordinateur ou machine de Turing ne sert pas à expliquer le fonctionnement de l'esprit, lequel ne peut non plus être adéquatement étudié en termes de comportements observables. (*input- output*), comme le montre le jeu d'imitation de Turing.

> That is, even if my thoughts occur to me in strings of symbols, there must be more to the thought than the abstract strings, because strings by themselves can't have any meaning. If my thoughts are to be *about* anything, then the strings must have a *meaning* which made the thoughts about those things. In a word, the mind has more than a syntax, it has a semantics. The reason that no computer program can ever be a mind is simply that a computer program is only syntactical, and

> minds are more than syntactical. Minds are semantical, in the sense that they have more than a formal structure, they have a content[352].

Searle affirme que les états mentaux sont aussi réels que les autres phénomènes biologiques que nous connaissons. Cette sorte de Naturalisme biologique est basée sur une position physicaliste naïve se rapprochent du physicalisme structurel auquel Fodor s'oppose et qui est décrit ainsi par ce dernier:

> La difficulté que rencontre le physicalisme structural, c'est que les possibilités psychiques d'un système dépendent non de son ***hardware*** (sa composition matérielle) mais de son ***software*** (son programme). Pourquoi le philosophe nierait-il la possibilité que des Martiens, dont la matière organique serait organisée autour du silicium, éprouvent de la souffrance, si le silicium est convenablement organisé? Et pourquoi le philosophe rejeterait-il la possibilité que des machines aient des croyances, si ces machines sont correctement programmées? S'il est logiquement possible que Martiens et machines possèdent des propriétés mentales, alors propriétés mentales et processus neurophysiologiques ne peuvent pas être identiques, quel que puisse être leur parallélisme[353].

Le projet épistémologique de comprendre l'esprit comme un mécanisme formel est soutenu déjà par les rationalistes et les l'empiristes. Le fonctionnalisme ou théorie computationnelle de l'esprit hérite de la tradition philosophique rationaliste l'idée selon laquelle pour tout comportement pourvu de sens il y a une théorie interne qui le soutient. Le fonctionnalisme s'insère naturellement dans le cadre du représentationalime que nous avons décrit où la pensée n'est rien de plus qu'un mécanisme de calcul. Searle donne à comprendre dans le passage qui nous citons ci-après que les thèses de l'IA forte sur l'esprit sont des thèses nettement fonctionnalistes:

> This view has the consequence that there is nothing essentially biological about the human mind. The brain just happens to be one of an indefinitely large number odd different kinds of hardware computers that could sustain the programs which make up human intelligence. On this view, any physical system whatever that had the right program with the right inputs and outputs would have a mind in exactly the same sense that you and I have minds. So, for example, if you made a computer out of old beer cans powered by windmills; if it had the right program it would have to have a mind. And the point is not that for all we know it might have thoughts and feelings, but rather that it must have thoughts and feelings, because that is all there is to having thoughts and feelings: implementing the right program[354].

---

352 *Idem*, p. 31.
353 J. Fodor (1981), *op. cit.* , p. 82.
354 J.R. Searle (1984), *op. cit.* , p.28-29.

La thèse fonctionnaliste de la "réalisabilité" multiple des états mentaux» est fondamentale pour les sciences cognitives et pour l'I.A. puisqu'elle permet d'affirmer que les ordinateurs peuvent, dés que convenablement programmés, avoir des états mentaux semblables à ceux des êtres humains. J. Searle n'est pas du tout d'accord avec les thèses sur la réalisabilité ni avec l'idée selon laquelle nous pouvons tester des théories psychologiques à partir des programmes informatiques.

Autrement dit, pour Searle, le programme informatique n'est pas un esprit; aucune théorie fonctionnaliste ne peut expliquer correctement l'esprit ni servir de bases pour donner un esprit à un ordinateur.

> Now, that is a very powerful conclusion, because it means that the project of trying to create minds solely by designing programs is doomed from the start. And it is important to re-emphatise that this has nothing to do with any particular state of technology or any particular state of the complexity of the program. This is a purely formal, or logical, result from a set of axioms which are agreed to by all (or nearly all) of the disputants concerned. That is, even most of the hard-core enthusiasts for artificial intelligence agree that in fact, as a matter of biology, brain processes cause mental states, and they agree that programs are defined purely formally[355].

Conclusion

Ainsi, nous concluons en disant que dans ses critiques à l'IA et à la science cognitive, Searle veut laisser clairement entendre que nous ne pouvons pas réduire l'esprit à des systèmes purement formels basés sur des règles et sur la notion de traitement d'information. L'esprit ne doit pas être compris, comme le veulent les fonctionnalistes, comme fonctionnellement équivalent aux états et transitions d'états d'un système formel, soit-il un ordinateur digital ou une machine de Turing.

Le caractère anti- fonctionnaliste des critiques de Searle à l'IA et à la science cognitive peut être résumé ainsi: la structure configurationnelle des états et transitions d'états du cerveau ou de la machine ne suffisent pas pour expliquer les rapports entre le cerveau et l'esprit, comment l'esprit fonctionne, et comment le cerveau humain a le pouvoir de causer l'esprit.Toute théorie sur l'esprit doit prend en considération non seulement le fait que l'esprit est causé par le cerveau, mais que l'esprit a un pouvoir causal sur le cerveau et que

---

355 J.R. Searle (1984), *op. cit.* , 39-40.

l'Intentionnalité des états mentaux permet au corps (le cerveau matériel) d'être en contact avec le monde, ce qui ne peut être executé par aucune procédure formelle.

Les critiques de Searle à l'IA et à la science cognitive ont comme cible les théories fonctionnalistes que leurs sont sous-jacents. La position de Searle concernant l'IA n'est pas l'anti-représentationnaliste, comme c'est le cas de Dreyfus, mais elle estanti fonctionnaliste. Ces deux auteurs ont été présenté ici pour montrer l'intérêt philosophique de l'IA. Il veulent montrer que l'IA est un *mythe* mais pour faire cela ils s'appuient sur son *logos*.

CONCLUSION GÉNÉRALE

Les théories scientifiques, comme l'héliocentrisme, la relativité etc. et les inventions, comme l'avion, la télévision, l'ordinateur etc. sont en général crées à partir des acquis d'un développement culturel antérieur qui n'exclut pas les erreurs, les mythes, les fausses croyances. L'IA en tant que rassemblement d'inventions, d'idées et de théories est un domaine de recherche marqué par la survivance d'une idée ancienne dans notre culture, parfois associée au monde de l'imaginaire (le *mythe* de l'IA), d'autres fois associée à un *logos*.

Kepler ayant trouvé dix-neuf hypothèses sur la trajectoire de Mars et calculé toutes ces trajectoires hypothétiques a réussi à trouver la vraie trajectoire elliptique. Nous ne sommes pas en mesure de savoir combien d'hypothèses sur le fonctionnement de l'esprit ou sur le fonctionnement des machines et des programmes en IA il sera nécessaire de concevoir pour prouver ou nier que l'intelligence artificielle est faisable.

Est-ce que nous pouvons avec nos connaissances et avec le développement des capacités formelles-symboliques des ordinateurs faire en sorte qu'une machine puisse penser? Cette question formulée philosophiquement serait posée ainsi: "quelles sont les limites de la pensée à produire une explication sur elle-même permettant de la reproduire au moyen de représentations formelles?"

Nous avons montré que l'IA, tout comme la philosophie grecque ancienne, se dégage peu à peu de son caractère mythologique pour se trouver des principes de rationalité à partir de l'intersection de connaissances diverses. Le *logos* de l'IA est explicité par un assemblage de théories et techniques dans le domaine de l'informatique, de la neurologie, de la linguistique, de la psychologie cognitive,de la logique et de la philosophie. Ce *logos* est capable de la soutenir en tant que discours sur les possibilités concrètes de créations d'êtres artificiels (ou si on veut de programmes) intelligents.

Lorsqu'une critique à l'IA touche à un aspect quelconque de ces éléments théoriques qui la constituent, situés quelque part à l'intersection de l'informatique, des études sur le cerveau et des sciences humaines, c'est le *logos* de l'IA qui est en évidence.

De la même façon que les milésiens, Thalès, Anaximandre et Anaximène élaborent toute sorte d'explications construisant une nouvelle image de l'univers qui met de côté tout recours au mythe, ceux qui travaillent en IA manient le *logos* (discours) sur les machines intelligentes afin d'éliminer toute rhétorique mythologique qui contribue à une mauvaise

compréhension de leurs propos. Les chercheurs en IA inventent et réinventent l'IA comme les pré-socratiques on inventé et réinventé pas à pas la philosophie.

Le fait que l'IA se fonde sur un logos n'exclut pas le fait qu'elle possède un caractère mythique lié à l'univers de l'imaginaire humain. Le passage du *mythe* au *logos* de l'IA serait fait à partir de certaines conceptions philosophiques rationalistes de la concrétisation et de la réussite de plusieurs de ses projets théoriques et technologiques considérés utopiques.

Nous entendons que le *mythe* de l'IA est lié à son *logos*, mais pour définir et réaliser un projet dans ce domaine on a dû abandonner le mythe. Si, lorsqu'il a rêvé pour la première fois de voler comme les oiseaux, l'homme était resté sur des considerations mythologiques, mettant en rapport ses limitations à celles d'Icare et d'autres personnages mythiques, il n'aurait certainement pas conçu (par le rassemblement de connaissances et expériences concrètes) l'art et les outils pour voler, mais se serait limité a créer d'autres mythes capables d'entretenir son rêve.

Nous allons refaire de façon abrégée et conclusive le chemin que nous avons parcouru dans les quatre chapitres en explicitant pour chacun d'eux nos conclusions générales.

Dans le chapitre I nous avons caractérisé comme mythe de l'IA toutes les idées sur des êtres artificiels qui ne font pas partie de la culture rationaliste. Le *mythe* tel que nous l'avons exposé révèle, si on veut, des éléments pré-historiques, ancrés dans la culture occidentale. Ces éléments précèdent l'IA proprement dite, faisant en sorte que l'idée d'une "Intelligence artificielle" ne paraisse pas si étrange, en dépit des limitations théoriques et techniques trouvées dans ce domaine. Le *mythe* de l'I.A remonte à des époques lointaines, les récits, les poésies, les légendes et plusieurs documents de valeur littéraire évoquent le dessein de l'homme de créer de répliques de son corps et de son esprit à partir de éléments de la nature.

La phase mythique est interrompue seulement aux XVIIe et XVIIIe siècles lorsque Descartes, ainsi que Hobbes, Hume, Leibniz et Kant, ont rendu possible le modèle épistémologique de l'homme-machine, et l'idée de la pensée en tant que mécanisme de calcul permettant de mieux comprendre l'homme comme un système mécanique. Nous avons caractérisé comme *logos* tout un ensemble de connaissances scientifiques, philosophiques et techniques qui constituent en elles-mêmes les principes de rationalité présupposés par la notion d"Intelligence Artificielle". Le *logos* représente, un abandon des modèles mythiques; il se manifeste par le triomphe de la conception philosophique de la raison, conçue comme une sorte de calcul, et par le développement de la programmation et de la technologie informatique. Ces deux moments, l'un sur le plan des idées et l'autre sur

le plan de la technique, constituent respectivement les bases empiriques et rationnelles pour la constitution de l'IA.

Le *logos* de l'IA se construit à partir de deux approches, l'approche ascendante, liée aux théories et hypothèses de caractère biologique, généralement associées à la cybernétique, et l'approche descendante, dont les théories et hypothèses privilégient l'aspect formel et la construction de modèles logiques capables de rendre compte des procédures logiques qui sont sous-jacentes aux raisonnements, au langage naturel et sont en rapport avec la résolution de problèmes. Cette dernière approche est liée à la tradition représentationnaliste (que nous avons discutée dans le chapitre II) selon laquelle il est possible d'expliquer le fonctionnement de la pensée au moyen de représentations formelles.

Nous avons pu conclure à partir des discussions de notre premier chapitre que l'IA est liée à des mythes (légendes, récits, croyances sans justification etc.), mais qu'elle se développe à partir d'un logos, que lui donne sa rationalité, sa cohérence et sa justification en tant que thème du domaine scientifique. La notion de logos permet de faire ressortir la signification historique et philosophique de la recherche sur les machines intelligentes.

Nous avons montré dans le chapitre II que l'IA a comme base plusieurs conceptions, lesquelles sont dérivées des courants philosophiques divers, caractérisés dans ce mémoire, de façon très générale, par le nom de tradition représentationnaliste (ou tradition computationnelle/ calculationnelle selon H. Dreyfus). Lorsque nous avons dit que l'IA est en rapport avec cette tradition, nous voulons souligner qu'elle est liée à des idées sur la représentation et sur l'esprit issues elles-mêmes des philosophie rationaliste et empiriste.

La conception selon laquelle la pensée a pour base une théorie ou un ensemble de règles existe depuis Platon. Une telle conception a inspiré les thèses représentationnalistes en IA. Tout cet héritage philosophique est sous-jacent aux travaux de la branche descendante de l'IA, aussi appelée l'approche analytique et formelle. Toutefois, cette dernière approche et l'approche concurrente, dite connexionniste, font toutes les deux, appel à l'idée de représentation comme moyen de créer des programmes intelligents.

Selon certaines conceptions de IA, l'esprit fonctionne selon un mécanisme de calcul complexe. Les machines ont les mêmes propriétés que la pensée rationnelle d'exécuter des opérations sur des symboles. Une autre idée courante, c'est qu'il est possible, au moyen de représentations formelles, de rendre compte de nos raisonnements et de notre pensée. Cette idée est dérivée de la conception selon laquelle l'intelligence humaine et l'ordinateur digital adéquatement programmé partagent la propriété d'être des systèmes capables d'avoir de comportements intelligents.

Toutes les conceptions mentionnées en haut sont héritières de la tradition philosophique représentationnaliste, laquelle s'est inspirée en partie du succès des sciences physiques dans l'explication de la nature au moyen du calcul.

L'IA et la philosophie partagent des questions telles que: 1) L'esprit peut-il être considéré vraiment comme un système de calcul? 2) L'esprit est-il un mécanisme qui manipule de symboles? 3) La pensée peut-elle être décomposée en des éléments plus simples nous permettant de comprendre nos raisonnements et nos comportements intelligents?Descartes, Hobbes, Leibniz et al. essayent de discuter et de répondre de façons diverses à ces questions et concluent que lorsque nous pensons nous employons certaines règles, nous procédons par des opérations de calcul et dans ces termes nous utilisons des représentations.

Les travaux en IA montrent, en général, que les règles et les représentations jouent un rôle fondamental dans la production et dans la compréhension des comportements intelligents. La notion de règle est liée à la notion de pensée en tant que mécanisme de calcul. En IA, comme dans la conception moderne de l'esprit, les règles relèvent d'une façon organisée de voir le monde et de résoudre des problèmes. Les conduites intelligentes, de ce point de vue, doivent, par hypothèse, être régies par des règles. En IA l'idée est que si nous découvrons ces régles nous pouvons expliquer et décrire formellement certains comportements intelligents et on est capable de les programmer sur un ordinateur digital.

L'IA s'inspire non seulement des idées philosophiques de Descartes, Hobbes, Leibniz et Hume et du succès de la physique moderne, mais aussi du développement des mathématiques et de la logique contemporaine, lesquelles constituent aussi des éléments importants pour l'IA.

Nous avons vu que depuis Platon, les comportements intelligents et l'esprit sont analysés en termes de règles et qu'ils sont exprimés en termes d'une théorie quelconque qui constitue aussi leur explication[356]. Les philosophes fonctionnalistes vont dans la même direction que Platon. Suivant la tradition représentationnaliste, ils croient être possible d'établir une théorie formelle qui explique le fonctionnement de l'esprit en termes de machines de Turing.

Le fonctionnalisme est à la base d'un ensemble d'idées, appelé dans plusieurs travaux récents de "néo-mécanisme", lequel caractérise l'IA et les recherches cognitivistes.

---

356 Cf. H.L. , Dreyfus (1979) op. cit. pp. 67 et 176-177.

Pour le fonctionnalisme nous pouvons expliquer les propriétés mentales, au moyen des formalisations; ces propriétés sont analysées en tant que fonctions abstraites et considérées comme complètement indépendantes de la base matérielle du système. Il existe un niveau théorique intermédiaire pour décrire le fonctionnement de l'esprit, et ce niveau de description intermédiaire, entre les descriptions des phénomènes mentaux et physiques, a un caractère formel.

Il n'est pas question selon Fodor, par exemple, de décrire les états mentaux comme identiques au cerveau; pour lui, ils sont des propriétés formelles des systèmes physiques ayant un certain type d'organisation. Les états mentaux sont des fonctions abstraites indépendantes de la réalité physique du système qui les réalise.

Le fonctionnalisme définit un état mental par ses relations causales avec d'autres états mentaux. Il est possible d'expliquer les phénomènes mentaux en termes a) d'une individualisation des états mentaux qui composent un événement mental donné; b) d'une analyse des relations que les états mentaux entretiennent les uns avec les autres et c) de la spécification fonctionnelle donnée par le rôle causal de ces états.

Le fonctionnaliste affirme que nous aurons compris le fonctionnement de l'esprit lorsque nous aurons conçu un programme qui soit l'équivalent fonctionnel de celui-ci. Si nous pouvons, par les moyens formels dont nous disposons, concevoir des programmes ou des machines capables de réaliser les mêmes processus que ceux produits par l'esprit humain, alors il est théoriquement possible de concevoir de systèmes capables d'avoir des états mentaux semblables à ceux des êtres humains.

Les deux, le fonctionnalisme et la tradition représentationnaliste sont en rapport avec les thèses les plus importantes en IA.

Le courant philosophique fonctionnaliste est une continuité du projet métaphysique représentationnaliste d'analyser l'esprit à partir de principes universels de calcul. Elle est aussi la théorie sur l'esprit la plus proche des recherches en IA.

Nous concluons à partir des discussions de notre deuxième chapitre que les thèses en IA sont basées sur des conceptions représentationnalistes (comme moyen d'explication de la pensée et des comportements intelligents), et d'une conception mécaniste de la pensée (le fonctionnalisme). Ces deux éléments constituent les fondements philosophiques du *logos* de l'IA et permettent d'affirmer que l'IA et la philosophie sont reliées entre elles.

Dans les chapitre III et IV nous avons présenté deux approches critiques distinctes sur l'IA défendues par Dreyfus et Searle. Nous avons travaillé sur leurs points de vue philosophiques sur l'IA à partir de la discussion de leurs travaux les plus importants sur ce

sujet. Notre but a été de montrer que l'IA est un domaine de recherche susceptible de critique philosophique et que cela permet en même temps de réitérer notre explication de l'IA en tant que logos[357] et de confirmer les rapports entre l'IA et la philosophie.

Dreyfus et Searle ont posé, tous les deux, des objections de caractère philosophique à l'IA pour signaler les limitations théoriques dans ce domaine. Ils ont montré, par des arguments quelquefois très proches, qu'étant donné les moyens théoriques employés par ceux qui travaillent sur l'IA nous ne pouvons pas dire qu'un ordinateur se comporte de façon intelligente (Dreyfus) ou qu'il comprend le langage naturel (Searle).

Dreyfus analyse les résultats et les présuppositions de l'intelligence artificielle en utilisant une approche épistémologique d'inspiration phénoménologique, tandis que Searle fait une analyse des limitations sémantiques des systèmes dits intelligents à partir d'une approche analytique.

Les critiques de Dreyfus et de Searle ont trois points en commun: (1) l'analyse de la question des règles par rapport au comportement intelligent, (2) la mise en relief du corps (ou du cerveau) dans l'analyse de l'esprit humain et des comportements intelligents et (3) l'analyse de la notion de traitement d'information.

Pour Dreyfus aussi bien que pour Searle la description d'un comportement intelligent au moyen de règles ne signifie pas que les règles interviennent dans l'exécution d'un tel comportement. Par exemple, nous utilisons le langage sans faire appel à des règles. Lorsque ces règles sont intériorisées, on n'a plus besoin d'elles. Les règles servent à décrire notre compétence linguistique mais elles ne suffisent pas à orienter ni à expliquer notre comportement (performance) linguistique.

Les deux philosophes s'attaquent à la notion de traitement d'information employée en IA et montrent que lorsque nous agissons de façon intelligente nous ne traitons pas de l'information dans le sens strictement formel que cette expression possède. Pour Searle cette notion de la Théorie de la Communication a été sémantisée indûment, tandis que pour Dreyfus, elle ne correspond pas à ce que nous faisons lorsque nous agissons de façon intelligente.

Pour Searle et pour Dreyfus le corps joue un rôle très important dans nos comportements; pour le premier le corps ou le cerveau doit être en interaction avec le monde pour qu'on puisse comprendre le langage et agir de façon intelligente. Pour le

---

357 Les chapitre III et IV sont en rapport avec le première chapitre dans le sens où ils prouvent que c'est en tant que logos que l'IA fait objet de critique philosophique et de discussion scientifique.

second, c'est le rapport du corps avec le monde qui oriente la plupart de nos conduites intelligentes et nous permet de résoudre des problèmes.

Le corps joue un rôle important pour l'intelligence humaine, car il permet une interaction avec le monde et les autres êtres. Il nous permet d'apprendre, de percevoir et de comprendre une situation quelquonque. Le corps nous permet, entre autres choses, de fixer notre attention sur des objets qui sont dans le champ de notre attention, tout en laissant de côté ce qui est accessoire ou ce qui ne nous intéresse pas. Nous analysons un problème ou une situation sélectivement et faisons le tri de ce qui est pertinent ou non à partir des expériences vécues et de nos besoins et pour le faire, nous utilisons notre corps. La compréhension du langage et de nos comportements intelligents demandent des habiletés et des facultés qui ne sont pas accessibles aux machines car elles n'ont pas un corps comme le nôtre.

Une chose qu'a retenu notre attention dans les critiques de Dreyfus et de Searle a été que lorsque ces deux auteurs s'efforcent de mettre en évidence les mythes et les limites de l'IA, ils se rapportent aux limites philosophiques qui sont derrière l'IA: les limites du représentationnalisme et du fonctionnalisme.

Dans le chapitre III nous avons présenté quelques critiques de Dreyfus à l'IA. Nous avons vu comment Dreyfus explique que ce sont des présuppositions basées sur le représentationnalisme et sur le mécanisme qui servent à expliquer le fait que les recherches en IA continuent à être faites, en dépit de leurs échecs continus. Ces présuppositions (assomptions) d'ordre biologique, psychologique, épistémologique et ontologique, lui permettent de faire une réflexion philosophique et une critique épistémologique sur les idées fondamentales qui constituent la base de l'IA et de la recherche en Simulation Cognitive.

La présupposition biologique est identifiée, par l'affirmation que le cerveau fonctionne selon le modèle d'un ordinateur digital, la présupposition psychologique est fondée sur l'hypothèse que l'esprit fonctionne en termes binaires comme un ordinateur digital et selon des règles préétablies et qu'il est ainsi possible de programmer de façon convenable l'ordinateur pour qu'il puisse produire les mêmes sorties que l'esprit; la présupposition épistémologique consiste à présumer que tout comportement intelligent peut être formalisé et reproduit sur un ordinateur digital; la présupposition ontologique est caractérisée par la croyance que les traits essentiels des comportements intelligents peuvent en principe être analysés et décomposés en des éléments discrets, lesquels sont susceptibles d'être traités sur un ordinateur digital.

Toutes les présuppositions discutées par Dreyfus sont fondées sur l'idée que l'esprit est un mécanisme de calcul et que les processus cognitifs demandent un traitement de l'information et fonctionnent selon les mêmes principes formels qu'une machine abstraite ou qu'un ordinateur digital.

La plupart des présupposés mentionnés ont en commun l'idée qu'il est possible de formaliser les processus cognitifs au moyen de représentations formelles. Ils sont en rapport avec l'idée existente depuis Platon, selon laquelle il est possible d'analyser la conduite humaine au moyen des règles.

Selon Dreyfus, si on attribue à l'intelligence artificielle un statut de science on tombe immédiatement dans une pétition de principe, car, l'IA est fondée sur des présuppositions qui ne suffisent pas à la rendre cohérente, mais qui sont admises comme des axiomes. Les critiques philosophiques menées par Dreyfus, dans *What Computers Can't Do* ont reçu déjà beaucoup d'objections, dans plusieurs articles, mais aucun livre n'a été écrit jusqu'à maintenant pour critiquer ponctuellement ces conclusions et tous ces arguments. Un des mérites philosophiques de Dreyfus en rapport avec son travail sur l'IA est qu'il a été un des premiers philosophes, sinon le premier, qui a eu le courage de travailler sur l'IA en abordant un nombre vaste de domaines, avec une grande proximité du travail expérimental et théorique dans ce domaine.

Les conclusions que soulève Dreyfus, dans *What Computers Can't Do*, sur le besoin d'une plus grande collaboration entre l'homme et les machine au lieu d'essayer d'imiter les propriétés cognitives humaines sont en rapport avec certaines idées défendues précédemment dans les recherches sur les *systèmes experts*.

Un autre aspect important des critiques de Dreyfus est celui qui met en question le rôle des règles dans l'explication du comportement intelligent Pour lui les règles sont importantes pour l'explication scientifique des phénomènes naturels, mais l'intelligence, et la pensée humaine ne peuvent pas être expliquées complètement en termes de règles. L'ordinateur, à la différence des êtres humains, est obligé de suivre des règles strictes, mais les êtres humains lorsqu'ils agissent de façon intelligente ne suivent pas de règles. L'intelligence et la pensée humaine échappent complètement aux règles formelles et aux procédures de programmation informatique.

Les ordinateurs digitaux pour des motifs techniques et théoriques sont programmés avec des règles et manipulent des *faits atomiques* isolés en termes d'éléments discrets. Cela explique pourquoi les chercheurs en Intelligence Artificielle sont obligés de décomposer le

monde et les comportements en termes de concepts atomistiques élémentaires, mais ces stratégies formelles ne permettent pas de reproduire l'intelligence humaine.

Selon Dreyfus cette confiance inébranlable dans le calcul, c'est-à-dire dans les représentations formelles comme moyen de comprendre l'esprit en créant des programmes intelligents dérivent de quelques présuppositions mentionnées comme étant sous-jacentes aux recherches en IA. Ces présuppositions sont le fruit de la tradition représentationnaliste (computationnelle/calculationnelle).

Pour Dreyfus les comportements intelligents et l'esprit humain résistent à être réduits à des représentations formelles. Pour lui les études cognitives basées sur des conceptions mécanistes comme le fonctionnalisme ne servent pas d'espoir à la recherche en IA. D'ailleurs, sur ce sujet il démontre le même scepticisme que Searle (dans *Minds Brains and Science*) il affirme que théoriquement on peut arriver à créer des êtres artificiels avec une intelligence comparable à la nôtre, à la condition que de tels êtres puissent être constitués d'un corps semblable au corps humain. Selon Dreyfus, les propriétés formelles de l'esprit et la façon dont le cerveau est organisée ne permettent pas d'expliquer les comportements intelligents. Toute explication de l'intelligence doit d'abord compter sur une compréhension du monde et des besoins humains en rapport avec cette intelligence.

Les situations humaines ainsi que nos comportements ne sont jamais dénués d'ambiguïté. Ce qui est simple et proche des êtres humains relève de la pragmatique des rapports humains. Il n'y a pas de règle ni de formalisme capable de rendre compte de cette pragmatique. Selon Dreyfus, pour élaborer des programmes capables de traiter des contextes pragmatiques, il faut que les machines aient une certaine capacité d'apprendre, mais pour apprendre il faut qu'elles soient capables de vivre en situation dans un monde humain et se comporter comme des êtres humains. Pour Dreyfus les conduites humaines, à la différence de celles des machines, ne dépendent pas de la formalisation, ou d'un programme quelconque.

L'attitude de Dreyfus tend plutôt vers une assimilation critique des études en intelligence artificielle, en prenant une posture philosophique qui accepte l'intelligence artificielle comme un projet de recherche falsifiable. La réussite de l'intelligence artificielle serait selon lui la réussite du projet de la métaphysique représentationnaliste traditionnelle d'analyser et expliquer complètement le monde et l'esprit humain par des moyens formels.

La critique des présuppositions biologique, psychologique, épistémologique et ontologique sous-jacentes aux travaux en IA ont un caractère épistémologique et servent à dénoncer également l'attitude très peu critique manifestée par les chercheurs en Intelligence

Artificielle et en science cognitive en sur-valorisant leurs travaux et en ne reconnaissant pas leurs limites théoriques. Dreyfus entend qu'il faut démystifier les entreprises technologiques dans ces champs et reconnaître les limites technologiques et théoriques des projets en IA.

Selon Dreyfus, nous ne pouvons pas décomposer les comportements intelligents ni, l'esprit en termes d'éléments formels indépendants capables d'être traités comme des données discrètes sur un ordinateur digital. L'intelligence, ou les comportements intelligents pour lui sont en rapport avec des situations humaines et liés à un monde construit *par* et *pour* les êtres humains.

Le monde humain *résiste*, à être compris complètement en termes des représentations formelles. Dreyfus est dans cette perspective un philosophe anti-représentationnaliste il se considère lui-même, dans ce sens, un anti-formaliste ou anti-mécaniste[358].

Cependant e la critique anti- représentationnaliste de Dreyfus présente une difficulté à signaler: étant donné que (a)l'IA est fondée sur des conceptions représentationnalistes et (b) qu'elle dépend aussi du choix d'une théorie de l'esprit qui soit en même temps compatible avec ce représentationnalisme et avec les présuppositions qui lui sont sous-jacentes, nous somme amenés à conclure que l'anti-représentationnalisme de Dreyfus ne laisse aucune voie ouverte pour un développement possible de l'IA, car l'IA est liée à des techniques de programmation qui font obligatoirement appel à des représentations. Nous ne voyons pas comment l'IA peut être opérationnelle sans compter sur des modèles représentationnalistes. Même si pour quelques chercheurs de l'IA il est possible d'étudier la cognition sans faire appel à des représentations[359], les représentations sont fondamentales dans ce domaine.

Nous avons conclus à partir du chapitre III de notre travail qu'étant clairement anti-représentationnaliste et mettant en jeu la tradition philosophique représentationnaliste comme base à la compréhension des comportements intelligents en IA, Dreyfus montre, en même temps, les rapports de l'IA avec la philosophie et l'intérêt philosophique qu'un tel thème peut avoir.

Dans le chapitre IV nous avons remarqué que pour Searle l'affirmation que les machines peuvent penser ne pose pas de problèmes, dès qu'on est pleinement conscient qu'on parle *métaphoriquement*. Dans un sens *littéral* seulement des êtres possédant un cerveau avec le pouvoir causal égal à celui du cerveau humain peuvent penser.

---

358 A.Y., Walworth, "The Prospects for Artificial Intelligence, thèse de Doctotorat, University of California at Berkeley, 1989, voir chapitre IV, " Dreyfus's Anti-representationalism" 56-104 et Dreyfus (1979), p. 56.

359 T.Winograd, et F. Flores,*Understanding Computers and Cognition: a New Foundation for Design*, Ablex Publishing Corporation, Norwood, N.J, 1986, 207p.F. Varela, *op.cit.*.

Pour Searle, il ne suffit pas de créer un système analogue à certains processus formels de l'esprit, pour pouvoir attribuer de l'Intentionnalité aux machines. La conception d'un système capable d'être fonctionnelement équivalent au cerveau humain, n'assure pas qu'un tel système soit capable d'avoir des états mentaux.

Les critiques de Searle à l'IA et à la science cognitive sont fondées sur ses thèses sur l'esprit et sur le langage. Pour lui, un être ayant un esprit peut manipuler des significations (des entités sémantiques). Searle entend que le cerveau cause l'esprit, par conséquent seul un être possédant un cerveau biologique avec les mêmes propriétés intentionnelles que le cerveau humain peut manipuler des entités sémantiques. La syntaxe qui régle le fonctionnement des ordinateurs digitaux ne suffit pas à rendre compte de la sémantique. Pour Searle, l'esprit ne peut pas être expliqué ni être dupliqué par le moyen des programmes d'ordinateurs.

Selon Searle, quand nous pensons, nous faisons plus qu'exécuter des opérations formelles. L'esprit, pour lui, n'est pas un mécanisme syntaxique; au contraire, il se caractérise par sa capacité de donner signification et d'articuler Intentionnellement syntaxe et sémantique. Pour cette raison, selon Searle, les ordinateurs ne peuvent pas penser. L'activité de l'esprit exige des contenus représentationels; la pensée ne résulte pas de manipulations formelles dénués de signification.

Le fait que les machines ne sont pas capables d'être programmées pour avoir de l'Intentionnalité est en rapport avec le fait qu'elles sont des systèmes formels syntaxiques qui ne peuvent jamais rendre compte du contenu sémantique des informations qu'ils traitent. Selon Searle, il y a une différence entre les propriétés de l'esprit humain et celles d'un ordinateur digital. Nos états mentaux ne sont pas le résultat d'une simple manipulation des symboles par le cerveau.

Les ordinateurs digitaux sont par définition des machines basées sur des opérations formelles et syntaxiques qui manipulent des symboles sans contenu selon des régles strictes. Ce sont des machines à programme qui se comportent de façon purement formelle (syntaxique). Si la manipulation de symboles par une machine digitale permet de simuler certains aspects formels de l'intelligence ou de la compréhension d'une langue cela ne veut pas dire que la machine est intelligente ou qu'elle comprend vraiment cette langue.

Nous avons montré que les critiques de Searle à l'I.A et à la science cognitive peuvent être comprises comme des critiques au fonctionnalisme. Searle est radicalement contre l'idée fonctionnaliste selon laquelle l'esprit humain ne dépend d'aucune structure biologique spécifique, c'est-à-dire celle d'un cerveau humain. Le cerveau humain pour le

fonctionnaliste n'est qu'une sorte d'ordinateur à programmes parmi d'autres. Pour Searle, au contraire le cerveau est à la base même de l'existence de l'esprit.

Le fonctionnaliste affirme que l'esprit (ou le fait d'avoir des états mentaux) ne dépend pas du matériel (*hardware*) constituant les systèmes mais du logiciel (*software*). Il est théoriquement possible que les machines puissent avoir des états mentaux semblables à ceux de l'être humain; pour cela, il suffit que les machines soient programmées adéquatement. Searle, par contre, entend que les états mentaux ne peuvent pas être réalisés dans n'importe quelles structures physiques. Pour lui, le fait d'avoir un programme n'est ni équivalent ni suffisant pour avoir un esprit. Searle affirme que la base matérielle d'un système capable d'avoir des états mentaux doit être spécifique: elle doit avoir la structure biologique du cerveau humain.

Un aspect important des critiques de Searle au sujet de l'Intentionnalité est qu'en plus d'être causée par le cerveau, elle ne peut jamais être le résultat de l'exécution d'un programme pour machine de Turing ou d'un programme informatique quelquonque. L'IA ne peut pas dupliquer l'intelligence humaine à l'aide de manipulations formelles (syntaxiques) de symboles. Les états computationels de la machine ne peuvent, selon Searle, être comparés aux états intentionnels humains; l'exécution d'un programme informatique n'est pas une condition suffisante de l'Intentionnalité.

Il est implicite dans les critiques de Searle que nous devons abandonner toute analyse *intermédiaire* de type fonctionnaliste sur le fonctionnement de l'esprit. Il n'y pas de place, selon lui, pour des théories qui proposent l'existence d'un programme informatique, ou machine de Turing, entre l'esprit et le cerveau. Searle à la différence de Dreyfus ne critique pas l'IA dans ses nombreux domaines, il s'intéresse surtout à la recherche portant sur le langage naturel.

Searle s'oppose à l'idée fonctionnaliste selon laquelle l'ordinateur peut servir de modèle aux études sur la cognition humaine. Cette idée est liée à la croyance en une analogie entre l'esprit et les machines abstraites telle que la machine de Turing. Cette analogie a deux points d'appui: la notion de *traitement de l'information* et la notion d'*emploi de règles*. Il affirme que nous ne pouvons pas réduire l'esprit à des systèmes purement formels basés sur des règles et sur la notion de *traitement d'information*. L'esprit ne doit pas être compris, d'après lui, comme le veulent les fonctionnalistes, en termes d'états et transitions d'états d'un système formel, soit-il un ordinateur digital ou une machine de Turing.

Il faut encore rappeler que pour Searle il existe une distinction importante, celle entre IA forte et IA faible. Nous avons vu que pour Searle l'IA faible défend que l'ordinateur

digital est un instrument important capable d'aider dans la compréhension de l'esprit, tandis que pour l'IA forte l'ordinateur n'est pas un simple instrument pour l'explication de la pensée: toute machine adéquatement programmée peut littéralement penser et, par exemple, comprendre le langage naturel. C'est à l'IA forte que ses critiques sont adressées.

Cependant la thèse de Searle contre l'IA forte présente une difficulté à signaler: nous pouvons constater que dans les travaux de l'IA il n'existe pas un sens *fort* d'intelligence Artificielle. Il faut remarquer que l'affirmation que les machines sont ou ne sont pas intelligentes dépend de la définition générale donnée au concept d"intelligence". Il existe un sens *fort* et un sens *faible* d"intelligence". Le sens fort d'"intelligence" est en rapport avec notre intuition, le jugement, des aptitudes essentiellement pratiques (savoir faire) etc. , (Zone 4 du tableau annexe) Le sens *faible* d'"intelligence", par contre, est celui le moins réfractaire à la programmation. Sont des exemples d"intelligence faible": certaines tâches intellectuelles de caractère logique et associative, celles qui demanderaient de l'apprentissage au moyen de règles etc. (Zones 1 à 3 trois du tableau annexe, p.218). Toutes les tâches qui sont simulées ou programmés en IA demanderaient de l'intelligence seulement dans le sens faible. Les critiques de Searle sont en vérité, adressées aux chercheurs de l'approche descendante qui font de la recherche en Intelligence Artificielle en exploitant le sens faible d'intelligence.

C'est le sens faible d'intelligence qui est en général employé par Herbert Simon, Alan Newel, Marvin Minsky, Alan Turing, John MacCarty et bien d'autres chercheurs en IA. La plupart des machines actuelles ont, disons, une "intelligence" faible et c'est dans le sens faible qu'il faut comprendre la plupart des auteurs quand ils parlent d'«Intelligence Artificielle» et des recherches elles-mêmes. Le fait que les chercheurs exploitent seulement le sens faible de l'intelligence ne signifie pas qu'ils ne s'intéressent pas à développer de machines fortement intelligentes. Cependant le fait qu'ils aspirent, un jour, à pouvoir concevoir des machines fortement intelligentes ne veut pas dire que les recherches faites actuellement IA soient basées sur des notions d'intelligence forte. La distinction fort/faible en IA est, comme l'affirment certains auteurs problématique[360].

La conclusion que nous avons tirée du chapitre IV sur les critiques de Searle à l'IA, est que ses thèses sur l'esprit et le caractère anti-fonctionnaliste de ses critiques révèlent des rapports de l'IA avec la philosophie et l'interêt philosophique de ce thème.

---

360 J.G.Ganascia, *op.cit.* p. 221-223.

Tout au long de notre travail nous avons voulu montrer le rapport entre l'IA et la philosophie et plus précisement nous avons eu à l'esprit la question de savoir si l'IA est un thème d'intérêt philosophique ou si elle est un sujet essentiellement philosophique.

Nous avons vu que le *logos* révèle le caractère rationaliste de l'IA, tandis que le *mythe* n'est qu'un élément pré-historique qui peut réapparaître sur des travaux récents où la préoccupation avec la justification est mise de côte pour donner de la place à des spéculations exagérées sur les capacités des ordinateurs digitaux. Ce *logos* de l'IA présuppose une position représentationnaliste et le choix d'une théorie quelconque de l'esprit. Les critiques de Dreyfus et Searle mettent en évidence les rapports entre l'IA et la philosophie, le caractère respectivement, anti-représentationnaliste et anti-fonctionnaliste de leurs critiques à l'IA montrent que ce domaine de recherche demande une prise de position philosophique par rapport aux représentations et en particulier, par rapport à l'esprit.

Ainsi nous pouvons affirmer que c'est en tant que *logos* que l'IA constitue un objet de critique philosophique, qu'en plus d'être un thème philosophique, l'IA est un sujet *essentiellement* philosophique.

L'IA est un domaine d'étude qui nous oblige à poser des questions philosophiques importantes telles que: Qu'est-ce que l'esprit? Qu'est-ce qu'un raisonnement et la rationalité? Qu'est ce que la signification? Cela explique d'une part l'intérêt de beaucoup de philosophes pour le thème, d'autre part le rapport de l'IA avec la pensée philosophique.

Nous pensons que l'IA est un sujet essentiellement philosophique[361]. L'IA a en elle-même un sens philosophique; cela s'explique surtout par le bouleversement qu'elle peut provoquer dans nos habitudes et attitudes intellectuelles, c'est-à-dire sur notre vision de monde, sur la façon dont nous pensons les capacités de notre esprit, les techniques et notre propre pensée.

Nous sommes d'accord avec plusieurs auteurs lorsqu'ils affirment que l'étude de l'IA permet la redéfinition de certains problèmes philosophiques classiques et la découverte de nouvelles questions dans le domaine de la philosophie[362].

Par exemple, nous constatons que la question de la représentation et les problèmes liés à l'explication des rapports entre le cerveau et l'esprit sont des questions qui intéressent les philosophes et les chercheurs en IA. Il est évident que les approches et les objectifs visés

---

361 Cf. ,C. Glymour, "Artificial Intelligence is philosophy" *in* J. H. Fetzer (ed.), *Aspects of Artificial Intelligence*, Dordrecht, Kluwer Academic Publishers 1988, 195-207.

362 Cf. H. Dreyfus(ed), Husserl, Intentionality and Cognitive Science, Mit Press, Cambridge, Mass. 1982. R. , Van Guclick, *op. cit.* , et aussi M. Ringle, *op. cit.*

par eux sur ces sujets sont tout à fait différents, cependant les recherches philosophiques sur IA et les questions de caractère philosophique posées par ceux qui travaillent dans le domaine de l'IA peuvent quelquefois être complémentaires. L'IA sert au philosophe comme thème de discussions philosophiques. Sont caractère essentiellement philosophique est mis en relief par le fait qu'elle nous incite à ré-évaluer et reprendre certaines questions philosophiques importantes.

Ainsi l'IA est liée à la philosophie non seulement par le fait qu'elle constitue un objet intéressant du point de vue philosophique, mais par le fait que l'IA est en rapport avec toute une tradition que lie l'esprit et l'intelligence à une sorte de calcul ou à des représentations formelles. Elle est un thème essentiellement philosophique en vertu des bases philosophiques du *logos* qui la constitue. Aujourd'hui, discuter sur les rapports entre l'IA et la philosophie ne provoquerait pas les sourires d' il y a dix ans.

BIBLIOGRAPHIE

ABBAGNANO, N., *Dicionário de Filosofia*, Editora Mestre Jou, São Paulo, 1982, 980p.

ABU-MUSTAFA, Y. et PSALTIS, D., "Des ordinateurs optiques à l'image du cerveau", *in Pour la Science*, mai, 1987, pp. 71-80.

ALBUS, J.S., *Brains, Behavior, and Robotics* , BYTE Books Publications inc., s./v.,USA, 1981, 352p.

ANDERSON, A.R., éd.,"Computing Machinery and Intelligence", *Mind*, 59, Englewood Cliffs, Prentice Hall N. J.,1964, pp. 4-30.

ANDERSON, A.R., éd, *Minds and Machines*, Prentice-Hall Inc., NJ, 1964, 114p.

ANDERSON, A.R., éd, *Pensée et machine*, Ed. du Champ Vallon,1983, Traduit de l'american par Patrice Blanchard *Minds and Machines*, Champ Vallon, 150p.

ANDLER, D., "Quelle est la place de l'intelligence artificielle dans la cognition?", *in Revue Internationale de Philosophie*, nº1, 1990, pp.62-86.

BODEN, M., *Artificial Intelligence and Natural Man*, Basic Books, Inc. Publishers, NY, 1987, 2e édition, 576p.

BOOTH, A. D. et Locke, éd., *Machine Translation*, NY, North-Holland Publishing Co, NY, 1967, 529p.

BORILLO, M., "Une machine spécualtive (Informatique, intelligence artificielle et recherche cognitive)", *Revue Internationale de Philosophie*, 1/1990, nº 172, p.47-61.

BUCHANAN, B.G., "Artificial Intelligence as an Experimental Science", *in* FETZER, J., *Aspects of Artificial Intelligence*, Kuwer Academic Publishers, Dordrecht, 1988.

CHAPEVILLE, F. , NOEL, E. (et al.), *Le Darwinisme aujourd'hui*, Le Seuil, Paris,1979, 189 p.

CHOMSKY, N., *Cartesian Linguistics*, Harper & Row, New York, 1966, 119 p.

COULOUBARITSIS, L. et HOTTOIS, G., éds, *Penser l'informatique: informatiser la pensé: mélanges oferts à André Robinet*, Bruxelles, 1987, 118p.

DENNETT, D.C., *Brainstorms, Philosophical Essays on Mind and Psychology*, Bradford Books Publishers Inc. Montgomery, Vermon, 1978, 208 p.

DENNETT, D.C., "When Philosophers Encouter Artificial Intelligence", *in* GRAUBARD, Stephen R. (éd.), *The Artificial Intelligence Debate, False Starts, Real Fondations*, MIT Press, Mass, 1989, p. 283- 295.

DESCARTES, R., *Le discours de la méthode*, Œuvres Philosophiques Tome I (1618-1637), Textes établis, présentés et annotés par Ferdinad Alquié, Éditions Garnier Frères, Paris, 1963, (cinquième partie), 829p.

DREYFUS, H.L., "Alchemy and Aritificial Intelligence", *in The RAND Corporation*, cal., decembre, 1965, p. 3244.

DREYFUS, H.L., *Intelligence Artificielle: mythes et limites*, Flammarion, Paris, 1984. 443p p. (traduit de l'anglais *What Computers Can't Do, The Limits of Artificial Intelligence*, Harper & Row, Publishers, NY, 1979, par Rose-marie Vassallo-Villaneau).

DREYFUS, H.L., "L'Intelligence Artificielle (IA): le problème de la représentation du savoir", *in Encyclopædia Universelle*, vol. I, PUF, Paris, 1989, pp.973-979.

DREYFUS, H.L., "Les ordinateurs peuvent-ils être vraiment intelligents?", *in Critique*, Paris, 1980, vol. 36, nº399-400, p.730-742.

DREYFUS, H.L., *What Computers Can't Do, The Limits of Artificial Intelligence*, Harper & Row, Publishers, NY, 1979, 354 p.

EVANS, T.G., "A Program for the Solution of a Class of Geometric Analogy Intelligence Test Question" *in* Minsky, M., éd., *Semantic Information Processing*, M.I.T., Mass., 1969, pp.346-347.

FISCHLER, M.A. et FIRSCHEIN, O., *Intelligence: The Eye, the Brain, and the Computer*, Addison-Wesley Publishing Co, Mass., 1987, 331p.

FODOR, J. A, *La modularité de l"esprit, Essay sur la psychologie des facultés*, Paris, Ed. Minuit, 1986, traduit de l'american *The Modularity of Mind*, 1983, par Abel Gerchenfeld, 178p.

FODOR, J., "Le corps et l'esprit", *in Pour la science*, mai, 1981, nº 43, p. 78-88.

FODOR, J., *Representations: Philosophical essays on the Foundations of Cognitive Science*, Brighton, Mass. : Mit Press, 1981

GANASCIA, J.G., *L'âme-machine. Les enjeux de l'intelligence artificielle*, Éditions du Seuil, Paris 1990, 304 p.

GLYMOUR, C., "Artificial intelligence is Philosophy" *in* FETZER, J., *Aspects of Artificial Intelligence*, Dordrecht, Kuwer Academic Publishers, 1988, pp.195-207

GOCHET, P., (Préface au numéro dédié à l'IA) *in Revue Internationale de Philosophie*, Université de Liège, 1/1990, nº 172, Difusion Presses Universitaires de France, Liège,Hassocks, Harvester Press, 1979, 244 p, , pp. 3-4.

HAUGELAND, J., *Artificial Intelligence: The Very Idea*, Bradford Book, MIT Press,Mass, 1985, 287p.

HEINEMANN, F., *A filosofia no século XX*, Fundação Calouste Gulbenkian, Lisboa, 1983, Traduction de A. F. Morujão, de l'allemand *Die Philosophie in XX. Jahrhundert*, Zweite,Auflage, E. K. Verlag, Stuttgart, 1963, 574p.

HINTON, G., SEJNOWSKY, T. et ACKLEY, D., " A Learning Algorithm for Boltzman Machines", *in Cognitive Science*, 1985, 9, pp. 147-169.

HOMERE, *Iliade*, Bibliothèque de la Pléiade, Gallimard, Paris, 1957. Traduction, introduction et notes de Robert Flacelière, 1140p.

JACOB F.,"L'évolution sans projet" in *Le Darwinisme aujourd'hui*, Le Seuil, Paris, 1979, pp.145-147.

JAPIASSU, H. et MARCONDES, D., *Dicionário de Filosofia*, Jorge Zahar Editora Ltda, Rio de Janeiro, 1990, 265p.

KANT, E., *Critique du jugement*, (1770), 2e partie, 1º section § 64-65 (traduction de l'allemand par G. Gibellin, 4e édition,Librairie Philosophique J.Vrin, Paris 1962.

LA METTRIE, J.O. *L'homme-machine*, (1746), Hol.,Ultrecht, 1966, 170 pages.

LADRIERE, J. *Les limitations internes des formalismes* , Gauthier-Villars, Paris, 1957. 715 p.

LAZORTHES, G., *Le cerveau et l'ordinateur*, Editions Privat, Toulouse, 1988, 161p.

LE NY, J. F., *Science cognitive et compréhension du langage*, Presses Universitaires de France, Paris,1989.

LEIBNITZ, G. W., *La monadologie*, éd. Boutroux, Delagrave, 1970, 231p.

LIGNONIERE, R., *Préhistoire et histoire des ordinateurs*, éd. Robert Laffont, Paris,1987, 356p.

LONGEART, M., "Intelligence Artificielle, mythe ou réalité?", *in Carrefour*, Société Philosophique de l'Outaouais", Hull, 1989, pp.149-152.

LONGEART, M., "Intelligence et Intentionnalité: critique de l'argument de Searle contre l'Intelligence Artificielle", *Dialogue* Revue Canadienne de philosophie, vol.XXX, 1991, pp.85-102.

LUCAS, J.R., "Minds Machines and Gödel", *in Philosophy*, nº 36, 1961, pp. 112-127.

LUCAS, J.R., "Minds, Machines and Gödel", *in* ANDERSON, A. R., *Minds and Machines*, Prentice-Hall Inc., NJ, 1964, pp.43-71.

LUCAS, M.A., "A tradução computadorizada e a tradução humana"[La traduction automatique et la traduction humaine], *in Boletim de Filosofia*, UFRJ-IFCS, Nº6, decembre, Rio de Janeiro, 1986, pp. 74-81.

MACCARTHY, F. J. et HAYES, P. J., "Some Philosophical Problems from the Standpoint of Artificial Intelligence", *in* B. Meltzer and D. Michie, éds., *Machine Intelligence*, vol. 4, Halstead, NY, , 1969, pp. 463-502.

McCORDUCK, P., *Machines Who Think. A personal Inquiry into the History and Prospects of Artificial Intelligence*, W. H. Fremman & Company, NY, 1979, 375 p.

MENDONÇA, W. P. , "Intelligence Artificielle et signification: À propos des limites et des possibilités des sciences cognitives", *in Revue Philosophiques*, vol XVII, Numéro 1, printemps, 1990, pp.3-19.

METROPOLIS, N. H. et ROTA G. C. éds., *A History of Computing in the Twentieth Century*, Academic Press inc., N.Y., 1980, 659 pages

METTRIE, J. O., *L'Homme machine*, Utrecht, Hol. et J. Pauvert eds,1966, 170p.

MICHIE, D., *Reflexions sur l'intelligence des machine*, Masson, Paris, 1990 237 p.

MINSKY, M. et PAPERT, S., *Perceptrons*, MIT Press, Cambridge, Mass., 1969, 292p.

MINSKY, M., "Mather, Mind, and Models", *in Semantic Information Processing*, MIT Press, Mass, 1980, pp.425-432.

MITCHAM, C., "Aspects philosophiques de la tecnique", *in Revue Internationale de Philosophie*, nº 161, 1987.

NEWEL, A., SHAW, J. C. et SIMON, H.A., "Chess Playing Programs and the Problem of Complexity", *in* FEIGEMBAUM, E. A. and FELDMAN, J., éd., *Computers and Thought,* McGraw-Hill, NY, 1963, p.10-39.

OETTINGER, A.G., *Automatic Language Translation*, Harvard University Press, Cambridge, 1960, 380p.

PEREZ, J.C., *De nouvelles voies vers l"intelligence artificielle*, Masson, Paris, 1988, 247p.

PERSONNAZ, L. et DREYFUS, G., "Les machines neuronales", *in La Recherche*, société d'Editions scientifiques, Paris, 1988, nº 204, novembre, vol.19, pp.1362-1371.

PLATON, *La République*, Garnier, Paris, 1958, 528 p. (Traduit par E. Chambry)

PYLYSHYN, Z.W., *Computation and Cognition: Toward a Foundationion for Cognitive Science,* Cambridge, Mass., 2º éd., 1985, 292p.

RENE, M., *in* PEREZ, J.C., *De nouvelles voies vers l'Intelligence Artificielle*, Masson, Paris, 1988, 247p., "préface à l'ouvrage", pp.5-8.

RINGLE, M., éd., *Philosophical Perspectives in Artificial Intelligence*, The Harvester Press, Brighton, 1979, 244p.

RYLE G., *La notion d'esprit: Pour une critique des concepts mentaux*, traduit de l'anglais *The concept of Mind* par Stern-Gillet, S. Payot, Paris, 1978, 314p.

SEARLE, J.R., *Minds, Brains, and Science*, Harvard University Press, Cambridge, Mass., 1984, 107p.

SEARLE, J.R., *Du cerveau au savoir*, Hermann, Paris, 1985. Traduit de *Minds., Brains, and Science*, par Catherine Chaleyssin, 143p.

SEARLE, J.R., et VANDERVEKEN, D. *Foundations of Illocutionary Logic*, Cambridge University Press, Cambridge, 1985, 227p.

SEARLE, J.R., *Expression and Meaning*, Cambridge University Press, Cambridge, 1979.

SEARLE, J.R., *Intentionality: an Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge, 1983,. 278p.

SEARLE, J.R., "Minds., Brains, and Programs", *in Minds Behavioral and Brain Sciences*, Harvard University Press, Cambridge, Mass.,1980, nº 3, p.417-424.

SEARLE, J.R., " Consciousness, unconciousness, and Intentionality " , *Philosophical Topics*, volume XVII, nº1, spring 1989, pp. 193-209.

SELFRIEDGE, O.G. et NEISSER, U., "Pattern Recognition by Machine" dans *Computers and Thought* , *in* FEIGENBAUN, E. A. and FELDMAN, J., éd., McGraw-Hill, NY, 1963. pp. 238-245.

SHANNON, C. WEAVER, W. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana,1949, 125 p.

SHAPIRO, S.C. et ECKROTH, D., *Encyclopedia of Aritficial Intelligence*, J. Wiley & Sons Inc., vol. I et II, , 1219p , USA, 1987.

SMITH, J-C., éd., *HistoricalFoundations of CognitiveScience*, Kluwer Academic Publishers, Dordrecht, 1990, 303p.

TAYSE, A., *Approche logique de l'intelligence artificielle* Tome I: "De la logique classique à la programation logique", Dunod Informatique, Paris, 1988, 386p.

TAYSE, A., *Approche logique de l'intelligence artificielle* Tome II: "De la logique modale à la logique des bases de données", Dunod Informatique, Paris, 1989, 427p.

UHR, L. and VOSSLER, C., " A Pattern Recognition Program that Generates, Evaluates and Adjusts its Own operations", *in* FEIGENBAUM, E. A. and FELDMAN, J. *Computers and Thought* , J. ed., McGraw-Hill, pp.248-251, New York,1963,

VARELA, F. *Connaître*, Éditions du Seuil, Paris, 1989, Traduction de l'anglais: *Cognitive Science. A Cartography of Current Ideas*, 1988, par Pierre Lavoie., 122p.

WALWORTH, A.Y, "The Prospects for Artificial Intelligence, thèse de Doctotorat, University of California at Berkeley, 1989, 362p.

WEIZENBAUM, J., *Puissance de l'ordinateur et raison de l'homme: du jugement au calcul. Editions d'informatique*, Paris, 1981, 195 p. Trad. de l'americain,*Computer Power and Human Reason: From Judgment to Calculation*,(W H Freeman, San Francisco1976.) par Marie Therese Margulici .

WIENER, N., *Cybernetics*, The M.I.T. Press, Cambridge, Mass., 1948, 212p.

WINOGRAD, T. et FLORES, F., *L'Intelligence Artificielle en question*, PUF, Paris, 1986, 295p.

WINOGRAD, T. et FLORES, F., *Understanding Computers and Cognition: a New Foundation for Design*, Ablex Publishing Corporation, Norwood, N.J, 1986, 207p.

WINSTON, P., *The Psychology of Computer Vision* , P. H. Winston ed. Mc Graw Hill, , 282 p., NY.,1975

WINSTON, P.H. *Intelligence artificielle*, Inter Editions, Paris, 1988, 528 p. Traduit de l'american, *Artificial intelligence,(* Mit Press, Cambridge, Mass.1979) par d'Annie Danzart et Jean-Michel Moreau.

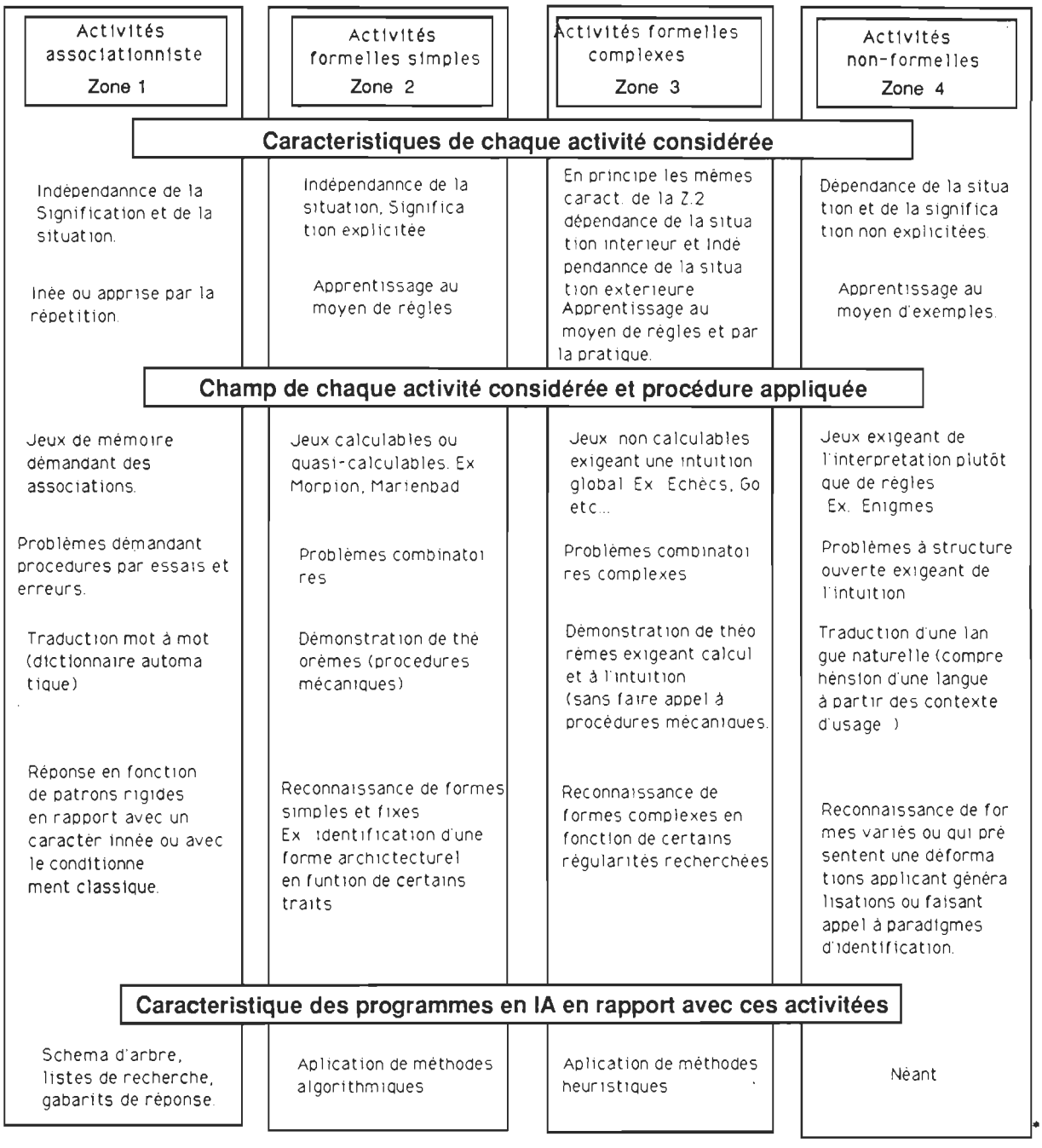
WINSTON, P.H., *Intelligence artificielle*, Inter Editions, Paris, 1988, 528p.

WITTGENSTEIN, L, *The Blue and Brown Books*, Basil Blackwell, Oxford, 1969,.192 p.

WTTGENSTEIN, L., *Investigations philosophiques*, Librairie Gallimard, Paris,1961 traduit de l'allemand par Klossomski, Pierre.

ANNEXE

Classification des activités intelligentes selon Dreyfus :



* Ce Tableau a été construit à partir du tableau conçu par H. Dreyfus. Cf. Dreyfus (1979), *op. cit.* ,p.292