

ACTION ET CONTEXTE.
DU TOURNANT COGNITIVISTE À LA PHÉNOMÉNOLOGIE TRANSCENDANTALE

Tom Dedeurwaerdere (FRS-FNRS/UCL)

Bibliographical reference

Tom Dedeurwaerdere. 2002. *Action et Contexte. Du tournant cognitiviste à la phénoménologie transcendantale*, Hildesheim/Zürich/New-York, Georg Olms Verlag, 237 p.

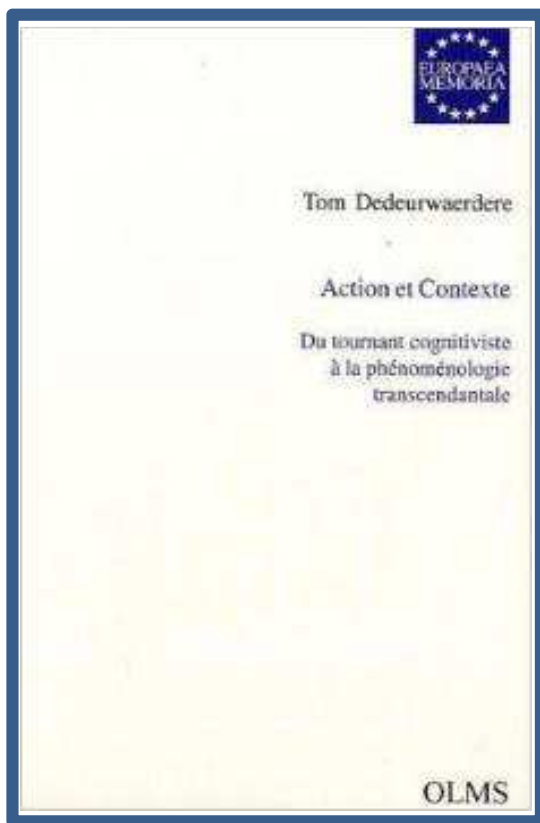
Self-archived author copy

This copy is for your personal, non-commercial use only.
For all other uses permission shall be obtained from the copyright owner.

ISBN-13: 978-3-487-11475-0, ISBN: 3-487-11475-5

Hildesheim/Zürich/New-York, Georg Olms Verlag

Copyright © - All rights reserved.



**ACTION ET CONTEXTE.
DU TOURNANT COGNITIVISTE À LA PHÉNOMÉNOLOGIE TRANSCENDANTE**

Tom Dedeurwaerdere

Hildesheim, 2002 (ISBN 3-487-11475-5), 237 p.

Dans la ligne de D.Fisette, P.Poirier et J.-L.Petit, l'ouvrage de T.Dedeurwaerdere s'inscrit dans le projet général d'une exploration des sources européennes des développements actuels en philosophie de l'esprit. En dialogue avec H.Putnam et J.Searle, mais aussi avec J.Ladrière et P.Engel, l'auteur montre le rôle central de la problématique classique de l'intentionnalité dans l'entreprise de naturalisation menée par les sciences cognitives. De la sorte, il parvient à déconstruire cette entreprise et à renouer au-delà des tentatives formalistes inspirées de la systémique avec les conclusions de la phénoménologie husserlienne, en particulier celles des ouvrages *Expérience et jugement* et *De la synthèse passive*. L'ouvrage montre donc que c'est en quelque sorte la tradition reprise et réinterprétée qui parvient à dépasser, à travers la mise en question des dimensions contextuelles et sémantiques, les modélisations contemporaines de l'intentionnalité.

Table des matières

Avant-propos	1
1. Première perspective : l'action et le contexte dans les sciences cognitives	1
2. Deuxième perspective : la théorie de l'action	4
2.1. La réversibilité entre l'action et le contexte	4
2.2. Limitations internes et limitations contextuelles de la modélisation.....	7
3. Perspectives de recherche	9
Introduction.....	11
1. La problématique des limitations de la naturalisation de l'intentionnalité en philosophie de l'esprit	12
2. L'approche des limitations à partir de la problématique de la modélisation	14
3. Aperçu de la structure de l'ouvrage.....	18
Chapitre 1 : Du computationnalisme aux modèles écologiques de la cognition.....	21
1. Introduction.....	21
2. Le modèle de l'ordinateur	23
3. De l'intelligence habituelle à l'intelligence productive	28
3.1. Les modèles connexionnistes : des mémoires associatives à l'auto-organisation.....	29
3.1.1. Les mémoires associatives	29
3.1.1.1. Un réseau de reconnaissance de visages.....	30
3.1.1.2. Le neurone binaire et l'apprentissage associatif	33
3.1.1.3. Des propriétés inattendues.....	35
3.1.2. L'apprentissage par auto-organisation.....	39
3.1.3. L'intelligence habituelle	42

3.2. L'extension des modèles connexionnistes.....	44
3.2.1. Le corps simplement habituel : le cas Schneider	45
3.2.2. L'adaptation contextuelle et l'activité en situation	50
3.2.2.1. La transformation de nos habitudes.....	50
3.2.2.2. La négociation active en situation	53
3.3. Conclusion :	
le connexionnisme et les formes d'intelligence.....	56
4. Au-delà des mécanismes internes : l'écologie de la pensée	59
4.1. L'écologie subjective	62
4.1.1. Les représentations partielles	64
4.1.2. Les représentations pratico-sociales.....	69
4.2. La détermination écologique	74
4.2.1. Limiter la charge représentationnelle :	
le couplage dynamique.....	77
4.2.1.1. Les modèles cybernétiques.....	77
4.2.1.2. Les systèmes décentralisés sous contrôle	
de l'environnement	79
4.2.1.3. Les boucles de rétroaction élargies.....	82
4.2.2. La structuration active de l'environnement :	
les dispositifs.....	85
4.2.2.1. Les dispositifs.....	85
4.2.2.2. La conscience réflexive	88
4.2.3. Les systèmes émergents	90
4.2.3.1. L'émergence directe et l'émergence indirecte.....	91
4.2.3.2. Les variables globales.....	94
Chapitre 2 : Les conséquences méthodologiques.	
Du fonctionnalisme à la modélisation contextuelle	99
1. Introduction.....	99
2. Le défi de la cognition : modéliser les computations.....	101

2.1. La théorie computationnelle de David Marr et ses critiques	101
2.1.1. La computation comme traitement d'information	102
2.1.2. Les critiques de l'approche computationnelle	105
2.1.3. Le problème du fonctionnalisme	110
2.2. L'articulation entre la computation et l'implémentation	112
2.2.1. L'explication fonctionnelle en biologie	113
2.2.2. Le concept de système téléonomique	117
2.2.3. L'émergence des capacités de traitement d'information à partir du cerveau	121
3. Le défi de la modélisation contextuelle : modéliser les représentations et les fonctions	125
3.1. L'interaction entre les représentations et la critique fonctionnaliste	125
3.1.1. La décomposition modulaire dans les études structure - fonction	127
3.1.1.1. La méthode de la dissociation en neuropsychologie clinique	127
3.1.1.2. La méthode de la soustraction en imagerie cérébrale	131
3.1.2. Les modules interactifs	133
3.1.2.1. L'interaction des modules dans le système visuel	134
3.1.2.2. Les solutions globales de coordination du mouvement	136
3.2. Le couplage fonctionnel et la critique par les théories de la clôture opérationnelle	142
3.2.1. Définir l'hypothèse de la clôture opérationnelle	143
3.2.2. La clôture dans les systèmes cognitifs	145
Chapitre 3 : Limitations internes et limitations contextuelles de la modélisation de l'intentionnalité	151
1. Introduction	151
2. La critique du traitement standard de la limitation	155

2.1. Du modèle standard à la conception sémantique.....	155
2.2. La prise en compte de la dimension sémantique des modèles.....	161
2.2.1. Les conditions d'élaboration des modèles particuliers ...	163
2.2.2. Les conditions de la productivité sémantique des formalismes.....	165
3. Du formalisme aux modèles particuliers	170
3.1. Les limitations internes de la productivité sémantique du formalisme en sciences cognitives.....	172
3.2. Modèles et interprétation dans la modélisation de l'intentionnalité.....	177
3.2.1. Le caractère contextuel de la modélisation de l'intentionnalité	178
3.2.2. Le rôle de l'interprétation	180
4. Pour un traitement réflexif de la contextualisation : la solution phénoménologique et sa radicalisation	184
4.1. Le dépassement de la question des limitations dans la phénoménologie génétique	186
4.1.1. L'horizon eidétique de la logique formelle moderne.....	188
4.1.1.1. La manifestation effective du sens : les orientations fondamentales de la logique formelle moderne	188
4.1.1.2. Les présuppositions idéalisantes de la logique formelle moderne.....	190
4.1.2. La réduction phénoménologique transcendantale	193
4.1.2.1. De l'horizon de la logique formelle à l'horizon du monde de l'expérience	193
4.1.2.2. L'expérience perceptive comme donation des choses dans leur être même	195

4.2. La structure d'éveil de la conscience comme forme de vie originaire.....	198
4.2.1. Les synthèses passives dans le flux de la conscience.....	199
4.2.2. La vie originaire comme vie en éveil dans le présent vivant du temps	201
4.3. L'habitus pré-réflexif et la radicalisation de la méthode phénoménologique.....	203
4.3.1. L'ouverture à la sphère de déterminabilité de la conscience	204
4.3.2. Une théorie radicale de la connaissance.....	208
Conclusion	211
1. Le premier axe : l'opération de modélisation de l'intentionnalité dans les sciences cognitives	211
2. Le deuxième axe : les limitations contextuelles de l'opération de modélisation	213
3. Contribution de la théorie écologique à l'hypothèse de la réversibilité asymétrique	214
Bibliographie	219
Index onomastique.....	229
Table des matières.....	233

Avant-propos

L'ouvrage que vous allez lire a un statut particulier, dans la mesure où il s'inscrit dans un projet de recherche collective¹. C'est pourquoi nous voulons tout d'abord le situer dans le contexte de ce projet de recherche. Nous réserverons ainsi l'exposé de la méthodologie de travail et des heuristiques de recherche sur lesquelles nous nous sommes appuyées pour l'introduction proprement dite. L'avant-propos a donc une certaine autonomie par rapport au reste de l'ouvrage. Il vise surtout à donner un aperçu des travaux qui ont eu une influence sur sa genèse, sans toutefois toujours y apparaître de façon toujours explicite.

1. Première perspective : l'action et le contexte dans les sciences cognitives

Dans cet ouvrage, nous posons la question du statut de l'action et du contexte dans les modèles contemporains de l'intentionnalité en sciences cognitives. La remise en question de certains des présupposés les plus fondamentaux du cognitivisme désormais "classique" dans différentes recherches comme celles de la nouvelle robotique², du néo-connexionnisme³, de l'éthologie cognitive⁴ ou encore de la psychologie du développement⁵ constitue notre point de départ. Ce qui est visé par ces recherches, ce n'est pas tant la proposition de nouveaux mécanismes formels de l'intelligence, qu'une meilleure construction du rapport entre la représentation et l'action dans les modèles de l'intelligence, ainsi qu'une prise en compte de l'importance du contexte.

1 Il s'agit d'un programme de recherche PAI (Pôle d'Attraction Interuniversitaire) financé par les services du premier ministre belge et intitulé "Théorie de la Norme et Régulation Démocratique" (cf. le document *Theory of the Norm and Democratic Regulation*, Louvain-la-Neuve, 2001, 165 pp. ; la version française de ce document est disponible sur le site du Centre de Philosophie du Droit à l'adresse <http://www.cpdr.ucl.ac.be>).

2 Cf. par exemple l'article séminal de ce courant de recherche, BROOKS R., "Intelligence without Representation", in *Artificial Intelligence*, 47 (1991), pp. 139–159.

3 Cf. VARELA F., THOMPSON E. et ROSCH E., *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, Paris, 1993.

4 Cf. HUTCHINS E., *Cognition in the Wild*, Cambridge (MA), 1995.

5 Cf. THELEN E. et SMITH L., *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge (MA), 1994.

A titre préliminaire, nous pouvons illustrer l'enjeu de ces recherches à partir de l'exemple du formalisme du réseau de neurones, au cœur de la remise en question du cognitivisme classique⁶. Or, on peut distinguer trois usages différents du formalisme du réseau de neurones en sciences cognitives. Ces trois usages indiquent d'ailleurs une progression dans la critique du cognitivisme classique. Tout d'abord, d'après l'interprétation classique de Fodor, les réseaux de neurones fournissent simplement un moyen d'implémentation différent des propriétés de l'intelligence, dont le caractère essentiel reste le traitement symbolique de l'information. Ensuite, dans l'interprétation connexionniste, l'usage du formalisme du réseau de neurones donne lieu à une remise en question radicale de la conception symbolique de l'intelligence, en mettant l'accent sur les processus non symboliques, comme les processus d'acquisition de connaissances, de mémorisation et de généralisation associative. Une troisième interprétation, finalement, va au-delà de cette attention renouvelée pour les processus non symboliques dans le fonctionnement de l'intelligence, pour remettre en question le représentationnalisme qui caractérise toujours l'option connexionniste. Deux observations permettent d'appuyer une telle interprétation. Tout d'abord, les mécanismes d'apprentissage dans les réseaux supposent un ajustement progressif du système cognitif à son environnement opératoire. Dans le modèle des réseaux de neurones, l'environnement n'est donc pas seulement une source de données pour un système de traitement d'informations, mais il a également un rôle de stabilisation d'un processus d'ajustement ou de coordination entre les activités cognitives d'un agent et son environnement opératoire. Ensuite, le processus d'ajustement permet de donner un rôle différent à l'environnement opératoire. En effet, l'environnement n'a pas seulement un rôle passif dans ce processus, il n'intervient pas seulement comme "constante" dans les équations, mais sa dynamique peut jouer un rôle actif dans le processus d'ajustement. A partir de ces deux remarques, on peut entrevoir que le formalisme des réseaux de neurones permet de donner un statut différent à la représentation, en la subordonnant à un processus de coordination interne à l'action, et, en même temps, permet d'intégrer le contexte dans la modélisation, à titre d'environnement opératoire des opérations cognitives avec ses propriétés dynamiques intrinsèques. Notre perspective s'inscrit dans cette troisième interprétation, qui se caractérise par une prise en compte du rôle de l'action et du contexte dans la modélisation.

La conception de l'intentionnalité comme action en contexte se retrouve également dans d'autres pratiques de formalisation novatrices que nous

6 Cf. par exemple, FISETTE D. et POIRIER P., *Philosophie de l'esprit. Etat des lieux*, Paris, 2000, pp. 299–301.

avons citées ci-dessus, comme l'approche de la cognition dans les termes du formalisme des systèmes dynamiques ou à partir de la théorie des automates autonomes. Toutefois, à la différence de la plupart de ces approches, nous ne considérons pas que l'unité de ces recherches réside dans la proposition d'une nouvelle théorie alternative de la cognition. Tout d'abord, il ne semble pas, si nous suivons entre autres le diagnostic de F. Keijzer et de S. Bem, que les différents modèles de la cognition s'entendent sur ou permettent de définir une telle théorie alternative⁷. Ensuite, la déconstruction de certaines notions de l'ancien paradigme, comme la notion de représentation ou la notion d'environnement comme source de données, même si elle ne permet pas de construire une nouvelle théorie, permet néanmoins de mettre en évidence un déplacement épistémologique qui vise à rencontrer les insuffisances de l'opération de modélisation de l'intentionnalité dans le paradigme classique. Cette hypothèse de travail rejoint le projet de Hilary Putnam dans *Représentation et réalité*, où il montre les insuffisances des différentes améliorations des modèles fonctionnalistes de l'intentionnalité en sciences cognitives quand ils continuent à adopter le même cadre épistémologique que celui des sciences de la nature. C'est ce qui amène cet auteur à formuler une hypothèse épistémologique alternative à titre de "programme de travail", qui devrait pouvoir mieux prendre en compte la spécificité du domaine de l'action intentionnelle. Sur les traces de H. Putnam⁸, nous pouvons formuler à titre préliminaire l'hypothèse d'un tel déplacement par une conditionnalité à deux niveaux :

- condition d'autoréférentialité : une théorie qui entend modéliser l'intentionnalité doit pouvoir s'interroger elle-même comme production intentionnelle ;
- condition d'usage : l'application des modèles de l'intentionnalité sera conditionnée par une certaine interprétation de ces modèles interne à leur usage dans le contexte concret de l'action.

Ces deux conditions épistémologiques renvoient le projet de la modélisation de l'intentionnalité à la question des conceptions de l'action en amont de la modélisation d'une part et à son application au domaine de l'action de l'autre. Même si les travaux empiriques discutés dans cet ouvrage ne sont pas les mêmes que ceux considérés par Putnam à l'époque de *Représentation et réalité*, il nous semble que cette double conditionnalité,

7 Cf. KEIJZER F. et BEM S., "Behavioral Systems Interpreted as Autonomous Agents and as Coupled Dynamical Systems : a Criticism", in *Philosophical Psychology*, 9 (1996), pp. 323–346.

8 Cf. PUTNAM H., *Representation and Reality*, Cambridge (MA), 1989, pp. 113–116.

qui vise à prendre en compte la spécificité du domaine de l'action, reste tout à fait pertinente pour dépasser les insuffisances épistémologiques de la modélisation de l'intentionnalité dans les sciences cognitives.

2. Deuxième perspective : la théorie de l'action

C'est par notre intérêt pour les développements récents en théorie de l'action, par le biais particulier des modèles cognitivistes de l'intentionnalité, que notre projet a croisé les recherches menées au Centre de Philosophie du Droit de l'Université catholique de Louvain. Plus précisément, cette deuxième source de notre travail est constituée par un ensemble de recherches menées de 1995 à 1997 sur la critique du formalisme en théorie de l'action et les conséquences de cette critique au niveau d'une théorie de la norme⁹. Les résultats de ces recherches nous ont permis de préciser nos hypothèses de travail sur deux plans : tout d'abord, sur le plan de la modélisation de l'intentionnalité, par une meilleure construction de l'opération de contextualisation et ensuite, sur le plan du déplacement épistémologique, à partir des concepts de limitation interne et de limitation contextuelle de la modélisation.

2.1. La réversibilité entre l'action et le contexte

Le premier résultat de ces recherches, une meilleure construction de l'opération de contextualisation en théorie de l'action, a été développé à partir de deux domaines particuliers, d'une part le formalisme de l'intentionnalité en sciences cognitives et d'autre part le formalisme de la communication dans la pragmatique communicationnelle.

Tout d'abord, la relecture par M. Maesschalck des développements récents en sciences cognitives, dans une perspective orientée vers la théorie de l'action, permet d'aller au-delà du langage métaphorique qui caractérise la

9 Pour les résultats de ces recherches cf. MAESSCHALCK M., *Normes et contextes. Les fondements d'une pragmatique contextuelle*, Hildesheim, 2001 ; ID., "Ressources compréhensives et contextualisation des normes. Une limitation du formalisme procédural", in *Ethica*, 11 (1999), pp. 33 à 58 ; ID., "Compétences langagières et ressources compréhensives", in *Science et Esprit*, 49 (1997), pp. 259-280 ; ID., "Pour une critique du formalisme en éthique. En hommage à Jean Ladrière", in *Science et Esprit*, 48 (1996), pp. 135-152 ; LENOBLE J., "Law and Undecidability : Towards a New Vision of the Proceduralisation of Law" in ARATO A. et ROSENFELD M. (eds.), *Habermas on Law and Democracy : Critical Exchanges*, Berkeley, 1998, pp. 37-81 ; ID., "La procéduralisation contextuelle du droit", in COPPENS PH. et LENOBLE J. (eds.), *Démocratie et procéduralisation du droit*, Bruxelles, 2000, pp. 97-124.

plupart des modèles du projet cognitiviste élargi dont nous sommes partis. En effet, c'est seulement par une construction de la rationalité du domaine de l'action pour elle-même que l'on peut aller au-delà d'une pratique de modélisation de l'intentionnalité qui se base sur la généralisation de notions intentionnelles véhiculées par le langage ordinaire. C'est bien cet usage métaphorique des notions intentionnelles qui a été critiqué à juste titre par des auteurs comme Paul et Patricia Churchland¹⁰. Toutefois, la critique de ces auteurs réduit la question de l'étude de l'intentionnalité à un choix entre d'une part des modèles qualitatifs basés sur le langage intentionnel du sens commun et d'autre part des modèles quantitatif rigoureux, qui permettent de remplacer les notions approximatives du sens commun par des notions plus scientifiques. Or, au lieu d'adopter une solution éliminativiste des notions du sens commun, comme le font ces auteurs, une solution plus prometteuse semble être de partir d'une construction de la rationalité de l'agir pour elle-même, à partir des développements récents en théorie de l'action. Ce que ces développements permettent d'entrevoir, que ce soit à partir d'une critique du fonctionnalisme comme chez H. Putnam ou P. Jacob, ou à partir d'une reconstruction analytique des principes de la cybernétique comme chez A. Rosenberg, c'est une spécification de la rationalité de l'agir intentionnel comme un niveau de causalité à part entière¹¹. En effet, comme l'écrit M. Maesschalck en synthétisant ces travaux, le domaine de l'agir "concerne plus largement l'ensemble des procédures de comportement animal qui supposent la *détermination* de deux niveaux d'effectivité : l'un physique et l'autre organisationnel. Dans un comportement, un programme dont la fonction est définissable pour elle-même est mise en œuvre (implémentation) par des éléments chimiques dont l'enchaînement constitue un événement définissable pour lui-même"¹². Une construction du domaine de l'action pour lui-même permet donc de distinguer entre la causalité intentionnelle d'une part, qui est définissable pour elle-même à titre de programme d'action inscrit dans l'organisation propre de tel ou tel être vivant, et la causalité physique, de l'autre, qui détermine le comportement de chacune des composantes de cette organisation.

Le domaine de l'agir se caractérise donc par l'association de *deux* niveaux de causalité, l'un physique et l'autre intentionnel. Toutefois, une telle définition du domaine de l'agir s'expose encore à une réduction fonctionnaliste de l'intentionnalité. En effet, elle nous laisse encore la

10 Cf. CHURCHLAND P.S., *Neurophilosophie. Vers une science de l'esprit cerveau*, Paris, 1999 ; CHURCHLAND P.M., *A Neurocomputational Perspective*, Cambridge (MA), 1989.

11 Cf. MAESSCHALCK M., *Normes et contextes*, op. cit., p. 176.

12 *Ibid.*

possibilité de définir la causalité intentionnelle dans les termes de mécanismes computationnels d'un système de traitement d'information. Même si on admet dans ce cas, comme dans le fonctionnalisme classique, que le niveau des mécanismes computationnels ne peut pas se réduire au niveau des processus neuro-chimiques du système nerveux central, on procède néanmoins à une autre réduction : une réduction des états intentionnels aux états computationnels d'un système de traitement d'information. Or, comme le note bien M. Maesschalck, une telle réduction suppose "qu'à des états computationnels déterminés devraient correspondre des attitudes propositionnelles déterminées"¹³, c'est-à-dire des états intentionnels déterminés. Cependant, la fixation d'une attitude propositionnelle ou d'un état intentionnel ne dépend pas de l'identification d'une suite d'états computationnels particuliers, mais d'une pratique interprétative qui "définit les conditions auxquelles une croyance est admise comme fixée"¹⁴. En effet, la capacité de se comprendre, de parler d'une même chose ou d'avoir une même croyance au sujet d'un état de choses n'implique pas d'avoir des contenus mentaux en tout point identiques. En fait, comme l'écrit M. Maesschalck, "la capacité d'interpréter ces croyances liées à des suites d'états computationnels différents réside précisément dans le refus d'une approche réductrice des croyances qui consisterait à expliquer les états intentionnels à partir d'une théorie non intentionnelle de l'intentionnalité"¹⁵. Par conséquent, on se retrouve devant la tâche "d'élaborer une théorie de l'intentionnalité qui demeure interne à son usage"¹⁶. La relation entre les états intentionnels et leurs contextes d'usage peut donc se comprendre dans les termes d'un rapport de conditionnalité réversible : la fixation d'un état intentionnel dépend de l'usage, et donc du contexte plus large des significations intentionnelles dans lequel il s'insère, mais ne se réduit pas pour autant à ce contexte d'usage. En effet, "la tâche interprétative consiste, de ce point de vue, à définir un certain rapport intentionnel à l'usage, non l'usage lui-même"¹⁷.

Ce premier résultat a également été développé dans un autre domaine, à partir d'une critique du formalisme de la communication dans la pragmatique communicationnelle de Jürgen Habermas. L'importance de ce deuxième ensemble de recherches est de pouvoir faire plus directement le lien avec la question du déplacement épistémologique. Tout d'abord, comme l'ont montré J. Lenoble et A. Bertin, la pragmatique communicationnelle tente bien de

13 *Ibid.*, p. 178.

14 *Ibid.*, p. 179.

15 *Ibid.*, p. 178.

16 *Ibid.*

17 *Ibid.*, p. 179.

prendre en compte la spécificité du domaine pragmatique de l'agir. Cependant, la prise en compte de cette spécificité ne se fait qu'au prix d'une réduction de celle-ci à un formalisme procédural de l'agir communicationnel¹⁸. La pragmatique communicationnelle n'intègre donc pas dans sa construction la réversibilité entre les règles formelles de procéduralisation et leur contexte d'application. En s'appuyant sur la philosophie du second Wittgenstein, cette réversibilité a été formulée en un premier temps dans les termes d'un ajustement entre la règle et son monde dans l'effectuation même de la règle¹⁹. Toutefois, cette réversibilité reste incomplète tant que l'opération de contextualisation de la règle n'est pas construite pour elle-même à titre d'opération de la raison communicationnelle. Une construction plus fine de la réversibilité, en s'appuyant sur la théorie de l'indécidabilité de Pierre Livet, a alors permis de mettre en évidence différents niveaux de contextualité internes aux opérations de la raison communicationnelle, qui constituent en quelque sorte autant de paliers de stabilisation de l'ajustement entre le monde et la règle²⁰.

La théorie des niveaux d'indécidabilité de Livet permet de donner une construction plus fine de la réversibilité entre l'action et le contexte. Toutefois, la prise en compte du contexte reste encore interne au projet d'une formalisation spécifique. La spécificité du domaine de l'action exige de faire un pas de plus et de prendre en compte l'autonomie propre du contexte de l'action. Si l'on maintient la pertinence d'une démarche de formalisation pour la théorie de l'action, il faut encore pouvoir spécifier la relation entre un formalisme spécifique et le moment de son application au domaine de l'action.

2.2. Limitations internes et limitations contextuelles de la modélisation

C'est précisément l'articulation entre formalisation et contexte d'application que la critique par M. Maesschalck de la théorie de Livet a permis de formuler de façon plus adéquate, à partir des travaux de J. Ladrière sur les limitations des formalismes. Ce que montrent en particulier ces travaux sur les limitations sémantiques des formalismes, c'est que la productivité sémantique des formalismes mathématiques est toujours liée à un champ spécifique d'expérimentation. En effet, comme l'écrit J. Ladrière, "lorsqu'on construit un système formel, c'est en général avec l'intention de

18 Cf. BERTEN A. et LENOBLE J., *Dire la norme*, Paris, 1990.

19 *Ibid.*, p. 188.

20 Cf. LENOBLE J., *Droit et communication*, Paris, 1994, pp. 42–45.

représenter dans ce système une théorie non formalisée”²¹. Or, “ainsi que les théorèmes de limitation le montrent, il n'est pas possible d'obtenir une représentation formelle adéquate d'une théorie dès que celle-ci a une certaine ampleur”²². On ne pourra donc pas définir un ensemble de règles de correspondance entre la théorie formelle et la théorie intuitive, pour l'ensemble des contextes d'application du formalisme. Le formalisme, par son autonomie opératoire, peut aller au-delà de son contexte d'application naturelle, mais son application à de nouveaux contextes dépendra d'un nouvel ensemble de règles de correspondance qui permettront de définir le champ d'expérimentation dans lequel il pourra s'effectuer. “L'opération de la raison, comme l'écrit J. Lenoble en synthétisant les résultats de ces recherches, ne se laisse pas réduire au seul niveau de l'opérateur pur. L'effectuation de la raison nécessite un second niveau de conditionnalité qui va conditionner l'usage sur lequel débouchera le produit de l'opération formelle. Ce second niveau de conditionnalité est constitué des ressources offertes par le contexte comme ‘structures de capacitation’ rendant possible l'application du ‘possible formel’ qui résulte de l'opération formelle de la raison”²³. L'opération formelle de la raison donne donc lieu à une double limitation, d'une part une limitation interne, liée aux limitations de la fécondité d'un formalisme donné, comme chez Livet, et d'autre part une limitation contextuelle, liée aux conditions d'usage ou de capacitation de la raison formelle dans le monde de l'action.

Par les concepts de limitation interne et de limitation contextuelle nous retrouvons bien la condition d'autoréférentialité et la condition d'usage que nous avons formulées ci-dessus à la suite de Putnam. Cependant, là où cette première formulation se situe encore dans la perspective d'une critique générale de l'opération de la raison, la formulation plus précise à partir des théorèmes de limitation devrait nous permettre d'articuler le déplacement épistémologique lié à la spécificité du domaine de l'action à un formalisme spécifique et à un champ expérimental spécifique, celui des modèles de l'intentionnalité en sciences cognitives.

21 LADRIÈRE J., “Les limites de la formalisation”, in PIAGET J. (dir.), *Logique et connaissance scientifique (Encyclopédie de la Pléiade)*, Paris, 1967, pp. 312–333, p. 313.

22 *Ibid.*

23 LENOBLE J., “La théorie de la loi et l'hypothèse de la procéduralisation contextuelle”, in *Les Carnets du Centre de Philosophie du Droit*, n° 70 (1999), p. 14.

3. Perspectives de recherche

La critique épistémologique de la conception mentaliste du contexte à partir du concept de limitation contextuelle de la formalisation permet de réinterpréter autrement l'intérêt des sciences cognitives pour la théorie de l'action. En effet, la limitation contextuelle de la modélisation laisse ouverte la possibilité d'un usage de la modélisation différent de son usage dans un cadre d'échange cybernétique, où le contexte est traité comme un simple réceptacle. "De fait, comme le note M. Maesschalck, un autre aspect de l'approche cognitive de l'intentionnalité pourrait être mis en évidence : son rapport interne au champ de l'effectuation, c'est-à-dire la mise en situation de l'action dans un *environnement opératoire*"²⁴. De ce point de vue, "le contexte dans l'approche cognitive reçoit un autre statut : c'est un milieu qui *interagit* avec l'action et dont le 'comportement' doit être pris en compte par celle-ci, parce que l'action utilise des procédures du milieu susceptibles d'inférer à leur tour sur le développement interne de l'action"²⁵. La prise en compte de la cohérence de l'action à l'intérieur d'un environnement autoprogrammé introduit une asymétrie dans la relation entre l'action et son contexte : "un même résultat a donc une réversibilité et une asymétrie : il est un effet produit, mais en retour il est une information adressée"²⁶.

Le déplacement épistémologique que nécessite une construction de l'opération de modélisation en sciences cognitives permet donc de formuler l'hypothèse d'une relation d'interdépendance réversible et asymétrique entre l'action et le contexte. Dans le cadre des recherches menées au Centre de Philosophie du Droit, cette hypothèse a été formulée à partir d'une critique épistémologique des modèles fonctionnalistes de l'intentionnalité en sciences cognitives d'une part et du formalisme de la communication chez P. Livet de l'autre. Toutefois, elle nous amène également à envisager le projet d'une reconstruction positive de l'opération de modélisation, qui prend en compte l'hypothèse de la réversibilité asymétrique. Or, il nous semble que les modèles du cognitivisme élargi dont nous sommes partis peuvent constituer un point de départ approprié pour un tel prolongement empirique du travail conceptuel en théorie de l'action, dans la mesure où ces modèles soumettent également le contexte des systèmes cognitifs à l'opération de modélisation. C'est par une confrontation de ces deux perspectives, la théorie de l'action d'une part et les sciences cognitives de l'autre, que nous pouvons alors envisager une recherche dont l'objet est de vérifier la fécondité de

24 MAESSCHALCK M., *Normes et contextes*, op. cit., p. 184.

25 *Ibid.*

26 *Ibid.*, p. 185.

l'hypothèse de la réversibilité asymétrique dans le cadre d'un champ expérimental précis, tel qu'il se définit à partir des modèles de l'intentionnalité en sciences cognitives.

A partir de cette mise en perspective de la recherche, nous pouvons indiquer les trois temps de la vérification de notre hypothèse. Tout d'abord, il nous faudra reconstruire les différentes étapes épistémologiques qui permettent un tel traitement de la contextualité dans les modèles de l'intentionnalité en sciences cognitives. Ce sera l'objet du premier chapitre de cet ouvrage. Ensuite, nous essaierons de tirer les conséquences de la reconstruction de l'opération de modélisation de l'intentionnalité au niveau de la méthodologie de recherche. Ce sera l'objet du deuxième chapitre. Finalement, afin d'éviter une reprise formaliste et non réflexive de la contextualisation, il faudra dépasser la position cognitiviste à l'égard des limitations de sa propre entreprise de modélisation et radicaliser notre questionnement sur les limitations. Ce sera l'objet du troisième chapitre. Toutefois, au préalable, il nous faut encore préciser la méthodologie de travail et les heuristiques de recherche sur lesquelles nous nous sommes appuyées. Ce sera l'objet de l'introduction, qui nous permettra de définir un vocabulaire de travail plus proche de la pratique de modélisation en sciences cognitives, avant d'entrer dans le vif du sujet.

Notre ouvrage, bien entendu, n'a pu aboutir que grâce au soutien de nombreuses personnes et institutions. Tout d'abord, j'adresse mes plus chaleureux remerciements à Marc Maesschalck, pour sa confiance, son attention constante et ses conseils. Son expertise de différents domaines de la philosophie et des sciences humaines et sociales a été décisive, tant pendant la thèse de doctorat dont est issu cet ouvrage, que pendant les recherches ultérieures qui ont permis de le finaliser. Ensuite, je tiens à remercier vivement Jacques Lenoble, pour son accueil généreux et les collaborations dans le cadre du Centre de Philosophie du Droit, Institut de recherche avancée de l'Université catholique de Louvain. Puis ma reconnaissance va aux collègues et amis du Centre de Philosophie du Droit, en particulier, à Raphaël Gély, qui m'a soutenu et aidé tout au long de la finalisation du texte, à Francisco Padilla Olivares, Vincent Sohet, Laurence Blésin et Valérie Kokoszka, pour leurs contributions multiples. Merci encore à Bernard Feltz, qui a bien voulu assumer la direction de la thèse, à Marc Crommelinck pour ses conseils judicieux et à Andy Clark et à William Bechtel pour leur accueil lors de mon séjour de recherche à la Washington University à St. Louis (MO). Qu'il me soit permis enfin de dédier cet ouvrage à mes enfants Samuel et Mathilde.

Introduction

Dans cet ouvrage nous proposons une reconstruction critique de la problématique du contexte en théorie de l'action, à partir des approches naturalistes de l'intentionnalité dans les sciences cognitives contemporaines. Dans la philosophie de l'esprit, surtout d'inspiration anglo-saxonne, la problématique de la naturalisation de l'intentionnalité est abordée dans le cadre d'une discussion sur les critères qui permettent de définir l'intentionnalité. Au point de départ de ces discussions, on trouve une certaine acception commune de l'intentionnalité, comme étant la propriété des états mentaux par laquelle ils sont dirigés vers des objets et doués d'un certain contenu¹. En même temps, malgré cette acception commune, la question de la définition précise de la visée intentionnelle et de la représentation du contenu reste controversée et fait l'objet de débats en philosophie de l'esprit. C'est à partir de ces débats que l'on peut également tenter de circonscrire les limitations de l'approche naturaliste de l'intentionnalité.

Toutefois, l'étude de la naturalisation de l'intentionnalité à partir de la philosophie de l'esprit, dont nous ne nions évidemment pas la pertinence, a le désavantage de situer le travail du philosophe à l'extérieur de la pratique scientifique et ainsi d'externaliser en quelque sorte la problématique de la naturalisation. Pour éviter une telle externalisation nous envisagerons dans cet ouvrage une autre voie, qui est celle d'une reconstruction critique de l'opération de modélisation de l'intentionnalité telle qu'elle s'élabore à l'intérieur de la perspective naturaliste. Un tel renversement de perspective nous permettra d'étudier les limitations à l'intérieur de la pratique scientifique, telles qu'on peut les mettre en évidence à partir de la modélisation de l'intentionnalité en sciences cognitives.

Ce renversement serait évidemment injustifié si dans les sciences cognitives on ne trouvait pas déjà d'emblée un souci pour les limitations de leur propre démarche. C'est pourquoi, au préalable, nous voulons confronter notre perspective aux discussions sur les limitations dans la philosophie de l'esprit. Ensuite, nous indiquerons de quelle façon ces limitations se retrouvent dans les sciences cognitives. Pour le premier point, nécessairement

1 Cf. ENGEL P., "La philosophie de l'esprit", in MEYER M. (ed.), *La philosophie anglo-saxonne*, Paris, 1994, pp. 529–564, p. 550.

approximatif, nous nous baserons sur un aperçu de la naturalisation de l'intentionnalité en philosophie de l'esprit dans l'article "Consciousness, Intentionality, and Pragmatism"² de Richard Rorty. Pour le deuxième point, qui introduit à proprement parler à notre problématique, nous partirons de l'article de Henri Atlan "Projet et signification"³.

1. La problématique des limitations de la naturalisation de l'intentionnalité en philosophie de l'esprit

Dans "Consciousness, Intentionality, and Pragmatism", Rorty indique deux axes majeurs selon lesquels se sont déroulées les discussions autour du concept d'intentionnalité depuis la publication en 1949 du livre pionnier de Gilbert Ryle, *The concept of mind*. Le premier axe de discussion mis en évidence par Rorty concerne la problématique de la conscience. Sur cet axe, on peut situer d'un côté des philosophes comme John Searle et Ernst Nagel qui défendent la nécessité de tenir compte du caractère irréductiblement subjectif de l'intentionnalité. De l'autre on peut situer des philosophes comme Daniel Dennett et Hilary Putnam qui argumentent en faveur d'une autonomie de l'étude de l'intentionnalité par rapport à la problématique de la conscience. Searle par exemple se fait le défenseur du premier point de vue quand il affirme :

"Personne n'a jamais fait l'analyse de ses douleurs ou de ses soucis les plus pressants pour en conclure ensuite que ça n'était là que les états d'une machine de Turing, ou qu'on pouvait définir entièrement ces états par leurs causes et leurs effets, ou que s'attribuer à soi-même de tels états revenait à adopter une certaine position [*stance*] à l'égard de soi-même"⁴.

D'après Searle, la conscience est une propriété intrinsèque de l'intentionnalité. Par conséquent, si l'on veut avoir accès aux contenus intentionnels qui sont visés dans l'acte intentionnel, on doit adopter le point de vue de celui qui fait l'expérience de la visée des contenus⁵, ou encore,

2 RORTY R., "Consciousness, Intentionality, and Pragmatism", in SCOTT M., CHRISTENSEN S. et TURNER D. (eds.), *Folk Psychology and the Philosophy of Mind*, New Jersey, 1993, pp. 388–404.

3 ATLAN H., "Projet et signification dans les réseaux d'automates : Le rôle de la sophistication", in JANICAUD D. (ed.), *L'intentionnalité en question. Entre phénoménologie et sciences cognitives*, Paris, 1995, pp. 261–288. Ce recueil édité par Dominique Janicaud reprend les articles d'un colloque sur la phénoménologie et les sciences cognitives qui a eu lieu à Nice en 1992.

4 SEARLE J., *L'intentionnalité. Essai de philosophie des états mentaux*, trad. par Claude Pichevin, Paris, 1985, p. 311.

5 Cf. *ibid.*, p. 311.

comme le dit Rorty, on doit adopter le point de vue de la “première personne”⁶ (*first-person view*).

Rorty oppose cette défense de la “première personne” aux arguments de Dennett et de Putnam qui réclament une autonomie de la problématique de l'intentionnalité par rapport à la problématique de la conscience. D'après ces auteurs, l'accès conscient aux contenus de signification n'est qu'une problématique secondaire par rapport à l'étude des significations dans leur contexte d'effectuation. Ainsi, suite aux travaux de Putnam dans les années 70, s'est développée une approche fonctionnaliste de l'intentionnalité qui défend le point de vue selon lequel “les états mentaux sont des états fonctionnels, qui relient les entrées et les sorties d'un système, de façon équivalente aux états fonctionnels d'un programme ordinateur”⁷. Dans cette perspective, “la question de savoir si les ordinateurs peuvent être conscients devient la question de savoir si les ordinateurs peuvent être programmés pour rendre compte de leurs propres états internes”⁸. La notion de conscience est alors réduite à celle d'une croyance qui porte sur le contenu de nos propres états intentionnels.

C'est dans le contexte d'une telle reconceptualisation de l'intentionnalité, comme croyance effective en un certain contenu de signification, que l'on peut situer le deuxième axe de discussion dégagé par Rorty, qui concerne la critique du fonctionnalisme. L'école fonctionnaliste identifie les croyances comme étant des représentations mentales internes, reliant une certaine entrée à une certaine sortie. Mais, le déplacement de la question de la conscience vers celle de l'intentionnalité comme croyance effective laisse encore ouverte la question de l'identification de ces croyances. En effet, une expérience de pensée célèbre de Putnam montre que l'identification des rôles fonctionnels internes ne détermine pas de façon univoque le contenu sémantique de la croyance en question. Dans ses travaux plus récents, Putnam en tire la conclusion que l'attribution d'une croyance à une personne doit également tenir compte des relations avec son environnement, au-delà de la spécification du rôle fonctionnel interne. Cette position, qui est bien rendue par la phrase “les significations ne sont pas dans la tête”⁹ (*meanings are not in the head*), est qualifiée par Rorty de holiste. D'après Rorty, on retrouve également de telles positions holistes chez des auteurs comme Dennett ou Donald Davidson. Pour Dennett, l'attribution d'une croyance à une personne est liée à l'adoption d'une perspective intentionnelle (*intentional*

6 RORTY R., *op. cit.*, p. 393 ; notre traduction.

7 *Ibid.*, p. 390.

8 *Ibid.*, p. 391.

9 *Ibid.*, p. 394.

stance), d'après laquelle on attribue à cette personne les croyances qu'elle devrait rationnellement avoir. Dans les travaux de Davidson, ce holisme est encore radicalisé, et l'attribution de significations intentionnelles dépend non seulement d'une certaine perspective intentionnelle, mais également du réseau des relations qui relie un état intentionnel aux états intentionnels d'autres personnes.

Cette reconstruction de l'histoire récente de la philosophie de l'esprit, autour de la double critique du concept de conscience et de l'individualisme méthodologique de la perspective fonctionnaliste, est ensuite utilisée par Rorty pour défendre une position qu'il qualifie de pragmatique et d'anti-essentialiste. Nous n'entrerons pas ici dans ce débat. Ce qui devrait nous intéresser davantage, c'est la double limitation de la modélisation de l'intentionnalité qui se dégage des discussions en philosophie de l'esprit. La première est la réduction, dans les sciences cognitives, de la conscience à une croyance effective qui porte sur le contenu de nos propres états intentionnels et la deuxième est l'abstraction du caractère holistique de l'attribution des significations, en spécifiant la signification dans les termes des propriétés fonctionnelles d'un système computationnel.

2. L'approche des limitations à partir de la problématique de la modélisation

En sciences cognitives, la problématique des limitations ne se pose pas de la même façon qu'en philosophie de l'esprit. En effet, ce qui est au centre de la démarche scientifique, c'est d'abord une certaine pratique de naturalisation de l'intentionnalité, et les limitations porteront en premier lieu sur les conditions de possibilité de cette naturalisation. Ensuite, tant que les méthodes utilisées à l'intérieur de la perspective naturaliste n'excluent pas la possibilité de la modélisation d'une certaine propriété de l'intentionnalité, celle-ci n'apparaîtra pas comme une limite intrinsèque de la démarche. Dans ce travail, nous proposons de partir de cette problématique des limitations de la modélisation.

Pour introduire à cette problématique, nous proposons de partir de l'article "Projet et signification" de Henri Atlan, dans lequel on trouve ces deux temps de notre réflexion. Dans cet article, Atlan développe une théorie naturaliste de l'intentionnalité qui vise à rendre compte de la problématique des significations et de la visée intentionnelle, tout en articulant cette théorie à une réflexion sur les limitations d'une approche scientifique de l'intentionnalité. Dans l'introduction de l'article, Atlan précise tout d'abord qu'en sciences cognitives l'intentionnalité est réduite à une forme particulière

de causalité. Elle sera modélisée, comme nous l'avons vu ci-dessus, dans les termes de propriétés fonctionnelles d'un système computationnel. On distingue deux propriétés de base des systèmes computationnels. D'une part, ils ont la capacité de réaliser un certain but (le programme) et, d'autre part, ils ont la capacité d'opérer sur certains contenus informationnels (les instructions et les données). Comme le remarque bien Atlan, ces deux propriétés ne sont pas des propriétés indépendantes, mais se déterminent de façon réciproque. En effet, les instructions et les données n'ont de signification que par rapport à la fonction qu'elles assurent dans le fonctionnement global de la machine¹⁰. On retrouve donc bien le caractère fonctionnel des significations, mis en évidence par l'école fonctionnaliste. D'autre part, et ici Atlan se distance du fonctionnalisme, la relation peut également fonctionner dans le sens inverse. Une machine intentionnelle doit pouvoir adapter sa partie but (le programme) si elle est confrontée à un nouveau champ de significations¹¹.

Après ces précisions conceptuelles, Atlan discute un premier niveau de limitation de la modélisation de l'intentionnalité en sciences cognitives, qui concerne le problème de la modélisation de la conscience. D'un côté, la conscience "créatrice de projet est reconnue comme une sorte de causalité efficiente particulière et, comme telle, un objet spécifique des sciences de l'homme"¹². En sciences cognitives, la conscience intentionnelle est donc réduite à une forme particulière de causalité et sera modélisée à l'intérieur des sciences cognitives comme une procédure computationnelle d'un certain type, c'est-à-dire comme un cas particulier de procédure intentionnelle. Plus précisément, Atlan cherche à construire un modèle de la conscience comme étant une procédure capable de "s'observer elle-même"¹³ et de créer "indéfiniment de la nouveauté"¹⁴. Toutefois, il souligne également que même si un modèle pouvait rendre compte de ces propriétés, ce modèle resterait un modèle de la conscience "au sens limité et réduit"¹⁵. En effet, la modélisation de la conscience proposée par Atlan se situe délibérément "dans le contexte matérialiste d'une biologie physico-chimique"¹⁶ et opère une réduction, en ce sens qu'on "met entre parenthèses les phénomènes de

10 Cf. ATLAN H., "Projet et signification", *op. cit.*, p. 261.

11 Cf. *ibid.*, p. 262.

12 *Ibid.*, p. 281.

13 *Ibid.*, p. 284.

14 *Ibid.*, p. 285.

15 *Ibid.*, p. 284.

16 *Ibid.*, p. 263.

conscience tels que nous en faisons par ailleurs l'expérience immédiate, ou encore l'expérience médiatisée par la méditation philosophique¹⁷.

Ce qui apparaît donc comme une limitation dans la perspective des sciences cognitives, ce n'est pas la possibilité de la modélisation de la conscience en tant que telle, mais c'est l'attitude naturaliste qui caractérise la méthode et qu'Atlan définit comme une mise entre parenthèse de l'expérience subjective. Le maintien de cette attitude naturaliste pour étudier la conscience est justifié chez Atlan par le refus de toute perspective qui poserait "l'intentionnalité au départ comme fait de conscience (ou de langage) fondateur"¹⁸, comme dans la "philosophie phénoménologique qui croyait pouvoir fournir des fondements absolus aux sciences de la nature à partir de l'expérience fondatrice du sujet transcendantal"¹⁹. Plutôt que de donner lieu à une donation de sens absolue, l'attitude phénoménologique est, d'après Atlan, marquée par une incomplétude symétrique à celle des sciences de la nature. Ni les sciences cognitives, ni la phénoménologie, ne peuvent retrouver entièrement ce qu'elles mettent entre parenthèses, respectivement l'expérience vécue et le monde des sciences de la nature. Les deux attitudes sont deux attitudes que l'on peut adopter par rapport à une même réalité, mais qui sont irréductibles l'une à l'autre. Cette réalité, c'est la conscience intentionnelle "créatrice de significations", envisagée d'un côté comme procédure effective et de l'autre comme expérience vécue.

Après avoir dégagé l'attitude naturaliste comme étant l'attitude qui caractérise la modélisation de l'intentionnalité en sciences cognitives, on peut envisager les conséquences de cette abstraction méthodologique, qui concernent les limites liées aux modèles proposés au sein de l'attitude naturaliste. Dans "Projet et signification", Atlan met en évidence deux limitations internes qui sont liées à la modélisation de l'intentionnalité comme procédure effective : d'une part une limitation liée à la clôture des systèmes intentionnels et de l'autre une limitation liée au problème de la finitude des programmes. Le modèle qui sert de base à la discussion de Atlan est le modèle formel des réseaux d'automates. De tels réseaux sont constitués d'unités élémentaires, les automates, qui par leur organisation en réseau peuvent donner lieu à l'émergence de propriétés d'ensemble. A partir de ce choix de formalisme, Atlan se pose ensuite la question de savoir si "une intentionnalité même limitée peut émerger dans un processus auto-organisateur comme propriété globale d'un réseau d'automates"²⁰.

17 *Ibid.*

18 *Ibid.*, p. 263.

19 *Ibid.*, p. 264.

20 *Ibid.*, p. 263.

Or, un tel projet rencontre des difficultés au moins à deux niveaux. D'abord, les significations qui peuvent émerger du fonctionnement interne du réseau ne peuvent être reconnues comme telles que par un observateur externe au système qui *interprète* ces significations en termes fonctionnels. Par exemple, un réseau qui classe des données peut réaliser une fonction de reconnaissance de formes, mais l'identification de cette fonction dépend d'un observateur externe au réseau. En introduisant la notion d'observateur, on ne doit pas pour autant sortir de l'attitude naturelle. En effet, comme le souligne Atlan à partir de l'exemple du réseau de reconnaissance des formes, l'attribution d'une fonction au réseau "est une projection de nos propres expériences cognitives de reconnaissance de formes [...] sur le comportement observé dans le réseau"²¹. Ensuite, pour éviter une régression à l'infini, il faut bien supposer quelque part l'existence d'un observateur qui est capable d'interpréter ses propres fonctions cognitives. Cet observateur sera précisément l'observateur qui possède la conscience intentionnelle, envisagée comme visée effective. Toutefois, on ne dispose pas encore d'une telle théorie de la conscience intentionnelle. Et, à cause de la complexité des réseaux, il se peut que l'on n'arrive jamais à fixer les significations de façon univoque. Ainsi, une machine qui possède réellement la conscience intentionnelle ne serait plus programmable, parce que, pour identifier une machine qui est capable de créer indéfiniment des significations nouvelles, il faudrait un programme infini²².

Suite aux réflexions d'Atlan on peut donc dégager, de façon provisoire, un premier niveau de limitation de la modélisation de l'intentionnalité en sciences cognitives. Il porte sur le traitement de la conscience intentionnelle en sciences cognitives comme "procédure effective" et renvoie à une abstraction méthodologique qui consiste à mettre entre parenthèses l'expérience subjective. Cette limitation concerne la complexité intrinsèque des modèles formels qui veulent rendre compte, à l'intérieur de l'attitude naturaliste, des propriétés de visée intentionnelle (orientation vers un but) et de représentation de contenus (création des significations)²³ des opérations intentionnelles.

21 *Ibid.*

22 Cf. *ibid.*, p. 287.

23 Dans un autre article sur les réseaux auto-organiseurs, Atlan discute une deuxième limitation interne liée à la complexité des réseaux. Il s'agit de ce qu'il appelle la sous-détermination des théories par les faits. La sous-détermination est engendrée par le fait que des réseaux de neurones différents, avec des histoires différentes, peuvent néanmoins évoluer vers un même état observable. On pourra donc avoir "un nombre considérable de théories [*neuronales*] qui prédisent les mêmes faits observables" (ATLAN H., "Self-organizing Networks : Weak, Strong and Intentional, the Role of their Underdetermination",

3. Aperçu de la structure de l'ouvrage

Le modèle formel des réseaux d'automates d'Atlan occupe une position assez marginale dans le champ des sciences cognitives²⁴. Si nous avons choisi ce formalisme comme porte d'entrée à notre problématique, c'est parce qu'il permet d'illustrer la problématique de l'origine des significations et des visées intentionnelles, telle qu'elle se pose en rapport à la modélisation de l'intentionnalité en sciences cognitives. Toutefois, le modèle d'Atlan, par son caractère formel, ne permet pas d'identifier une opération de modélisation spécifique de l'intentionnalité. Le formalisme des réseaux d'automates vise en effet à rendre compte de la dynamique de l'émergence des significations dans différents systèmes auto-organisés, comme les systèmes cellulaires élémentaires ou les systèmes immunitaires des organismes supérieurs²⁵.

Les modèles formels nous donnent une première idée de l'opérationnalisation de l'intentionnalité en sciences cognitives, mais il faut dépasser ce niveau formel si l'on veut obtenir des modèles plus spécifiques. Pour définir une telle opération spécifique de modélisation, nous procéderons en deux étapes. En un premier temps, nous parcourons l'histoire récente des sciences cognitives, à partir des fils conducteurs de l'opérationnalisation (la computation), de la modélisation des représentations (les contenus de significations) et de la modélisation des fonctions (les buts poursuivis). Comme point de départ de ce parcours — qui se veut plus conceptuel que historique — nous suivons les discussions autour des différents mécanismes computationnels proposés dans le cognitivisme classique et le connexionnisme. A partir de là, nous considérerons les élargissements conceptuels nécessaires pour dépasser une approche purement computationnelle ou opératoire de la cognition. Comme nous le verrons, pour pouvoir rendre compte des dimensions représentationnelles et fonctionnelles des opérations cognitives, il faut dépasser les frontières artificielles entre le système et l'environnement et s'intéresser à l'inscription de la cognition dans son contexte sémantique et fonctionnel. La théorie de la cognition qui émergera de ces élargissements s'inscrira dans le courant de l'inscription corporelle et contextuelle de la cognition (*embodied and embedded cognition*) et sera qualifiée dans cet ouvrage de théorie écologique de la cognition.

in *La Nuova Critica*, 19–20 (1992), pp. 51–70, p. 58).

24 Cf. VAN DE VIJVER G., “Self-organization and Naturalism : Perspectives and Deadlocks of the New Connectionism of Atlan and Varela”, in *The Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology*, 6 (1989), pp. 177–198, p. 182.

25 Cf. ATLAN H., “Projet et signification”, *op. cit.*, pp. 273–274.

En un deuxième temps, qui fera l'objet du deuxième chapitre, nous essayerons de tirer les conséquences méthodologiques d'une telle caractérisation de l'opération de modélisation de l'intentionnalité en sciences cognitives. Tout d'abord, nous chercherons à mettre en évidence l'originalité de la pratique de modélisation en sciences cognitives, dans son rapport aussi bien au réductionnisme biologique qu'au réductionnisme algorithmique de la cognition. En nous basant en particulier sur les travaux méthodologiques de D. Marr, nous étudierons de façon détaillée l'approche computationnelle qui caractérise de façon générale la modélisation en sciences cognitives et nous verrons en quoi cette modélisation se distingue d'un simple mécanisme de l'esprit. Cette caractérisation générale de la méthode de modélisation donne encore lieu à différents courants particuliers qui s'opposent dans le champ des sciences cognitives, que ce soit dans des paradigmes d'inspiration fonctionnaliste ou d'inspiration biologique. C'est ce qui nous amènera, dans un deuxième temps, à reprendre la question plus spécifique des conséquences méthodologiques de l'opération de modélisation de l'intentionnalité telle qu'elle aura été définie dans le cadre de la théorie écologique. Nous verrons que ni les méthodes d'inspiration fonctionnaliste, ni les méthodes d'inspiration biologique ne prennent en compte d'une façon adéquate la problématique du contexte représentationnel et fonctionnel au sein même de la modélisation.

Finalement, dans le troisième chapitre, nous reprendrons la problématique des limitations de la modélisation. Tout d'abord, nous serons amenés à déplacer la question des limitations telle qu'elle a été définie par Atlan. Celui-ci prend seulement en compte la condition de pertinence du formalisme ou sa limitation interne. En effet, la caractérisation de la limitation donnée par Atlan est une caractérisation générale qui est commune aux sciences cognitives et aux sciences de la nature. Elle concerne les conditions d'opérationnalisation de l'intentionnalité à travers la définition d'une méthodologie empirico-formelle et met en évidence la réduction naturaliste de la temporalité vécue de l'expérience subjective. A partir de là, Atlan prend uniquement en compte les limitations qui portent sur la complexité formelle des modèles de l'intentionnalité et les difficultés à rendre compte de cette complexité au sein de l'attitude naturaliste. De cette façon, Atlan ne prend pas en compte les limitations contextuelles de la modélisation. Celles-ci sont liées à la capacité sémantique des formalismes à investir des champs particuliers de modélisation, comme le champ des modèles de l'intentionnalité à la différence des modèles naturalistes en physique ou en biologie par exemple. Dans cette première partie du dernier chapitre, nous étudierons les conséquences de ce déplacement de la

problématique des limitations vers la problématique de la cohérence contextuelle de l'opération de modélisation. Nous verrons que ce déplacement nous amène à prendre en compte le rôle de l'interprétation contextuelle des modèles au sein même de l'entreprise de la naturalisation.

Toutefois, cette première approche de la limitation contextuelle se limite encore à un point de vue structurel et ne clarifie pas pour elle-même l'opération de contextualisation sur laquelle elle s'appuie pourtant. Elle appelle un second moment, qui fera l'objet du dernier paragraphe du troisième chapitre. Ce second moment tentera de montrer l'apport d'une radicalisation du questionnement sur les limitations. La radicalité de cette critique consistera à s'appuyer sur un concept réflexif de contextualisation, tel qu'il se dégage du dépassement phénoménologique de la position cognitiviste à l'égard des limitations contextuelles de sa propre modélisation de l'intentionnalité. Il nous faudra donc vérifier si et comment le modèle phénoménologique parvient à frayer un accès réflexif à la dynamique de la modélisation, caractérisée par les opérations d'élargissement, de modification, d'extension et, finalement, de stabilisation en rapport à un usage contextuel déterminé. De la sorte, nous parviendrons à dépasser les modélisations contemporaines de l'intentionnalité et à renouer au-delà des tentatives formelles inspirées de la systémique avec les conclusions de la phénoménologie husserlienne, en particulier celles des ouvrages *Expérience et jugement* (1928/1938) et *De la synthèse passive* (1918–1926).

Chapitre 1

Du computationnalisme aux modèles écologiques de la cognition

1. Introduction

Les sciences cognitives se présentent comme un regroupement de différentes disciplines dont le but commun est d'élaborer des modèles scientifiques du fonctionnement de la pensée. Leur émergence est intimement liée à l'avènement de l'ordinateur dans les années cinquante, par lequel s'est popularisée l'idée d'une conception de la pensée en termes de traitement d'information. L'esprit, en tant qu'être pensant, peut être conçu comme une machine qui encode des informations de l'environnement et qui ensuite, par des transformations successives de ces informations, extrait des connaissances générales sur le monde. Le concept central des sciences cognitives sera alors le concept de computation, concept qui généralise le concept mathématique de calcul et qui désigne tout processus causal spécifié par une suite systématique d'opérations élémentaires de traitement de données¹.

La conception de la pensée comme machine, comme pur automatisme obéissant à des lois causales, n'est évidemment pas neuve. Que l'on pense au modèle purement mécanique de la perception et de la mémoire — inspiré par la mécanique des fluides — de René Descartes² ou encore à la psychologie associationniste des sensations d'un des pionniers de la psychologie expérimentale, Wilhem Wundt — basée sur un parallélisme psychophysique³. S'y ajoute aujourd'hui, à côté de la mécanique ou de

1 Cf. CUMMINS R. et SCHWARZ G., "Connexionnisme, computation et cognition", in ANDLER D. (ed.), *Introduction aux sciences cognitives*, Paris, 1992, pp. 374–394.

2 L'extension du modèle mathématique de la mécanique corpusculaire à la physiologie et à la psychologie sous l'influence de René Descartes, spécialement dans le *Traité de l'homme*, aura une influence considérable à la fin du XVII^e et tout au long du XVIII^e siècle. Cf. MÉTRAUX A., "Die Mikrophysik der Wahrnehmung und des Gedächtnisses in der französischen Aufklärung", in FLOREY E. et BREIDBACH O. (eds.), *Das Gehirn, Organ der Seele ?*, Berlin, 1993, pp. 129–150.

3 Sur l'utilisation du schème de l'association en psychologie et sa naturalisation

l'association, le schème de la computation. La force de ce nouveau schème, c'est de pouvoir penser un troisième niveau de détermination causale, à côté du niveau de la détermination physique ou mécanique et de la détermination neurophysiologique ou associationniste, un niveau de détermination propre à la pensée.

Dans ce chapitre, nous partirons en un premier temps de l'étude de la cognition à partir du concept de computation. Ensuite, nous parcourrons les différents élargissements conceptuels nécessaires pour dépasser une modélisation qui se limite au seul aspect computationnel ou opératoire. De cette façon, nous serons amenés à distinguer trois étapes importantes dans l'évolution des modèles théoriques en sciences cognitives : les modèles cognitivistes classiques, les modèles connexionnistes et, finalement, les modèles écologiques de la cognition.

Tout d'abord, dans le premier paragraphe, nous donnons un bref aperçu des modèles cognitivistes classiques. Ces modèles s'inspirent d'une analogie avec l'ordinateur et définissent la computation dans les termes d'un raisonnement logique sur des objets représentés symboliquement. Comme en intelligence artificielle, le raisonnement se présente sous la forme de règles conditionnelles (enchaînements de "si ..., alors ...") et les objets manipulés sont représentés de façon explicite dans des bases de données.

Malgré la fécondité des modèles qui ont été élaborés à partir des hypothèses du cognitivisme classique, ces modèles éludent complètement la question de l'origine des contenus intentionnels. De plus, la question de l'émergence de nos capacités de traitement d'information à partir du cerveau est évacuée par une philosophie fonctionnaliste, selon laquelle la description fonctionnelle ou informationnelle de la cognition est indépendante de la description matérielle. C'est précisément le mérite d'une seconde génération de modèles, les modèles connexionnistes, d'avoir proposé un système de représentation et des mécanismes de calcul qui permettent de sortir de ces impasses. D'après ces modèles, qui feront l'objet du deuxième paragraphe de ce chapitre, les représentations sont encodées de façon diffuse dans l'activité d'un réseau de neurones et les calculs sont de simples associations par similarité. Ces processus associatifs permettent d'expliquer l'émergence des représentations dans des contextes d'apprentissage bien précis et la capacité d'utiliser ces représentations dans des contextes suffisamment semblables au contexte d'apprentissage. Dans ce deuxième paragraphe, nous verrons également comment on peut étendre les modèles connexionnistes pour

physiologique au XIX^e siècle, voir par exemple BORING E., *A History of Experimental Philosophy*, New York, 1929.

rendre compte de l'utilisation des représentations dans des contextes entièrement nouveaux.

Dans les modèles connexionnistes, le monde et le corps sont de simples sources d'entrée du système. L'apprentissage se fait sur fond d'un monde et d'un corps profondément passif, le seul élément actif étant la pensée. Cette hypothèse est partagée par les modèles cognitivistes classiques et les modèles connexionnistes. A partir d'une telle présupposition, on obtient seulement une spécification des mécanismes computationnels qui caractérisent l'intentionnalité — qu'il s'agisse de mécanismes de calcul, d'encodage de représentations ou d'apprentissage —, mais on ne rend pas compte de la dimension représentationnelle (création de significations) et fonctionnelle (le problème de la spécification de la fonction d'ensemble) des activités intentionnelles.

Dans le troisième paragraphe nous aborderons la problématique de la dimension représentationnelle et fonctionnelle. Nous verrons que les systèmes cognitifs interviennent activement dans la définition des entrées et des sorties — les contenus intentionnels —, selon leurs intérêts et leurs dispositions propres. D'autre part, le contexte fonctionnel extérieur au système joue également un rôle actif dans la résolution des problèmes. En particulier, l'environnement et le corps deviennent des partenaires actifs qui prennent une part de la réalisation des fonctions à leur compte, limitant ainsi la charge computationnelle des systèmes cognitifs. La théorie de la cognition qui résulte de ces développements sera qualifiée dans ce travail d'écologique. Elle considère que les représentations que construit le système cognitif sont relatives à un milieu de vie (*Umwelt*) particulier et ne peuvent être étudiées qu'en se référant aux relations que le système cognitif entretient avec les ressources naturelles et artificielles de ce milieu de vie. De cette façon on obtiendra, au terme de notre parcours, une théorie de la cognition qui permet de croiser l'opérationnalisation de l'intentionnalité avec une modélisation de la dimension représentationnelle et fonctionnelle.

2. Le modèle de l'ordinateur

L'émergence des sciences cognitives est intimement liée à l'avènement de l'ordinateur. Aussi bien l'idée d'un niveau de détermination causale propre à la pensée que les concepts de computation et de représentation ont reçu une première formulation en sciences cognitives à partir du développement de l'ordinateur. On peut caractériser ces premiers modèles de façon simplifiée à partir de trois hypothèses :

1. la pensée se rapporte au cerveau comme le logiciel (*software*) au matériel (*hardware*),
2. les computations sont des règles proches des langages formels de la logique,
3. les représentations sont encodées de façon explicite dans des bases de données⁴.

Ces trois hypothèses sont inspirées d'une analogie entre le cerveau et l'ordinateur⁵ et circonscrivent les modèles des sciences cognitives qu'on qualifie aujourd'hui de cognitivisme classique.

La première hypothèse, le rapport entre la pensée et le cerveau, définit un "fonctionnalisme au sens strict"⁶, selon lequel la description informationnelle ou fonctionnelle de la pensée est indépendante de la description matérielle. Dans le deuxième chapitre de ce travail, nous discuterons la question du rapport entre la pensée et le cerveau et montrerons en quoi le fonctionnalisme strict ne peut être maintenu. Les systèmes cognitifs sont des systèmes biologiques et la biologie impose des contraintes à l'organisation et au fonctionnement des systèmes cognitifs.

Dans ce paragraphe, en partant de la deuxième et de la troisième hypothèse, nous regarderons de plus près la définition de la cognition dans les termes de règles logiques et de représentations. Nous verrons qu'une telle conception de la cognition est à la base de nombreux programmes d'intelligence artificielle, comme les systèmes experts ou les programmes de conversation. Ensuite, nous indiquerons brièvement comment les hypothèses de base du cognitivisme classique se retrouvent également dans d'autres disciplines comme la psychologie cognitive et les neurosciences cognitives.

4 Cf. ANDLER D., "Calcul et représentation : les sources", in ANDLER D. (ed.), *Introduction aux sciences cognitives*, *op. cit.*, pp. 13–14.

5 L'ordinateur auquel se réfèrent ces premiers modèles a l'architecture devenue classique dans nos ordinateurs personnels, dite architecture de von Neumann. Cette architecture consiste précisément dans la séparation de la conception logique de la machine (*software*) et la conception de ses circuits (*hardware*), la première hypothèse du cognitivisme classique. D'un point de vue théorique toutefois, cette distinction est arbitraire. Toute opération exécutée par programme, c'est-à-dire implémentée en logiciel, peut aussi être implémentée par des circuits électriques, c'est-à-dire en matériel. Le poids du logiciel (du programme) dans nos ordinateurs personnels est si important, simplement parce qu'on a décidé de réaliser en matériel seulement quelques opérations logiques simples (cf. LOBELLE M., *Systèmes informatiques*, syllabus dactylographié, Faculté des Sciences Appliquées, Louvain-la-Neuve, 1992, p. 10). L'analogie avec l'ordinateur, sur laquelle se base le cognitivisme classique, est donc une analogie avec un certain type d'ordinateur, ayant l'architecture de von Neumann.

6 ANDLER D., *op. cit.*, p. 14.

Pour illustrer les hypothèses de base du cognitivisme classique, considérons d'abord un programme typique d'intelligence artificielle inspiré de ce courant. Il s'agit du programme SHRDLU de Terry Winograd — appelé aussi “le monde des blocs”. Ce programme est capable de mener une conversation concernant l'activité d'un robot (fictif) qui déplace des blocs visualisés sur un écran. En observant le dialogue entre une personne devant l'écran et l'ordinateur, on pourrait recueillir le fragment de conversation suivant (en écrivant les réponses de l'ordinateur en capitales et des explications sur le programme en italiques) :

- “Prends un gros cube rouge.
- OK [*le robot déplace quelques blocs pour pouvoir prendre le cube rouge*].
- Saisis la pyramide.
- JE NE COMPRENDS PAS DE QUELLE PYRAMIDE TU PARLES [*il y a trois pyramides dans le monde des blocs*].
- ...
- Que contient la boîte ?
- LA PYRAMIDE BLEUE ET LE CUBE BLEU.
- Qu'est-ce qui soutient la pyramide ?
- LA BOÎTE [*le programme suppose que la pyramide est celle dont on vient de parler*].
- ...”⁷.

Comme on peut s'en rendre compte à partir de ce morceau de dialogue, le programme SHRDLU est capable de raisonner sur les objets du “monde des blocs” et sur le langage utilisé pour se laisser instruire. Il utilise aussi des heuristiques en se basant sur le contexte ou les connaissances hypothétiques de l'interlocuteur.

L'objectif de ce genre de programme en intelligence artificielle est de construire des systèmes artificiels capables de comprendre le langage naturel. Un autre exemple bien connu (mais moins élaboré) de tels programmes est le programme ELIZA, qui simule un entretien avec un psychiatre non-directif. Ce programme répond par exemple à “Je suis ...” par “Depuis quand êtes-vous ... ?”. A “Je souhaite ...”, il répond “Que signifierait pour vous de ...” et à “Tout le monde ...”, il répond “Pensez-vous à quelqu'un en particulier ?”⁸. Ce programme est devenu célèbre parce que

7 WINOGRAD T. et FLORES F., *L'intelligence artificielle en question*, trad. de J.-L. Peytavin, Paris, 1989, p. 173.

8 Cf. *ibid.*, p. 188.

les personnes s'impliquaient émotionnellement dans l'entretien et semblaient parfois oublier qu'elles se trouvaient devant un écran d'ordinateur⁹.

Formellement, les programmes d'intelligence artificielle ont tous la même structure générale. La cognition s'y présente comme un raisonnement sur des objets (blocs, mots du langage) représentés symboliquement. Le monde réel, simulé ou représenté, est supposé exister en lui-même et être décomposable en unités élémentaires (les objets) ayant certaines propriétés et relations entre elles. Le raisonnement se présente sous la forme de règles conditionnelles (si telle et telle chose se présente, alors applique la règle r)¹⁰.

Un exemple extrême de cette conception est le projet de construire un énorme système expert, capable de reproduire la connaissance courante d'une personne normale. Ce projet, connu sous le nom CYC (pour *encyclopédie*), est entièrement basé sur l'encodage de données sous la forme de représentations symboliques et sur des raisonnements qui prennent la forme de règles de déduction explicites. Une information particulière est encodée dans des cadres (*frames*) comme le suivant¹¹ :

Jodoigne

Province : (Brabant Wallon)

Résidents : (Anne-Sophie, Louis-Michel)

Pays : (Belgique)

Le principe d'encodage est le même dans toute la base de données. Chaque cadre (*frame*) a des clefs (*slots*) — les trois sous-rubriques dans l'exemple donné ci-dessus —, chaque clef a comme valeur une liste d'entités et les clefs peuvent référer à d'autres clefs (par exemple, la clef *pays* peut référer à un nouveau cadre avec plus d'informations et ainsi de suite). Cet appareil de cadres, de clefs et d'entités est ensuite augmenté d'un langage plus puissant, permettant l'expression de relations logiques plus complexes, comme par exemple la règle : pour toutes les entités, si l'entité est un X alors elle a la propriété Y. L'idée de base du projet CYC est que, dès que le système expert aura dépassé un certain seuil de quantité d'informations stockées, il sera capable de découvrir par lui-même de nouvelles relations entre les entités et de répondre à des questions inédites d'une façon intelligente. Selon cette conception, l'intelligence est essentiellement une affaire de mémoire et de déductions logiques à partir des relations qui existent entre les entités mémorisées.

9 Cf. *ibid.*

10 Cf. VERLEYSSEN M. et BLAYO F., *Les réseaux de neurones artificiels*, Paris, 1996, pp. 19–20.

11 D'après un exemple donné par Andy Clark dans *Being There*. Cf. CLARK A., *Being There : Putting Brain, Body and World together again.*, Cambridge (MA)/London, 1997, p. 2.

A côté de leur utilisation dans la conception de systèmes artificiels, les hypothèses cognitivistes sont également à la base des modèles en psychologie cognitive. Ces modèles peuvent être caractérisés d'une façon générale par le concept de traitement d'information. D'après ces modèles, la psychologie cognitive étudie le traitement d'information dans le système cognitif humain, comme l'informatique étudie le traitement de l'information dans les ordinateurs¹². Un exemple de modèle en psychologie directement inspiré du cognitivisme classique est le modèle fonctionnel de la mémoire, basé sur trois catégories fonctionnelles distinctes : la mémoire déclarative, la mémoire procédurale et la mémoire de travail. La mémoire déclarative encode les représentations du monde (images, concepts, propriétés, relations, événements et modèles mentaux généraux), la mémoire procédurale encode les opérations mentales et les savoir-faire en termes de règles conditionnelles (si ..., alors ...) et la mémoire de travail encode de façon passagère les entités ou règles dont on se sert à l'instant¹³. Comme en intelligence artificielle, on a de nouveau un traitement séparé des contenus du système cognitif — encodés explicitement dans la mémoire déclarative ou procédurale — et des opérations de transformation de ces contenus — encodées dans la mémoire de travail.

Les neurosciences cognitives ont également été très influencées par les premiers modèles cognitivistes. Nous pouvons citer en particulier les études sur le cortex visuel. L'activité des neurones du cortex visuel a été interprétée comme effectuant la détection de traits des objets de l'environnement (leur orientation, leur vitesse de mouvement, etc.). Cette interprétation est directement liée à une conception de la cognition comme traitement logique d'informations encodées de l'environnement. L'encodage de l'information se ferait par la détection de traits et ensuite cette information serait traitée d'après les règles logiques explicites des niveaux supérieurs (classification, mise en mémoire et action)¹⁴.

Les premiers modèles cognitivistes, aussi bien en intelligence artificielle, en psychologie cognitive ou en neurosciences cognitives, sont aujourd'hui contestés par de nombreux modèles alternatifs qui mettent en question les règles utilisées (enchaînement de si ..., alors ...) et critiquent l'utilisation naïve d'un concept de représentation et d'un concept de mémorisation copiés sur le modèle de l'ordinateur. Dans les paragraphes suivants, nous confrontons les modèles classiques à ces modèles alternatifs.

12 Cf. HABERLANDT K., *Cognitive Psychology*, Cambridge (MA), 1994, p. 23.

13 Cf. *ibid.*, pp. 139–165.

14 Cf. VARELA F., THOMPSON E. et ROSCH E., *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, Paris, 1993, p. 79.

L'aperçu succinct des modèles classiques dans ce premier paragraphe avait seulement pour but de nous introduire au cheminement qui nous conduira, dans les paragraphes suivants, du calcul logique des modèles classiques à l'association habituelle des modèles connexionnistes et, ensuite, à l'inscription corporelle et contextuelle de la cognition.

3. De l'intelligence habituelle à l'intelligence productive

Les modèles cognitivistes classiques étudient une forme d'intelligence qu'on peut qualifier d'intelligence symbolique. Dans ces modèles, la cognition est définie dans les termes d'un raisonnement logique sur des objets représentés symboliquement. Ce type de modèles semble bien convenir, par exemple, pour l'étude de raisonnements qui utilisent des représentations verbales explicites ou des représentations mathématiques symboliques. Cependant, si l'on veut étendre le domaine d'étude des sciences cognitives au-delà de la pensée simplement logique, il faut revoir aussi bien les règles computationnelles proposées, que le système de représentation des données copié sur le modèle de l'ordinateur.

La nécessité de concevoir des hypothèses alternatives par rapport au cognitivisme classique apparaît de façon manifeste si l'on veut s'intéresser à la pensée qui intervient au niveau des comportements habituels. Ce qui semble essentiel dans ce type de comportements, c'est de pouvoir produire, de façon flexible, des réactions immédiates dans des contextes familiers, réactions qui ne font pas appel à des déductions explicites qui suivent les règles de la logique formelle. Cette opposition entre l'intelligence symbolique et l'intelligence habituelle peut être illustrée à partir d'un exemple donné par Hubert Dreyfus dans son analyse de l'acquisition des savoir-faire habituels¹⁵. Si l'on devait appliquer des règles explicites lorsqu'on roule en voiture — une règle pour la distance de sécurité à maintenir avec la voiture devant soi, en fonction de sa vitesse propre ; une règle pour ralentir quand il y a un vélo ; etc. — on se conduirait comme un débutant maladroit et notre comportement aurait très peu en commun avec un conducteur compétent. Plutôt que de décider et de calculer ce qu'il faut faire pour chaque situation, l'expert routier agit de façon habituelle. Il peut discriminer des types de situations, voir la ressemblance de la situation actuelle avec une certaine situation typique et y associer les comportements adéquats¹⁶.

15 Cf. DREYFUS H.L., "La portée philosophique du connexionnisme", in ANDLER D. (ed.), *Introduction aux sciences cognitives, op. cit.*, pp. 352–373.

16 Cf. *ibid.*, pp. 357–358.

Dans ce paragraphe, nous étudierons une série de modèles, les modèles connexionnistes, qui permettent de mieux rendre compte des formes d'intelligence habituelle. D'après ces modèles, les représentations sont encodées de façon distribuée dans l'activité d'un réseau de neurones et les calculs sont de simples associations par similarité. Ce qui est au centre des modèles connexionnistes, c'est la capacité d'apprentissage et de généralisation implémentée dans des réseaux de neurones. Nous verrons qu'on peut définir un algorithme d'apprentissage qui permet au réseau de mémoriser un grand nombre de données d'un échantillon d'apprentissage, en regroupant des données similaires dans une représentation prototypique approximative. Ensuite, des processus d'association par similarité permettent d'appliquer ces représentations prototypiques également dans d'autres contextes, si ces contextes sont suffisamment semblables au contexte de l'apprentissage. L'étude des processus d'apprentissage et de généralisation par similarité fera l'objet d'une première section de ce paragraphe.

Les modèles connexionnistes ne peuvent cependant pas rendre compte de la productivité de l'intelligence habituelle par rapport au contexte. Dans une deuxième section, nous étudierons alors les limites contextuelles des modèles connexionnistes. D'abord, nous nous intéresserons à la possibilité d'utiliser les représentations habituelles également dans des contextes entièrement nouveaux, c'est-à-dire à la possibilité de transformer une habitude en passant d'un contexte à l'autre, et, ensuite, nous examinerons la possibilité d'utiliser les représentations habituelles dans des situations où la réponse du système par rapport au contexte reste partiellement indéterminée. Finalement, la capacité de productivité de nos habitudes par rapport au contexte permet également d'envisager sous un nouvel angle la problématique de l'origine du contenu intentionnel et de sortir de l'impasse créée par les modèles cognitivistes classiques. En effet, la productivité des habitudes permet l'apparition d'une certaine nouveauté dans les réseaux de neurones. Toutefois, cette nouveauté n'est pas une nouveauté absolue. Comme nous le verrons, elle prend plutôt la forme d'une créativité ou d'une spontanéité située.

3.1. Les modèles connexionnistes : des mémoires associatives à l'auto-organisation

3.1.1. Les mémoires associatives

Les modèles connexionnistes introduisent un système de représentation et des mécanismes de calcul très différents par rapport au cognitivisme classique. Dans ces modèles, les représentations sont encodées de façon

diffuse dans l'activité d'un réseau de neurones et les calculs sont de simples associations par similarité. Quels pourraient être les avantages de ces modèles par rapport au cognitivisme classique pour étudier les fonctions cognitives ?

3.1.1.1. Un réseau de reconnaissance de visages

Considérons, par exemple, notre capacité de discrimination des sensations¹⁷. Par un ensemble limité de détecteurs spécifiques, on est capable de discriminer tout un éventail de couleurs, de goûts et d'odeurs. Cette richesse d'analyse des subtilités sensorielles est de loin supérieure aux descriptions verbales que l'on pourrait donner de ces sensations. Apparemment la discrimination et l'encodage des sensations ne se font pas d'après un système représentationnel symbolique accessible directement au langage naturel.

Malgré le fait que la puissance d'analyse sensorielle soit grande, le principe à la base de notre capacité de discrimination est simple. Dans le cas du goût par exemple, on dispose exactement de quatre types de détecteurs, que l'on peut appeler les détecteurs pour le sucré, le salé et l'amer¹⁸. Quand un aliment touche les détecteurs sur la langue, ceux-ci sont activés. Une pêche, par exemple, produira une réponse forte au niveau du détecteur du sucré, presque pas de réponse au niveau du détecteur du salé et de l'amer, et une faible activation du salé. Le point important à noter est que le résultat de l'activation n'est pas un mélange des différents goûts, mais un vecteur d'activité avec quatre composantes distinctes. C'est ce principe d'encodage, l'encodage vectoriel, qui permet de conférer une puissance de discrimination aussi grande à notre sensation de goût. Par exemple, si chaque composante peut discriminer dix niveaux distincts d'activité, le nombre total de goûts que l'on pourra discriminer s'élève à $10 \times 10 \times 10 \times 10 = 10.000$. Avec seulement quatre récepteurs différents de goût et dix niveaux d'activité pour chaque récepteur, on pourra discriminer 10.000 goûts différents. Le même principe est utilisé pour la discrimination des couleurs (à partir des trois types de cônes sur la rétine) et la discrimination des odeurs (on possède au moins 6 types distincts de récepteurs olfactifs).

La discrimination sensorielle n'est pas spécifique des humains. Un saumon peut retrouver la rivière où il est né, rien qu'à partir des odeurs, et ceci même s'il a vécu de nombreuses années dans l'océan et qu'il doit

17 Cf. CHURCHLAND P.M., *The Engine of Reason, the Seat of the Soul*, Cambridge (MA), 1995, pp. 21–27.

18 Cf. *ibid.*, p. 21.

remonter de nombreux branchements de la rivière. Les chiens et les souris également ont une capacité spectaculaire de discrimination des odeurs. Par contre, l'être humain excelle dans la discrimination des visages et leurs expressions émotionnelles. Un visage est une entité complexe, mais un visage familier sera reconnu à partir de différents angles en moins d'un quart de seconde¹⁹. Comment pourrait-on rendre compte de cette capacité de discrimination en suivant l'idée de la segmentation en traits ? Et quels seraient les traits du vecteur d'activité : la longueur du nez, la distance entre les yeux ?

Figure 1.1. Un réseau de neurones artificiels pour la reconnaissance des visages
(source : CHURCHLAND P.M., 1995, p. 40).

Remarquons d'abord que le cerveau humain ne semble pas posséder de cellules sensorielles qui détectent les traits d'un visage. Nous disposons seulement, pour encoder des visages, des cellules de la rétine qui sont sensibles à la couleur et à l'intensité lumineuse. Ces cellules se soucient peu de visages. Néanmoins, malgré cette absence de détecteurs, il y a une région du cerveau qui répond spécifiquement à des visages. Un dommage spécifique à cette région (dû à un accident vasculaire ou à un traumatisme crânien par exemple) cause l'étrange condition pour un patient de ne plus éprouver le moindre sentiment de familiarité à la vue de visages connus²⁰. Il

19 Cf. *ibid.*, p. 27.

20 Ce trouble psychique est appelé dans la littérature la prosopagnosie, du grec *prosôpon*

ne reconnaît plus les visages familiers et a perdu la capacité d'apprendre à reconnaître de nouveaux visages. Et ceci, malgré le fait que ce patient ne présente aucun défaut au niveau des yeux et qu'il peut reconnaître sans problème la plupart des objets qui ne sont pas des visages. Il y a donc une population de neurones spécifique, à un endroit précis du cerveau, qui est spécialisée dans la reconnaissance des visages.

Revenons maintenant à la discrimination des visages. Comment une population neuronale pourrait-elle réaliser cette discrimination ? Pour répondre à cette question, considérons un modèle de reconnaissance des visages, développé par le groupe de Garry Cottrell, sur base du principe de l'encodage vectoriel²¹. Ce modèle consiste en un énorme réseau de neurones, disposé en trois couches, tel qu'on l'a illustré par le dessin de la figure 1.1. La première couche, la rétine, représente les entrées du système dans un réseau carré de 64 fois 64 pixels dont chaque élément permet d'encoder 256 niveaux différents d'activation. Cette première couche à haute résolution permet l'encodage reconnaissable de photos de visages réels. Les éléments de la première couche projettent vers une couche intermédiaire, constituée de 80 éléments, permettant d'encoder les traits. Finalement, ces éléments projettent vers une couche de sortie, ayant 8 éléments. Les 3 premiers éléments de cette couche ont été ajustés pour discriminer respectivement, selon qu'ils sont activés ou non, entre les catégories visage ou non-visage, visage masculin ou non, visage féminin ou non. Les cinq éléments suivants sont ajustés pour attribuer le nom correct de la personne dont on voit le visage, sous la forme d'un code binaire conventionnel qu'on a assigné au nom.

Les performances de ce réseau sont remarquables. Après une période d'apprentissage, le réseau est capable de reconnaître sans faute les visages de l'échantillon d'apprentissage (cf. figure 1.2.) et leur genre, de reconnaître presque sans faute des photos des mêmes visages, mais photographiés à partir d'autres points de vue, et de donner, avec une réussite de quatre sur cinq, le genre de visages dont les photos ne faisaient pas partie de l'échantillon d'apprentissage. La capacité de discrimination du réseau permet donc non seulement de mémoriser des photos de visages, mais aussi de reconnaître des photos similaires des mêmes visages et de généraliser des propriétés de ces visages (ici masculin / féminin) à des photos de visages jamais vus.

(visage d'une personne) et *agnôsia* (absence de connaissance).

21 Cf. COTTRELL G., "Extracting Features from Faces Using Compression Networks", in TOURETSKY D. *et al.* (eds.), *Connectionist Models : Proceedings of the 1990 Summer School*, San Mateo (CA), 1991. Pour une discussion de ce modèle, cf. CHURCHLAND P.M., *op. cit.*, pp. 40-55.

Figure 1.2. Une sélection des images présentées au réseau lors de la phase d'apprentissage (source : CHURCHLAND P.M., 1995, p. 40).

3.1.1.2. *Le neurone binaire et l'apprentissage associatif*

La capacité de discrimination du réseau est contenue dans le principe de fonctionnement des neurones individuels du réseau artificiel. Chaque neurone artificiel fonctionne comme un dispositif à seuil (cf. la figure 1.3) : si la somme des entrées, modulées par la force ou le *poids* de leur connexion, dépasse un certain seuil, le neurone est activé (valeur 1). Sinon il reste en position de repos (valeur 0)²². Le neurone fonctionne donc comme un classificateur linéaire. Il partage l'espace des entrées en deux classes distinctes : l'ensemble des entrées qui donnent la valeur 0, si leur somme est plus petite que le seuil, et les entrées qui donnent la valeur 1, quand la

22 Ce neurone formel est celui introduit par McCulloch et Pitts dans une des premières descriptions du système nerveux comme machine logique (cf. McCULLOCH W.S. et PITTS W., "A Logical Calculus of the Ideas Immanent in Nervous Activity", in *Bulletin of Mathematical Biophysics*, 5 (1943), pp. 115–133). Il y a de nombreuses différences entre ce neurone formel (et d'autres utilisés par après) et les neurones réels (cf. CHURCHLAND P.M., *A Neurocomputational Perspective*, Cambridge (MA), 1989, p. 184). Néanmoins, la ressemblance est suffisamment forte pour qu'en principe les opérations exécutées par le neurone formel puissent l'être également par un neurone réel.

somme dépasse le seuil²³. Chaque neurone dans un réseau opère une telle classification. A partir de deux neurones on aura quatre classes, correspondant aux réponses (0,0), (1,0), (0,1) et (1,1) ; et ainsi de suite.

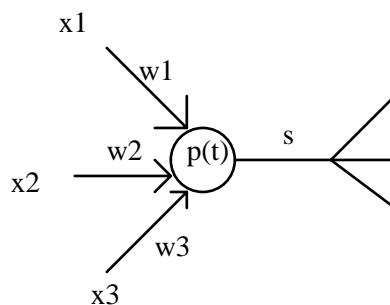


Figure 1.3. Exemple de neurone formel : x_1 à x_3 les entrées venant des autres neurones, w_1 à w_3 les poids de connexions, $p(t)$ le niveau d'activité du neurone et s la sortie. Si p dépasse un certain seuil le neurone s'active : $s = 1$, sinon il reste inactif : $s = 0$. Le niveau d'activité $p(t)$ est la somme des entrées multipliées par leurs poids respectifs : $p(t) : w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$ (figure de l'auteur).

Le réseau de neurones permet donc de regrouper des données en classes distinctes. Chaque couche du réseau de Cottrell produit un tel regroupement de données. De l'entrée à la couche intermédiaire, on classe d'abord les 4096 pixels de l'image d'entrée en 80 classes séparées. Ensuite, de la couche intermédiaire aux classes de la couche de sortie (les visages, les catégories associées).

A côté du regroupement des données en classes distinctes, on doit encore s'assurer que la construction des classes ne soit pas arbitraire. C'est le rôle des règles d'apprentissage. L'apprentissage dans les réseaux de neurones artificiels se fait selon une règle formulée déjà en 1949 par Donald Hebb²⁴. La règle de Hebb propose une stratégie d'apprentissage par association : si deux neurones dans un réseau sont activés fréquemment ensemble, alors le poids synaptique de leur connexion augmente, sinon il diminue. Dans le cas du modèle de Cottrell, on présente un échantillon de visages (cf. figure 1.2),

23 Certaines données présentées à l'entrée ne sont pas linéairement séparables. On peut remédier à ce problème en construisant une fonction seuil non-linéaire (au lieu de sommer les entrées, on peut introduire des termes non-linéaires pour atténuer l'effet des données trop éloignées de la moyenne par exemple). On peut montrer que, à partir d'une fonction de décision non-linéaire et un réseau à plusieurs couches, on peut approximer n'importe quelle fonction de séparation de classes (cf. VERLEYSSEN M. et BLAYO F., *op. cit.*, p. 66).

24 Cf. HEBB D., *The Organization of Behavior*, New York, 1949.

en fournissant en même temps la bonne réponse (visage ou simple objet, masculin ou féminin, nom associé au visage). Si la sortie du réseau est la solution désirée, ou suffisamment proche de cette solution, alors les poids synaptiques augmentent, sinon ils diminuent²⁵.

3.1.1.3. Des propriétés inattendues

Le modèle de Cottrell est bien capable de reproduire quelques propriétés connues de notre capacité de discrimination de visages. Parmi celles-ci retenons l'apprentissage de nouveaux visages, la capacité de classifier correctement des photos similaires d'un même visage et la possibilité de généraliser les propriétés des visages, comme le genre, à des données nouvelles. Ce modèle démontre, par l'exemple, la performance des réseaux de neurones dans la mémorisation de données dans des catégories et leur pouvoir de généralisation de ces catégories à des données inédites.

En plus de ces propriétés immédiates, le réseau artificiel de Cottrell a encore quelques propriétés inattendues. Une première propriété concerne les traits qui sont encodés par le réseau de la couche intermédiaire, le vecteur de traits. On pourrait s'attendre à ce que chaque élément de ce vecteur soit spécialisé dans un trait particulier du visage, comme la longueur du nez, la largeur de la bouche, la séparation des yeux, etc. Mais, en reconstruisant les stimuli d'activation spécifique de chacun des 80 éléments, on obtient un résultat très différent. En fait, chaque élément encode une structure ressemblant à un visage entier, mais seulement selon certains traits holistiques pour lesquels le langage ordinaire n'a pas de vocabulaire adéquat²⁶ (cf. figure 1.4). La discrimination de visages par le réseau fonctionne donc bien par segmentation en traits, comme dans le cas des sensations de couleur, d'odeur et de goût, mais chaque trait contient une information globale du stimulus d'entrée et pas une information partielle.

25 En toute généralité, le calcul des erreurs dans les réseaux de neurones se fait par une méthode statistique. D'abord, on calcule le carré des erreurs (ainsi on élimine les signes négatifs puisqu'on s'intéresse aux valeurs absolues de l'erreur) et ensuite on prend la moyenne sur les différentes composantes du vecteur de sortie. La modification des poids synaptiques se fait par une méthode de gradient. L'on isole une connexion particulière et on se demande si une modification du poids de cette connexion (en gardant les autres constants) augmenterait ou diminuerait l'erreur (cf. VERLEYSEN M. et BLAYO F., *op. cit.*, pp. 63–64). Dans le cas d'un réseau à différentes couches (comme c'est le cas ici), le calcul du gradient de l'erreur se fait de proche en proche dans le sens rétrograde des connexions. Pour cette raison, la méthode de correction utilisée dans le réseau de Cottrell est appelée méthode de la "rétropropagation du gradient" (cf. *ibid.*).

26 Cf. CHURCHLAND P.M., *The Engine of Reason, the Seat of the Soul*, *op. cit.*, p. 47.

Figure 1.4. Echantillon de réponse de la couche intermédiaire. Chaque réponse contient une information globale du stimulus d'entrée (source : CHURCHLAND P.M., 1995, p. 48).

Une deuxième propriété intéressante du réseau, c'est la façon dont les catégories globales, masculin/féminin et visage/non-visage, sont encodées par le réseau. Si l'on représente les valeurs possibles du vecteur de traits dans un espace, on constate que cet espace est divisé en deux sous-régions par un plan, une région donnant la réponse "masculin" et l'autre région la réponse "féminin"²⁷ (cf. figure 1.5). Pendant la phase d'apprentissage, les connexions du réseau se sont ajustées pour partager l'espace des réponses du vecteur de traits en deux classes linéairement séparables, une correspondant à la catégorie "masculin" et l'autre à la catégorie "féminin". Le point central de chaque sous-région est une réponse moyenne ou *prototypique* de la catégorie. Les données suffisamment proches de cette réponse prototypique seront classées dans la même catégorie. De la même façon, les visages à mémoriser sont encodés par une représentation prototypique de chaque visage.

D'autres réseaux ont été construits sur base des mêmes principes. Le réseau NETtalk, par exemple, est programmé pour lire correctement des mots anglais²⁸. Son mode de fonctionnement est tout à fait similaire au réseau de Cottrell. Pendant la phase d'apprentissage, on présente à l'entrée une série d'exemples de mots anglais et à la sortie la bonne prononciation de ces mots. Le réseau doit tenter d'associer à chaque mot les unités élémentaires de prononciation ou phonèmes correspondants. Après la période d'apprentissage, le réseau est capable de lire correctement les mots de l'échantillon d'apprentissage, ainsi que des mots semblables. Si on relie le

27 Cf. *ibid.*, p. 50.

28 Cf. SEJNOWSKI T. et ROSENBAUM C., "NETtalk : A parallel Network that Learns to Read Aloud", in *John Hopkins University Electrical Engineering and Computer Science Technical Report*, n° 86/01 (1986).

réseau à un synthétiseur de voix, qui traduit les phonèmes en sons acoustiques, on peut suivre le parcours d'apprentissage du réseau : le balbutiement initial, puis des mots prononcés partiellement et finalement une relativement bonne simulation de la prononciation normale.

Figure 1.5. L'encodage des catégories dans les réseaux de neurones. Le cube représente de façon schématique l'ensemble des valeurs possibles du vecteur de traits de la couche intermédiaire (ici dans un espace à 3 dimensions, en réalité 80 dimensions), les sous-régions représentent les données d'une même catégorie (source : CHURCHLAND P.M., 1995, p. 48).

A aucun moment, le programmeur de NETtalk n'a encodé de règles explicites de discrimination ou une structure de représentation pour les phonèmes. Tout ce qu'il a fait, c'est présenter des exemples. Après l'apprentissage, le fonctionnement du réseau est tout à fait similaire au réseau de Cottrell. D'abord, le réseau produit une discrimination des entrées dans un vecteur de traits (la couche intermédiaire). Ensuite, il construit une représentation des phonèmes (l'équivalent ici des catégories visuelles) en partageant l'espace de réponse du vecteur de traits en parties disjointes. Aussi bien la discrimination de l'entrée en traits, que la partition de l'espace de réponse du vecteur de traits en catégories, sont encodées de façon distribuée dans les poids des connexions synaptiques. On parle aussi d'un traitement parallèle et distribué de l'information (*parallel distributed processing*).

Il est intéressant de comparer ce réseau à un autre programme de prononciation, commercialisé sous le nom de DECTalk²⁹. Il s'agit d'un programme classique d'intelligence artificielle, qui encode de façon explicite des règles de conversion de texte en phonèmes et une liste d'exceptions à ces règles. Relié à un synthétiseur de voix, il est également capable de simuler la prononciation de l'anglais. Le point fort de NETtalk par rapport à ce programme, du point de vue des sciences cognitives, est sa capacité d'apprentissage. Dans DECTalk on doit encoder de façon explicite des règles formelles de conversion de texte en phonèmes, tandis que dans NETtalk cette conversion émerge de l'activité distribuée d'un réseau de neurones. NETtalk présente encore un avantage supplémentaire par rapport aux modèles classiques d'intelligence artificielle. Le programme NETtalk est moins fragile à l'endommagement. En effet, par le caractère distribué de la représentation des phonèmes à travers le réseau, une erreur locale dans les poids synaptiques n'entraînera pas d'erreurs énormes à la sortie. Tandis que, dans le cas de DECTalk, une erreur au niveau des règles ou de la base de données peut rendre le système complètement inopérant. En plus, il est possible de présenter une donnée incomplète au système NETtalk sans causer de grands problèmes. Le réseau classifiera cette donnée dans la catégorie correspondant au vecteur de traits le plus proche du vecteur de la donnée incomplète. Ces deux dernières propriétés, la fiabilité par delà l'endommagement et la possibilité de traiter des données incomplètes, ne sont pas spécifiques du programme NETtalk, mais des propriétés tout à fait générales du modèle des réseaux de neurones. Elles sont liées à l'encodage massivement distribué des représentations.

Le programme de reconnaissance des visages de Cottrell et le programme de prononciation de l'anglais NETtalk illustrent la performance des principes d'encodage et de calcul des réseaux de neurones. Dans le modèle des réseaux de neurones, les représentations (traits, catégories) sont encodées de façon distribuée dans l'activité d'un réseau et les calculs (la classification, la généralisation) sont de simples associations par similarité. A partir de ces principes simples, on peut implémenter de véritables mémoires associatives, capables de classer une série de données dans des catégories mémorisées à partir d'une série d'exemples et de généraliser cette classification à des données nouvelles. De plus, par le caractère distribué de l'encodage, les réseaux de neurones présentent une fiabilité par delà l'endommagement et la possibilité de traiter des données incomplètes.

29 Produit commercial du Digital Equipment Corporation (DEC). Cf. CLARK A., *Being There*, *op. cit.*, p. 55.

3.1.2. L'apprentissage par auto-organisation

Par leur capacité d'apprentissage et par l'encodage distribué des représentations (catégories, traits), les réseaux de neurones semblent déjà se rapprocher davantage de l'intelligence habituelle. Un problème subsiste cependant. La méthode d'apprentissage utilisée dans NETtalk et dans le modèle de Cottrell nécessite un professeur. En effet, pendant la période d'apprentissage, il y a toujours une série d'exemples à présenter dont on connaît à l'avance la réponse. Le réseau mémorise cette série d'exemples dans ses connexions et, après la mémorisation, sait associer une réponse à des données nouvelles par similarité avec les exemples de l'apprentissage.

C'est pour éviter ce recours à un professeur que certains chercheurs ont développé des réseaux capables d'auto-organisation. Un tel réseau peut produire de lui-même une sortie structurée, sans que quelqu'un ait donné une instruction quant à la structure à produire. Le réseau de Kohonen, par exemple, peut produire une représentation simplifiée d'une forme géométrique présentée à l'entrée, en respectant la topologie de cette forme (des points proches sur la forme géométrique seront également proches sur la représentation)³⁰. En présentant un carré par exemple, un réseau avec 100 noeuds produira une grille carrée avec 100 mailles. La seule règle d'apprentissage programmée est une règle locale, imposant une relation de proximité entre les vecteurs position des points individuels présentés à l'entrée et les vecteurs poids du réseau (de même dimension que le vecteur d'entrée, par construction du réseau de Kohonen). Après plusieurs itérations — en partant d'une distribution aléatoire — les valeurs des vecteurs de poids prendront la valeur des vecteurs au centre des mailles d'une grille carrée (cf. la figure 1.6). On ne sait rien à l'avance de la configuration de sortie globale qui émergera.

Le réseau de Kohonen peut produire une structure topologique simplifiée d'un carré, d'un triangle, d'un cercle ou même d'une figure tridimensionnelle. Chaque fois, la règle est une règle locale de similarité entre un vecteur du réseau et un vecteur position d'un point de la forme géométrique présentée à l'entrée. La programmation du réseau de Kohonen procède donc en quelque sorte par le mécanisme inverse de celui des mémoires associatives. On donne des instructions locales, mais on ne sait rien du résultat global après l'apprentissage. Dans le cas des mémoires associatives, on présente des instructions globales sur la configuration de sortie, mais on laisse les règles locales de transformation complètement

30 Cf. KOHONEN T., "Self-organized Formation of Topologically Correct Feature Maps", in *Biological Cybernetics*, 43 (1982), pp. 59–69.

indéterminées. On est donc en présence d'une forme d'auto-organisation au sens de l'émergence d'un ordre global à partir de stimulations et de modifications locales. De plus, en produisant des structures topologiques simplifiées à partir de formes parfois complexes, le modèle de Kohonen permet de percevoir des rapports géométriques difficilement appréhendables sur les formes de départ.

Figure 1.6. L'adaptation itérative des poids du réseau de Kohonen, à partir d'une distribution uniforme carrée (source : VERLEYSEN M. et BLAYO F., 1996, p. 98).

D'autres règles locales peuvent donner lieu à des réseaux auto-organisateurs. Le réseau de Héroult-Jutten par exemple tente de trouver une solution au problème bien connu du *cocktail party*³¹. Comment retrouver les sources indépendantes de deux signaux dont on ne recueille que des mélanges ? Pour résoudre ce problème, le modèle de Héroult-Jutten propose la règle locale suivante : pour chaque entrée, on modifie les deux vecteurs de sortie de façon à maximiser un critère d'indépendance entre les deux vecteurs. Après plusieurs itérations, les vecteurs de sortie correspondront aux

31 Cf. HÉRAULT J., JUTTEN C. et ANS B., "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé", in *Actes du Xe Colloque Gretsi*, Nice, 1985, pp. 1017–1022.

signaux des deux sources indépendantes. Rien qu'à partir de stimulations et de modifications locales, le réseau sait donc produire une sortie structurée globale, significative pour un observateur.

Il est intéressant de comparer le principe d'apprentissage par auto-organisation et l'apprentissage supervisé des mémoires associatives. Nous avons vu que le réseau de Kohonen produit une réduction de données : à partir d'une distribution de points représentant une figure géométrique il produit une structure géométrique plus simple.

Dans le cas du modèle de Cottrell, programmé pour la reconnaissance des visages, on constate un phénomène similaire. Chaque couche du réseau produit un regroupement des données. De l'entrée aux vecteur de traits, on classe d'abord les 4096 pixels de l'image d'entrée en 80 classes séparées. Ensuite, du vecteur de traits aux catégories, on regroupe les combinaisons possibles des traits autour de visages connus (le cas de la discrimination) ou autour de quelques catégories prototypiques (visage masculin prototypique, visage féminin prototypique).

Aussi bien le réseau de Cottrell que le réseau de Kohonen permettent de regrouper les données dans des classes statistiques. La différence entre la méthode d'apprentissage supervisé et l'apprentissage non-supervisé (l'auto-organisation) ne se situe pas au niveau du principe de classification, mais au niveau de l'optimisation des paramètres. En faisant varier les poids des connexions, les frontières entre les classes se déplacent. Par l'optimisation on fixe les frontières entre les classes de façon à approcher un critère de classification fixé à l'avance. Ce critère diffère pour les deux méthodes d'apprentissage. Dans le cas de l'apprentissage supervisé, on présente une série d'exemples qui doivent au moins être groupés dans la bonne classe. Tandis que dans le cas des réseaux basés sur l'auto-organisation, on doit simplement maximiser un critère de similarité topologique ou un critère d'indépendance entre classes. Dans ce cas, la structure des classes n'est pas connue d'avance.

Les méthodes d'apprentissage des réseaux de neurones peuvent se retraduire dans le langage mathématique de l'approximation des fonctions. Par l'algorithme d'apprentissage les réseaux réalisent, de façon itérative, une approximation d'une fonction de classification de données. Evidemment, l'intérêt de ces réseaux n'est pas simplement mathématique. Les méthodes d'approximation étaient déjà connues longtemps avant le développement des modèles de réseaux de neurones. Du point de vue des sciences cognitives, les réseaux de neurones ont avant tout un intérêt conceptuel. Ils présentent un exemple de structure purement matérielle pouvant réaliser le calcul d'une fonction de classification de données. De cette façon la classification et

l'optimisation ne sont plus le privilège des méthodes de calcul sophistiquées des mathématiciens, mais peuvent également s'inscrire potentiellement dans la matière.

3.1.3. L'intelligence habituelle

Les schèmes de l'apprentissage associatif et de la catégorisation autour de représentations prototypiques permettent de modéliser une forme d'intelligence qu'on peut qualifier, pour la distinguer de l'intelligence symbolique des modèles cognitivistes classiques, d'intelligence habituelle. Dans ces modèles, une réaction appropriée apprise dans un certain contexte pourra être reproduite de façon automatique dans des circonstances semblables. Remarquons que le caractère automatique ne préjuge pas du caractère conscient ou non des réactions. La seule chose qui importe, c'est que les réactions habituelles soient adaptées à la fonction représentationnelle qu'elles doivent réaliser. Certains processus sont complètement inconscients, comme la discrimination des visages en traits holistiques ou en prototypes et d'autres sont accessibles à la conscience, comme les données qualitatives de nos perceptions de visages : le teint, les contours, etc.³²

Une autre caractéristique de l'intelligence habituelle des modèles connexionnistes, c'est l'association directe de représentations globales aux stimuli. Dans ces modèles, on ne décompose pas d'abord le visage en fragments, que l'on recompose ensuite pour représenter le visage d'un tel ou d'un tel. On perçoit d'emblée le visage dans sa totalité, avec sa forme, un certain genre associé et des expressions émotionnelles par exemple. Au cours de l'apprentissage, le réseau construit à chaque étape de telles configurations globales qui évoluent progressivement vers une configuration finale, qui est la configuration d'équilibre qui déterminera les réactions

32 Le fait que certains processus de traitement d'information sont accessibles à la conscience attentionnelle et d'autres pas se comprend mieux si on réalise que le traitement de l'information et la reprise consciente de certaines informations sont deux fonctions psychologiques distinctes. On peut illustrer ceci grâce à l'exemple du système visuel. Selon Beckermann par exemple, on peut distinguer deux fonctions adaptatives différentes au niveau du système visuel. Le traitement de l'information sert à informer l'organisme de ce qui se passe dans son environnement et ne doit pas toujours être accessible à la conscience, tandis que certaines représentations intermédiaires, comme les données qualitatives de l'environnement (lumière, couleur, texture, etc...), sont toujours accessibles parce qu'elles servent à repérer les effets que l'environnement a sur le corps (ici les yeux en occurrence), effets que l'on peut influencer par des mouvements volontaires. Cf. BECKERMANN A., "Visuelle Informationsverarbeitung und phänomenales Bewußtsein", in METZINGER TH. (ed.), *Bewußtsein. Beiträge aus der Gegenwartphilosophie*, Paderborn, 1996, pp. 663–684.

habituelles futures. Par cette propriété, les modèles connexionnistes se rapprochent des théories de la *Gestalt*. Dans la *Gestalt* aussi, toute perception est un tout organisé que l'on perçoit directement comme tout³³. Aussi bien dans les modèles connexionnistes que dans les théories de la *Gestalt*, on a donc l'association automatique d'une certaine configuration ou forme globale à des données physiques au départ dissociées.

Les exemples de fonctionnement de réseaux de neurones que nous avons considérés dans ce paragraphe concernaient tous des fonctions perceptives, que ce soit l'association de perceptions visuelles ou auditives à des perceptions visuelles (des noms à des visages, des phonèmes à des mots) ou l'association de représentations visuelles simplifiées à des formes géométriques. Les modèles connexionnistes ne s'appliquent cependant pas uniquement aux fonctions perceptives. Les savoir-faire et les compétences pratiques prennent également appui sur des mécanismes d'association habituelle et peuvent donc se modéliser à partir des réseaux de neurones. Dans ces cas, les représentations sont des représentations de possibilités d'action relatives à certains contextes. En reprenant le vocabulaire gestaltiste, on peut parler de l'association d'un schéma corporel à une certaine situation³⁴.

A côté des représentations de notre environnement physique (la perception) et des représentations des possibilités d'action du corps (le schéma corporel), il y a encore une troisième catégorie de représentations où l'intelligence habituelle joue un rôle important : les représentations sociales. Les rôles sociaux par exemple, comme le rôle de parent ou d'étudiant, regroupent un ensemble de comportements habituels que l'on adopte dans des contextes appropriés. Souvent il s'agit de règles de comportement que l'on respecte sans pouvoir en donner une formulation explicite. Ils opèrent sur le sujet à travers l'inscription corporelle des habitudes corporelles adéquates³⁵. Ici aussi, on a l'association automatique de représentations, des représentations sociales, à certains types de situations.

33 Voir par exemple ROCK I., *The Logic of Perception*, Cambridge (MA)/London, 1983, pp. 32–36.

34 Pour l'application des modèles connexionnistes aux savoir-faire, voir par exemple DREYFYUS H. L., "La portée philosophique du connexionnisme", *op. cit.*

35 Cf. BOURDIEU P., *Le sens pratique*, Paris, 1980, p. 22. Cité dans MAESSCHALCK M., "Une éthique des styles de vie ? Questions au schématisme de Bourdieu", in *Science et Esprit*, 50 (1998), p. 162. On peut dire que les représentations sociales habituelles sont l'équivalent du niveau de *l'habitus* de Bourdieu, à ceci près que le schème de la grammaire générationnelle utilisé par Bourdieu pour définir *l'habitus* comme représentation est ici remplacé par le schème de la catégorisation connexionniste.

Nous n'approfondirons pas ici la possibilité d'appliquer les schèmes de catégorisation prototypique et d'apprentissage associatif à différents types d'intelligence habituelle. Dans la suite de ce paragraphe, nous voulons plutôt mettre en évidence un présupposé commun des modèles connexionnistes de l'intelligence habituelle, qui est la limitation des réactions habituelles à des contextes très spécifiques d'apprentissage. A partir des modèles connexionnistes, on ne peut pas rendre compte du caractère productif des représentations, c'est-à-dire de la possibilité de les utiliser en-dehors de leur contexte d'émergence. De plus, dans les modèles connexionnistes, les représentations sont complètement déterminées, ou saturées, par les contextes habituels. Ces modèles ne permettent pas de rendre compte de la spontanéité ou de l'improvisation souvent nécessaires dans des contextes pourtant habituels. L'intelligence habituelle (perceptive, corporelle, sociale) est productive et active en situation et pas simplement réaction cognitive automatique.

3.2. *L'extension des modèles connexionnistes*

Dans les modèles connexionnistes, les associations automatiques remplacent l'encodage explicite des représentations dans un langage symbolique. De cette façon, ces modèles étendent les modèles computationnels classiques à des formes d'intelligence qui sont en deçà de l'expression langagière et qui prennent la forme d'associations habituelles. Ces associations sont *intelligentes* en ce qu'elles n'associent pas simplement des réactions particulières à des stimuli particuliers de l'environnement, mais qu'elles associent de façon systématique une même représentation générale à toute une classe de stimuli de l'environnement ou tout une classe de situations d'action³⁶.

Cependant, que deviennent ces associations dans des environnements ou des situations qui sortent de nos contextes habituels ? Quelqu'un qui ne sait que faire quand il sort de ses habitudes ne sera sûrement pas qualifié d'intelligent. Est-ce que la notion d'intelligence ne suppose pas en plus une forme de productivité par rapport aux situations, au lieu de la simple

36 Il ne faut donc pas confondre ce type d'associationnisme computationnel et l'associationnisme physiologique du behaviorisme ou de la psychologie de l'introspection (basée sur un parallélisme psycho-physiologique (cf. WUNDT W., *Vorlesungen über die Mensch und Tierseele*, Hamburg/Leipzig, 1906, p. 253). Ici, les neurones sont des dispositifs à seuil et pas simplement des entités bio-chimiques. C'est précisément grâce aux propriétés de classification de ces dispositifs que les réseaux de neurones sont capables de discrimination et de catégorisation, c'est-à-dire capables de construire des représentations générales de leur environnement opératoire. Cette perspective est complètement absente des théories associationnistes non-computationnelles.

adaptation habituelle ? L'approfondissement du thème de la productivité du comportement intelligent montrera les limites des modèles connexionnistes. Tout contexte nouveau peut évidemment être pour le réseau connexionniste un contexte de départ pour un nouvel apprentissage, mais cela indique précisément qu'il ne sera pas reconnu d'emblée *comme* contexte. Il sera simplement un amalgame de données physiques disparates, reliées au hasard dans une configuration d'ensemble.

Pour illustrer le caractère productif de l'intelligence habituelle, nous reprendrons dans un premier temps les analyses de Maurice Merleau-Ponty dans la *Phénoménologie de la perception* concernant l'intelligence corporelle spatiale³⁷. L'intérêt de reprendre cette analyse réside dans le fait qu'il existe une grande ressemblance entre les limitations contextuelles des modèles connexionnistes et la problématique qui préoccupe Merleau-Ponty, comme l'a montré de façon convaincante David Hilditch dans ces travaux sur l'intelligence corporelle³⁸. Néanmoins, nous nous distancierons également quelque peu des pistes ouvertes par les travaux de David Hilditch, en indiquant une différence importante entre l'approche de Merleau-Ponty et l'opérationnalisation de la pensée dans les sciences cognitives. Dans une deuxième section, nous discuterons ensuite différentes pistes d'extension des modèles connexionnistes pour rendre compte de la productivité de l'intelligence habituelle. Cette extension ne nous amènera cependant pas à abandonner les mécanismes d'apprentissage des réseaux de neurones que nous avons étudiés. On peut plutôt considérer les mécanismes d'apprentissage des réseaux de neurones comme une première étape de construction de représentations dans des contextes bien définis. Les représentations habituelles simples pourront ensuite être transformées pour pouvoir faire face à des situations plus complexes.

3.2.1. Le corps simplement habituel : le cas Schneider

Merleau-Ponty, dans la *Phénoménologie de la perception*, a analysé en profondeur la productivité de l'intelligence habituelle. Considérons le corps en mouvement, par exemple. Ce corps n'est pas seulement corps habituel, mais aussi corps actuel, reprise active des représentations corporelles habituelles dans chaque nouvel espace : "On voit mieux, en considérant le corps en mouvement, comment il habite l'espace (et d'ailleurs le temps),

37 Cf. MERLEAU-PONTY M., *Phénoménologie de la perception*, Paris, 1945, pp. 114–172.

38 Cf. HILDITCH D., *At the Heart of the World : Merleau-Ponty and the Existential Phenomenology of Embodied and Embedded Intelligence in Everyday Coping*, Diss., St. Louis (MO), 1995, pp. 140–141.

parce que le mouvement ne se contente pas de subir l'espace et le temps, il les assume activement, il les reprend dans leur signification originelle qui s'efface dans la banalité des situations acquises"³⁹. L'espace représenté par le corps n'est ni un espace purement physique, ni l'espace abstrait de la géométrie. La représentation de l'espace prend la forme de possibilités générales de mouvement du corps (le schéma corporel), possibilités qui se rapportent au contexte de mouvement actuel, mais aussi à des contextes possibles. Chez Merleau-Ponty, c'est ce rapport à des contextes possibles, au-delà des contextes actuels, qui indiquera la différence entre une représentation spatiale purement habituelle (dans tel contexte j'ai telles possibilités de mouvement) et une représentation spatiale productive et dynamique : "Les psychologues disent souvent que le schéma corporel est dynamique. Ramené à un sens précis, ce terme veut dire que mon corps m'apparaît comme posture en vue d'une certaine tâche *actuelle* ou *possible*"⁴⁰.

Pour mieux mettre en évidence la productivité et l'activité du corps, Merleau-Ponty contraste le mouvement normal avec le cas d'un patient cérébro-lésé, Schneider, qui est capable d'exécuter des mouvements appropriés dans des contextes précis, mais qui n'est plus capable d'assumer activement ces mouvements⁴¹. En quelque sorte, le corps de ce patient est devenu entièrement habituel, capable seulement de produire des réactions automatiques⁴². La maladie de Schneider est causée par une lésion cérébrale dans la région occipitale du cerveau (la partie postérieure). D'après Merleau-Ponty, cette lésion a entraîné la perte de la capacité de produire des mouvements spontanés, actifs, mais a laissé d'autres fonctions intactes⁴³. Par exemple, Schneider est capable de mouvements concrets, dans des contextes

39 *Ibid.*, p. 119.

40 *Ibid.*, p. 116 ; nos italiques.

41 Cf. *ibid.*

42 Cf. HILDITCH D., *op. cit.*, pp. 141–151.

43 Le cas Schneider a suscité beaucoup de commentaires dans la littérature. Pour notre exposé nous pouvons suivre l'analyse de Merleau-Ponty. Cette analyse nous permet de préciser les limites contextuelles des modèles connexionnistes en adaptant — moyennant quelques modifications — le dialogue qu'entretient Merleau-Ponty avec la psychologie de la Gestalt (analogue aux modèles connexionnistes). D'un point de vue contemporain toutefois, il semble que le déficit de Schneider est beaucoup plus spécifique que ne le laisse comprendre Merleau-Ponty. Antonio Damasio, par exemple, suggère que le cas Schneider est analogue au fameux cas de Phineas Gage et qu'il s'agit d'une atteinte spécifique au centre d'intégration de l'émotion et de la raison. En fait, à l'époque de Merleau-Ponty, on surestimait le caractère holistique des fonctions corticales et on ne pouvait pas s'imaginer qu'une lésion cérébrale pouvait causer un déficit aussi spécifique. Cf. DAMASIO A., *Descartes' Error : Emotion, Reason and the Human Brain*, New York, 1994, p. 70.

familiers, habituels : “Le malade exécute, même les yeux fermés, avec une rapidité et une sûreté extraordinaire, les mouvements nécessaires à la vie, pourvu qu'ils soient habituels : il prend son mouchoir dans sa poche et se mouche, prend une allumette dans une boîte et allume une lampe. [...]”⁴⁴. Comme dans le cas des réseaux de neurones, l'association entre la situation et le schéma corporel correspondant se fait de façon automatique. La situation déclenche de façon automatique les habitudes corporelles et perceptives appropriées, et l'on agit⁴⁵.

Cependant, dans le cas du mouvement normal, on est capable non seulement de repasser par des chemins familiers, mais aussi de créer de nouveaux cheminements. On peut s'adapter à des situations qui sont indéterminées et se référer non seulement au monde actuel, mais aussi à un monde possible. Quand Schneider est confronté à de telles situations indéterminées, il est incapable de produire une réponse. Plus généralement, il est incapable de faire ce que Merleau-Ponty appelle des mouvements “abstraites”, des mouvements qui ne sont pas spécifiés à l'avance par une familiarité avec la situation⁴⁶. Le malade “est incapable d'exécuter des mouvements [...] qui ne s'adressent à aucune situation effective tels que mouvoir sur commande les bras ou les jambes, étendre ou fléchir un doigt. [...] Il ne peut exécuter des mouvements abstraits que si on lui permet de regarder le membre qui en est chargé ou d'exécuter avec tout son corps des mouvements préparatoires”⁴⁷. En quelque sorte, Schneider doit reconstituer ce genre de mouvements partie par partie, sans pouvoir y reconnaître une configuration d'ensemble qui suscite une réponse d'ensemble du corps⁴⁸. De nouveau, on constate une similarité avec les réseaux de neurones. On se souvient que pour les réseaux de neurones, tout contexte nouveau est simplement un nouveau contexte d'apprentissage, c'est-à-dire un amalgame de données disparates, classifiées dans des configurations sans signification particulière.

La maladie de Schneider possède donc certaines similarités avec les limitations contextuelles des réseaux de neurones. Dans les deux cas, il y a une dépendance étroite par rapport à des contextes habituels et les nouveaux contextes ne sont pas reconnus comme contexte, c'est-à-dire comme configuration globale évoquant des possibilités d'action, mais sont simplement une collection d'éléments disparates. Pour dépasser ces

44 MERLEAU-PONTY M., *op. cit.*, p. 120.

45 Cf. HILDITCH D., *op. cit.*, p. 143.

46 Cf. *ibid.*, p. 149.

47 MERLEAU-PONTY M., *op. cit.*, pp. 119–120.

48 Cf. HILDITCH D., *op. cit.*, p. 150.

limitations, il faut regarder du côté des alternatives, du côté du mouvement normal ou du corps productif. Pour Merleau-Ponty, l'attitude abstraite, ou le pouvoir de se rapporter à des contextes *actuels* et *possibles*, est liée à notre pouvoir de projeter un sens, à notre pouvoir de nous situer de façon spontanée dans l'action à travers nos projets : "Déjà la motricité, prise à l'état pur, possède le pouvoir élémentaire de donner un sens"⁴⁹. La projection de sens par le corps introduit une asymétrie forte entre le corps et le contexte. D'après Merleau-Ponty, c'est le schéma corporel qui garantit la cohérence des configurations globales de mouvement que l'on associe aux situations. "Toute théorie mécaniste se heurte au fait que l'apprentissage est systématique : le sujet ne soude pas des mouvements individuels à des stimuli individuels, mais acquiert le pouvoir de répondre par un certain type de solutions à une certaine forme de situations, les situations pouvant différer largement d'un cas à l'autre, les mouvements de réponse pouvant être confiés tantôt à un organe effecteur, tantôt à l'autre, situations et réponses se ressemblant dans les différents cas beaucoup moins par l'identité partielle des éléments que par la communauté de leur sens"⁵⁰.

Le caractère dynamique des représentations du corps, le pouvoir de les transférer d'un contexte à un autre, tient chez Merleau-Ponty à l'assomption de ces représentations dans une projection intentionnelle de sens (*Sinnggebung*). La projection intentionnelle, c'est le projet moteur du corps dans son ensemble, "à savoir la capacité de différenciation motrice du schéma corporel dynamique"⁵¹. C'est ultimement cette capacité de différenciation qui déterminera la spatialité du corps propre, le partage de l'espace selon le haut, le bas, le lointain et le proche, et non une quelconque spatialité présupposée au niveau de l'espace physique ou de l'espace abstrait de la géométrie.

L'assomption intentionnelle introduit une asymétrie forte entre le schéma corporel et le contexte. Non seulement on peut répondre par un même "mouvement de réponse" à des situations différentes, mais aussi les confier "tantôt à un organe effecteur, tantôt à un autre". La spatialité du corps est donc une propriété d'ensemble, *incarnée* tantôt dans tel, tantôt dans tel mouvement effectif et introduit une intégration d'un nouvel ordre. "Si j'ai l'habitude de conduire une voiture, je l'engage dans un chemin et je vois que 'je peux passer' sans comparer la largeur du chemin à celle des ailes, comme je franchis une porte sans comparer la largeur de la porte à celle de mon

49 MERLEAU-PONTY M., *op. cit.*, p. 166.

50 *Ibid.*, p. 166.

51 *Ibid.*

corps”⁵². La ressemblance entre “franchir la porte” ou “passer avec une voiture” ne tient pas à la ressemblance entre les situations, ni à la ressemblance entre les mouvements à effectuer, mais au fait que la voiture et le corps, dans les deux situations, sont des “puissances volumineuses, l'exigence d'un certain espace libre”⁵³. Par-delà les mouvements effectifs, et en quelque sorte avant tout mouvement effectif, il y a le projet moteur du corps dans son ensemble, l'exigence motrice ou encore une volonté de mouvement.

Nous aurons encore l'occasion, dans la deuxième partie de ce travail, de revenir sur la question de la projection intentionnelle de sens. Ce qui importe ici, c'est que la description du corps productif par Merleau-Ponty semble très éloignée de ce qu'on peut attendre des modèles en sciences cognitives. En sciences cognitives, on s'intéresse aux représentations ou états intentionnels effectifs. L'intentionnalité est conçue comme une opération effective sur un contenu représentationnel effectif. Le rapport au contenu est toujours un rapport à un contenu effectif possible et n'admet pas un rapport à un contenu simplement pensable sous forme d'une exigence. Ou, pour le formuler encore autrement, le rapport au contexte est un rapport à un contexte déterminable. Tandis que le projet moteur d'ensemble du corps, auquel fait allusion Merleau-Ponty, se rapporte à un contexte indéterminé.

Néanmoins, il nous semble que le contraste établi par Merleau-Ponty entre le cas du comportement du malade cérébral Schneider, cloué au monde, et le comportement d'une personne normale, capable de s'abstraire du monde, d'y projeter librement un sens, n'exclut pas l'exploration d'une piste intermédiaire. Selon les situations — passer la porte, passer un chemin étroit en voiture — notre réponse corporelle sera effectuée par tel ou tel mouvement. La spatialité du corps prend donc selon le contexte une incarnation différente. Ces différentes incarnations sont autant de médiations objectives du projet moteur et supposent un corps capable de ces médiations. L'étude des différentes médiations d'un même projet, cette fois-ci conçu comme projet dans l'effectif et pas simplement comme exigence, permet de mieux comprendre comment un sujet normal peut se rapporter d'une façon concrète et effective à différents contextes et par là de voir comment, quand le sujet est confronté à un contexte inédit ou non saturé, la productivité des représentations corporelles crée des médiations objectives nouvelles d'un même projet. C'est cette piste intermédiaire qui sera explorée dans la suite de ce paragraphe. Nous nous demandons comment on peut étendre les modèles connexionnistes classiques, trop liés à des contextes particuliers d'apprentissage,

52 *Ibid.*, p. 167.

53 *Ibid.*

pour mieux comprendre le caractère productif et actif de l'intelligence habituelle, intelligence que l'on retrouve sous la forme de la construction des catégories perceptives, des schémas corporels ou des représentations sociales.

3.2.2. L'adaptation contextuelle et l'activité en situation

Les réseaux de neurones ne permettent pas de rendre compte de la productivité et de l'activité de l'intelligence habituelle dans les comportements dont ils visent à rendre compte. Ils sont seulement capables de généralisations qui préservent la configuration globale de l'espace de réponse. Le programme NETtalk, par exemple, peut lire des mots incomplets ou jamais rencontrés. Il possède donc une certaine capacité de généralisation. Mais ces mots sont discriminés selon un vecteur de traits habituel et la sortie sonore utilise nécessairement les phonèmes de l'espace de sortie habituel. Dans ce type de généralisation on applique les représentations apprises, sans modification notable, à des cas nouveaux ou à de nouvelles situations. Il s'agit d'une forme de généralisation que l'on peut appeler, en empruntant le concept à Andy Clark, une généralisation contextuelle par similarité (*structure-preserving generalisation*)⁵⁴. Par de telles généralisations, on ne pourra jamais réellement produire de comportements nouveaux dans des contextes inédits ou modifier de façon dynamique la résolution d'un problème dans un contexte donné et au fur et à mesure que l'on découvre ce contexte.

3.2.2.1. La transformation de nos habitudes

La dépendance du contexte d'apprentissage des modèles connexionnistes peut d'abord être illustrée à partir d'un exemple. Supposons qu'un réseau connexionniste ait été programmé pour, disons, contrôler la marche d'un petit robot mobile. Considérons par exemple le robot à six pattes construit par Ken Espenschied et son équipe, pour marcher sur une surface plate présentant des obstacles fixes⁵⁵. Dans ce robot chaque patte a une certaine fréquence de mouvement et peut effectuer quelques

54 Cf. CLARK A., *Associative Engines : Connectionism, Concepts and Representational Change*, Cambridge (MA), 1993, p. 73.

55 Dans cet exemple, on s'inspire des robots mobiles de Ken Espenschied, sauf pour l'expérience de pensée du changement de contexte. Cf. QUINN R. et ESPENSCHIED K., "Control of a Hexapod Robot using a Biologically Inspired Neural Network", in BEER R. et al. (eds.), *Biological Neural Networks in Invertebrate Neuroethology and Robotics*, Boston, 1993.

mouvements simples : avancer, reculer et lever la patte. Le robot est également équipé de détecteurs tactiles pour évaluer sa position par rapport aux obstacles. Après la phase d'apprentissage, le réseau de neurones est capable d'associer une configuration motrice adéquate aux différentes situations que le robot rencontre dans son environnement. Ensuite, le robot est capable de marcher au hasard jusqu'au moment où il rencontre un obstacle fixe et le surmonter.

Qu'advient-il de notre robot si maintenant les obstacles se mettent à bouger d'une façon relativement régulière ? Dans cette nouvelle situation il faudrait pouvoir modifier certains paramètres, comme la fréquence de mouvement des pattes, tandis que d'autres, comme la hauteur à laquelle on lève les pattes quand on rencontre un obstacle, peuvent rester constants. Mais, comme les paramètres sont distribués sur l'ensemble des connexions du réseau de neurones, il est impossible de modifier certains paramètres, tout en laissant les autres paramètres intacts. La distribution de la connaissance sur l'ensemble du réseau rend la modification de fragments séparés, sans toucher aux autres fragments, impossible. Une telle modification locale demanderait de recommencer l'apprentissage à partir de zéro, cette fois en contexte mobile. Le caractère holistique de la réponse des réseaux de neurones, qui leur confère un caractère de réaction robuste et flexible dans leur contexte d'apprentissage, rend le réseau inopérant dans un contexte modifié de façon systématique.

Le transfert de la connaissance du contexte fixe au contexte mobile demande un réseau capable de transformer la structure de l'espace de réponse et pas seulement l'application d'une même structure à une nouvelle situation. Cette dernière capacité est désignée par Andy Clark par le concept de généralisation contextuelle transformatrice (*structure-transforming generalisation*)⁵⁶, en l'opposant à la simple généralisation contextuelle par similarité (*structure-preserving generalisation*). Les configurations globales qui émergent d'un réseau ne doivent pas seulement pouvoir être exportées d'un contexte à l'autre, mais aussi pouvoir être transformées à la rencontre d'un nouveau contexte.

On peut distinguer différents types de généralisation contextuelle transformatrice⁵⁷. La généralisation contextuelle peut par exemple prendre la forme de l'utilisation d'une ressource nouvelle dans un contexte habituel. Dans ce cas, on insère un facteur nouveau dans l'espace de réponse existant. Une personne qui apprend à utiliser un bâton pour marcher utilise un instrument nouveau pour améliorer sa marche, mais prend également appui

56 Cf. CLARK A., *Associative Engines*, op. cit., p. 73.

57 Cf. HILDITCH D., op. cit., pp. 159–76.

sur tout un ensemble de ressources existantes. Le résultat sera une nouvelle configuration globale de sortie qui combine les ressources nouvelles et les ressources existantes, dans un contexte d'action similaire.

En un sens plus fort, la généralisation contextuelle transformatrice implique également la possibilité d'investir des contextes systématiquement nouveaux. Au lieu d'insérer un nouveau facteur dans une configuration existante, on peut combiner et recombinaer sélectivement différentes parties d'habitudes existantes pour former de nouvelles réactions habituelles. Le transfert de la spatialité du corps propre à la voluminosité spatiale d'une voiture est de ce type : le conducteur de voiture prend appui sur la spatialité de son corps propre pour se positionner continuellement dans l'espace, mais recombine cette spatialité avec les propriétés spécifiques de la voiture. De même, quand on apprend la prononciation d'une nouvelle langue, on n'apprend pas seulement de nouvelles règles de conversion de texte en phonèmes, mais on produit un espace phonétique de sortie nouveau en combinant des phonèmes que l'on connaissait déjà avec des sons entièrement nouveaux. La transformation de nos habitudes peut donc se faire dans des contextes similaires ou dans des contextes systématiquement nouveaux. Dans les deux cas la nouveauté ne tient pas à la réorganisation complète de notre connaissance habituelle, mais à des améliorations locales⁵⁸. La transformation suppose donc de pouvoir séparer les configurations de réponse globales en parties utilisables de façon indépendante. Ensuite, on recompose ces parties dans de nouvelles configurations de sortie⁵⁹.

Le concept de généralisation contextuelle transformatrice permet ainsi de réfléchir sur les conditions d'incarnation du caractère dynamique de nos habitudes corporelles. Nous avons vu, en reprenant les analyses de Merleau-Ponty, que la spatialité du corps propre peut être transportée d'un contexte à l'autre grâce à la reprise, d'un contexte à l'autre, du projet intentionnel moteur : "S'habituer à un chapeau, à une automobile ou à un bâton, c'est s'installer en eux, ou inversement, les faire participer à la voluminosité du corps propre"⁶⁰. Maintenant, nous voyons que le transfert implique, selon le modèle de la transformation contextuelle, également la reprise de certains fragments anciens de réponse corporelle. "La communauté de sens" des configurations globales, appartenant à des contextes différents, implique donc également une certaine communauté au niveau des mouvements

58 Cf. CLARK A., *Associative Engines*, *op. cit.*, p. 85.

59 Dans le modèle des réseaux de neurones, la transformation prendra par exemple la forme d'un agrandissement du vecteur des traits ou d'une combinaison de morceaux de vecteurs pour en construire un nouveau. Cf. HILDITCH D., *op. cit.*, p. 168.

60 Cf. MERLEAU-PONTY M., *op. cit.*, p. 168.

effectifs. Du point de vue du projet comme exigence, le projet est compatible a priori avec tout contexte et tout type de réponse du corps. Mais du point de vue du projet effectif — et ici nous élargissons le point de vue de Merleau-Ponty — cela n'est plus le cas⁶¹. Le transfert du projet d'un contexte à l'autre prend appui sur les capacités d'innovation du corps effectif, la capacité d'utiliser des réactions habituelles déjà existantes dans de nouveaux contextes. Le caractère dynamique de l'intelligence habituelle au niveau des projets intentionnels doit trouver un ancrage au niveau de l'intelligence effective.

3.2.2.2. *La négociation active en situation*

Une deuxième façon d'étendre les modèles connexionnistes, à côté de la production de réponses nouvelles dans des contextes nouveaux, c'est d'envisager un rapport dynamique au contexte. Jusqu'ici nous avons toujours supposé que le contexte était donné d'avance de façon exhaustive et que le réseau calculait ensuite, de façon globale, la réponse par rapport à ce contexte. Toutefois, le contexte peut se modifier pendant l'exécution d'une certaine tâche et on découvre parfois le contexte seulement au fur et à mesure de l'exécution. Dans ces cas, le rapport au contexte est dynamique et non statique. De plus, même dans les contextes habituels, il y a toujours une certaine liberté par rapport au contexte. Certains paramètres de la réponse sont laissés indéterminés par le contexte et ils sont ajustés de façon active en cours de route. Le rapport avec le contexte se négocie de façon active, dans les limites circonscrites par les associations habituelles.

Un pianiste, par exemple, qui vient d'étudier une nouvelle partition peut la jouer en quelque sorte sans réfléchir, de façon immédiate. Cependant les différentes exécutions de la partition ne sont pas de simples reproductions habituelles identiques. Sinon le pianiste pourrait seulement rejouer la partition du début à la fin, sans accentuer un certain mouvement ou jouer des variations sur un thème⁶². L'apprentissage permet une certaine liberté par rapport à la partition, plutôt que la simple reproduction. Au cours de l'apprentissage, en s'appropriant la partition, celle-ci devient progressivement disponible pour une reprise créative au niveau de l'accentuation des notes, la

61 Comme le montre de façon convaincante R. Gély, Merleau-Ponty avait cependant déjà anticipé cet élargissement dans ses analyses du processus d'impossibilisation des parties et des facultés du corps, cf. GÉLY R., *La genèse du sentir*, Bruxelles/Paris, 2000, pp. 172–174.

62 Cf. HILDITCH D., *op. cit.*, p. 155. Pour une étude plus détaillée de l'approche cognitive de l'improvisation musicale, voir aussi KARMILOFF-SMITH A., *Beyond Modularity : A Developmental Perspective on Cognitive Science*, Cambridge (MA), 1992.

variation de la structure de la phrase musicale, etc. Le rapport au contexte habituel — ici la partition — est un rapport créatif et actif, plutôt que simple reproduction passive.

Dans une étude sur les pratiques cognitives dans nos comportements au quotidien, Jean Lave montre de façon plus détaillée comment on négocie de façon active la résolution de certains problèmes dans des contextes habituels⁶³. Si l'on s'aventure au supermarché par exemple, on doit faire preuve d'une grande capacité d'improvisation, malgré, par exemple, notre familiarité avec l'endroit ou un style habituel de comportement d'achat. D'un côté le contexte — l'environnement physique du supermarché — ne se montre qu'au fur et à mesure que l'on passe les différents produits en revue et, de l'autre, la résolution du problème — acheter ce dont on a besoin — se modifie en fonction de ce qu'on découvre dans ce contexte. Dans le langage du modèle des réseaux de neurones, on peut dire que non seulement les vecteurs d'entrée qui se présentent pour résoudre le problème se modifient au fur à mesure de l'activité, mais aussi la configuration de l'espace de sortie, les solutions élaborées. En étudiant nos réactions habituelles en situation, on découvre donc un double caractère dynamique. D'abord du côté des entrées du système, le contexte qui se modifie en permanence, et ensuite du côté des processus de transformation des entrées en sortie, la négociation active de la réponse en fonction de ce contexte. Selon Lave, le comportement d'achat sera le résultat d'un jeu dialectique entre la négociation active de l'acheteur et le contexte physique qui est découvert au fur et à mesure que l'on avance dans l'exécution de notre tâche.

Le point réellement nouveau qui intervient dans l'analyse de Lave, par rapport aux analyses précédentes, c'est l'intervention d'une négociation active de la réponse du sujet par rapport au contexte. Les capacités cognitives que nous avons étudiées jusqu'ici manifestaient toutes une dépendance beaucoup plus étroite par rapport au contexte. D'abord, au niveau de la généralisation habituelle par similarité, une dépendance par rapport au contexte d'apprentissage déterminant les réponses prototypiques et, ensuite, au niveau de la généralisation transformatrice, une dépendance par rapport au contexte systématiquement nouveau, qui détermine la façon dont il faut recomposer des fragments de réponse plus anciens en nouvelles réponses adéquates. Dans les comportements analysés par Lave, la réponse du système ne suit pas automatiquement l'évolution du contexte, mais est négociée de façon active. Pour désigner cette négociation active, Lave utilise la métaphore spatiale de la distance à franchir (*gap closing*), distance qui sépare la

63 LAVE J., *Cognition in Practice : Mind, Mathematics and Culture in Everyday Life*, Cambridge, 1988.

résolution actuelle du problème de sa résolution complète. Cette distance peut diminuer si les bonnes entrées se présentent au système, quand on trouve le bon produit dans un rayon du supermarché par exemple, mais aussi quand on décide de reformuler le problème, on renonce par exemple à tel ou tel achat. Cette reformulation active peut prendre la forme de calculs arithmétiques explicites de prix, de dialogues internes qui dirigent notre action vers de nouveaux produits ou même d'une réflexion autour des choix éthiques par rapport auquel le sujet veut s'engager⁶⁴.

Dans l'analyse de Lave, le sujet a une certaine liberté de comportement par rapport au contexte, liberté qui nécessite une négociation active en plus de la réaction habituelle globale par rapport à la situation. On retrouve cette négociation active dans l'improvisation musicale. La négociation prend dans ce cas la forme d'un ajustement progressif entre l'émotion évoquée par la musique et l'interprétation qui est tentée par le musicien. C'est-à-dire, on retrouve le rapport dialectique entre le sujet et le contexte analysé par Lave, mais sous la forme d'une exploration active du pouvoir expressif d'une partition⁶⁵. L'échange dynamique entre le système cognitif et le contexte, que ce soit par la négociation langagière, le calcul mathématique ou l'exploration émotionnelle, prend donc la forme d'un échange dialectique entre le sujet et le contexte. Au cours de la négociation, le rapport au contexte n'est pas un rapport d'adaptation contextuelle immédiate, mais une exploration active pour tenter de combler l'écart par rapport à la tâche à résoudre. En retour le contexte suscite une redéfinition permanente des pistes de solution que le système peut tenter et donc les stratégies d'explorations qu'il mettra en

64 Le passage à la représentation symbolique explicite, comme les symboles mathématiques ou langagiers, se fait donc au cours de nos comportements les plus habituels et pas uniquement en cas d'échec du rapport habituel avec l'environnement, comme dans la théorie des savoir-faire des Dreyfus. Cf. DREYFUS H.L. et DREYFUS S.E., "How to Stop Worrying About the Frame Problem Even Though It's Computationally Insoluble", in Pylyshin Z.E. (ed.), *The Robot's Dilemma : the Frame Problem in Artificial Intelligence*, 1987, pp. 95–111, p. 102.

65 Comme le précise David Hilditch, les émotions peuvent soit jouer un rôle actif, comme dans la négociation active, soit se réduire à une réaction passive de mon corps : "la conscience [émotionnelle] du corps peut soit émerger comme une interprétation personnelle de la situation par rapport à mes projets [...], ou (inversement) former le point de départ de l'exploration de la situation" (cf. HILDITCH D., *op. cit.*, p. 203 ; notre traduction). Dans le premier cas, on aura une négociation active du sujet avec le contexte : les émotions sont à la fois évoquées par la situation et renvoient aux projets du sujet dans cette situation. Les émotions ont un rôle actif par rapport à la situation. Tandis que dans le deuxième cas, les émotions déterminent à l'avance mon attitude globale par rapport à la situation. Elles font partie, de pair avec les réactions habituelles, de la détermination globale de la situation au sein de laquelle le jeu dialectique entre le sujet et le contexte pourra se dérouler. Elles sont purement passives ou automatiques.

oeuvre. De cette façon, par un aller et retour dialectique entre le contexte et le système, le système évolue progressivement vers un état d'équilibre qui correspond à une solution estimée satisfaisante par rapport au contexte en question.

3.3. Conclusion : Le connexionnisme et les formes d'intelligence

Les modèles cognitivistes classiques semblent inadéquats pour étudier des formes d'intelligence non symboliques, comme l'intelligence qui intervient au niveau des savoir-faire habituels. On se souvient de l'argument de Dreyfus qu'un conducteur de voiture n'agit pas en suivant des règles logiques explicites — comme dans le cognitivisme classique — mais plutôt par association habituelle d'un comportement adéquat à telle et à telle situation. Dans le contexte du modèle des réseaux de neurones, et plus particulièrement à partir du modèle des mémoires associatives, on semble être mieux armé pour surmonter ces problèmes. L'intelligence habituelle peut être modélisée à partir de quelques principes relativement simples. D'après le modèle des mémoires associatives, le conducteur peut mémoriser un grand nombre de situations différentes grâce à l'encodage vectoriel et regrouper les situations similaires autour de situations prototypiques grâce au pouvoir de classification des neurones. L'encodage vectoriel et l'association par similarité remplacent les bases de données et les règles de calcul des modèles classiques. En plus l'association par similarité permet une généralisation de nos comportements prototypiques à des situations nouvelles, sans devoir faire appel à des règles de raisonnement logique explicites. C'est réconfortant sur le plan intuitif, vu que dans la mesure où lorsqu'on demande à un expert (routier, cuisinier, chirurgien) de nous expliquer de façon explicite comment il a fait pour s'en sortir dans telle ou telle situation, on reste souvent sur sa faim. Le pouvoir associatif des mémoires associatives est plus grand que ce que l'on peut formuler de façon explicite dans le langage naturel⁶⁶.

66 Chez Dreyfus, la capacité de discrimination et de généralisation contextuelle correspond seulement au stade du débutant avancé. Pour devenir expert, il faut en plus pouvoir réduire l'espace de décision en construisant une hiérarchie de priorités. D'abord, au niveau de la compétence, la capacité de s'engager dans des plans d'action à partir d'une situation donnée et ensuite, au niveau de l'expertise, la capacité de reconnaître des stratégies globales adéquates par rapport à ces plans. Selon Dreyfus, les réseaux connexionnistes peuvent construire de telles hiérarchies de comportements typiques dont est capable l'expert. Cf. DREYFUS H.L., "La portée philosophique du connexionnisme", *op. cit.*, pp. 356–366 et p. 372.

Le modèle des réseaux de neurones semble donc plus satisfaisant que les modèles classiques d'intelligence artificielle pour modéliser l'intelligence habituelle. Il reste cependant un problème épineux, c'est la dépendance étroite des réactions habituelles par rapport au contexte d'apprentissage. Nous avons vu deux exemples de programmes d'apprentissage, l'apprentissage supervisé et l'apprentissage par auto-organisation. Dans la discussion de ces programmes, nous nous sommes surtout intéressés aux conditions d'arrêt des programmes ou, de façon équivalente, aux buts de l'apprentissage. Dans l'apprentissage supervisé, ce but est atteint si l'on peut correctement reproduire des catégories définies à l'avance (le genre, des phonèmes, ...). Dans le cas de l'auto-organisation, le but est atteint quand une règle locale de transformation entrée - sortie est satisfaite sur l'ensemble des entrées (des points voisins sur une figure présentée à l'entrée doivent être voisins sur la configuration de sortie par exemple). A partir de ces deux exemples de méthodes d'apprentissage nous avons vu comment le modèle des réseaux de neurones permet de donner des réponses satisfaisantes au problème de l'arrêt.

Mais, à côté de l'arrêt du programme, il faut aussi définir le contexte de fonctionnement, la relation à un domaine possible d'application. Le problème des modèles connexionnistes, c'est qu'ils sont très liés à un contexte particulier de fonctionnement. La seule généralisation dont ils sont capables est une généralisation par similarité. La dépendance étroite d'un contexte d'apprentissage particulier indique une première limitation contextuelle des modèles connexionnistes. Nous avons vu qu'il y a moyen d'étendre les modèles connexionnistes classiques pour inclure la généralisation des représentations à des contextes nouveaux. En effet, les mécanismes d'apprentissage des réseaux de neurones peuvent fonctionner comme une première étape de construction de représentations, dans des contextes bien définis. Ces représentations particulières pourront ensuite être transformées pour pouvoir faire face à des changements de contexte. Pour modéliser le transfert des représentations d'un contexte à un autre, il faut introduire une certaine productivité ou capacité d'innovation dans les réseaux. Cependant, celle-ci n'est pas une innovation complètement aléatoire, *ex nihilo*. Il s'agit d'une innovation conditionnée par des comportements habituels plus anciens.

Dans ce paragraphe, nous avons également étudié une deuxième extension des modèles connexionnistes, qui concernait les situations où le contexte ne détermine pas entièrement la réponse habituelle. Ni les modèles connexionnistes classiques, ni leur extension vers la généralisation transformatrice ne peuvent modéliser une indétermination de la réponse par

rapport au contexte ou la possibilité de moduler la réponse par rapport à ce contexte par une négociation active. L'étude de la négociation active du système cognitif dans certains contextes habituels indique une deuxième limitation contextuelle des modèles connexionnistes, limitation qui concerne la trop étroite dépendance de la réponse par rapport aux entrées du système dans ces modèles.

La combinaison de la négociation active avec des réponses habituelles suggère une architecture cognitive où interviennent deux composantes. Une composante habituelle qui dépend étroitement du contexte, d'une part, et une composante active qui tente de résoudre la part d'indétermination laissée par ce contexte au niveau de la réponse du système (le *gap closing* de Lave). Dans la composante habituelle interviennent des schèmes d'encodage statistique et des mécanismes de calcul par similarité ou par généralisation transformatrice. Dans la composante active interviennent différentes formes d'échange dialectique avec le contexte, comme des évaluations éthiques, des calculs mathématiques ou des explorations émotionnelles. Ces échanges impliquent, au moins dans certaines situations, la manipulation de représentations symboliques explicites, comme des nombres ou des expressions verbales, et donc l'utilisation de formes de représentation fondamentalement différentes de l'encodage vectoriel dans des réseaux de neurones. Les deux composantes de l'architecture cognitive supposent donc également deux systèmes représentationnels, un système basé sur l'encodage vectoriel statistique et un système d'encodage symbolique.

Une architecture en étages permettra alors de couvrir des processus allant de l'association prototypique dans les réseaux de neurones jusqu'aux calculs logiques explicites des systèmes formels, en passant par la généralisation transformatrice et les capacités de négociation active en situation. Une telle architecture reflète bien l'état actuel des recherches en sciences cognitives et développe de façon concrète la conception centrale de la cognition comme computation de fonctions de traitement d'information. Certains chercheurs pencheront davantage du côté d'un format de représentation symbolique pour l'encodage des informations, d'autres davantage du côté d'un format de représentation statistique distribuée, comme dans les réseaux de neurones. Néanmoins, comme nous avons tenté de le montrer dans ce paragraphe, les deux formats de représentation sont nécessaires pour rendre compte des capacités du système cognitif humain. D'un côté, les représentations symboliques ne peuvent pas rendre compte de la capacité de discrimination, de reconnaissance et d'adaptation contextuelle quasi automatique de l'intelligence habituelle. Les modèles connexionnistes permettent de reproduire ces propriétés. De l'autre, nous avons constaté

l'étroite dépendance des modèles connexionnistes par rapport au contexte. Si le transfert d'une représentation d'un contexte à l'autre peut encore se modéliser, moyennant quelques amendements, dans un cadre connexionniste, on ne pourra jamais rendre compte de certaines formes de négociation active de la réponse. On retrouve dans cette négociation des formes de représentation symbolique, comme les représentations verbales explicites ou les représentations mathématiques. Dans le cas des raisonnements abstraits ou formels étudiés par les programmes d'intelligence artificielle, on se détache plus encore du contexte effectivement présent et les systèmes cognitifs auront davantage recours à des représentations symboliques.

La présentation des systèmes cognitifs dans ce paragraphe, même si elle fournit un aperçu relativement large des différents schèmes d'encodage et de calcul, reste cependant très incomplète. En effet, nous nous sommes essentiellement intéressés à ce qui se passe à l'intérieur du système. De cette façon, nous avons pu définir différentes formes d'intelligence, comme l'intelligence symbolique, l'intelligence habituelle ou l'intelligence exploratoire (la négociation active). Chacune de ces formes se caractérise par des mécanismes computationnels différents et permet ainsi de définir différentes formes d'opérationnalisation de l'intentionnalité.

Toutefois, en nous intéressant uniquement à l'aspect opératoire, nous avons présupposé que les entrées du système sont des données de l'environnement et que les sorties sont entièrement le résultat des processus de traitement d'information dans le système. Si nous voulons comprendre les systèmes intentionnels, nous ne pouvons pas partir d'une telle définition artificielle des entrées et des sorties. Comme nous le verrons dans le paragraphe suivant, les systèmes cognitifs interviennent activement dans la définition des entrées et des sorties, selon leurs intérêts et leurs dispositions propres. De plus, le contexte du système joue également un rôle actif dans la réalisation des fonctions cognitives. Les modèles connexionnistes, aussi bien que les modèles d'intelligence artificielle, surestiment le poids de la représentation et de la computation. De cette façon, ils ne prennent pas en compte les vertus dynamiques de notre corps et de l'environnement naturel.

4. Au-delà des mécanismes internes : l'écologie de la pensée

Dans ce chapitre, nous nous sommes intéressés aux mécanismes computationnels qui caractérisent les opérations intentionnelles. Nous avons d'abord présenté les mécanismes de calcul logique et les principes d'encodage des représentations dans des bases de données dans le cognitivisme classique, et ensuite nous avons étudié plus dans les détails les

mécanismes d'association par similarité et d'encodage vectoriel dans le connexionnisme. Cette première perspective purement computationnelle est toutefois insuffisante si l'on veut modéliser également les aspects représentationnels et fonctionnels de l'intentionnalité. Dans le troisième paragraphe de ce chapitre, nous étudierons la problématique de la modélisation des représentations et des fonctions, par-delà le problème du format d'encodage des représentations ou des mécanismes de calcul des fonctions.

Une première façon d'aborder la problématique des représentations est de partir de l'hypothèse que les systèmes cognitifs visent à élaborer des représentations objectives de l'environnement naturel et social. Une telle manière d'envisager la problématique des représentations est une des présuppositions le plus partagées des sciences cognitives : les représentations manipulées par les systèmes cognitifs véhiculent une information objective sur l'environnement. Mais comment justifier une telle présupposition ? Nous avons vu que les systèmes cognitifs construisent des représentations par des mécanismes d'association ou des règles logiques (si ... alors ...), en partant de données de l'environnement. Cependant, les seules données de l'environnement captées par le système cognitif sont les énergies et les forces physiques qui excitent les neurones sensoriels (fréquences lumineuses ou sonores, forces mécaniques, chaleur, etc.). Comment maintenir le réalisme des représentations, qui typiquement se réfèrent à des objets et des événements dans l'environnement, si les seules données objectives dont le système dispose sont des énergies et des forces physiques ? On pourrait s'imaginer le raisonnement suivant. Les systèmes cognitifs interagissent de façon cohérente avec des objets distants de notre environnement. Il faut donc bien qu'ils arrivent à construire des représentations véridiques de ces objets. Les informations contenues dans les entrées sensorielles sont suffisantes pour reconstruire, par des inférences cognitives (associations, règles d'interprétation), une représentation des objets dont les stimuli sensoriels proviennent. Par l'interaction répétitive avec ces objets, on peut alors ajuster les représentations pour les rendre plus véridiques.

Ce raisonnement plausible sur le plan logique doit encore être vérifié sur le plan empirique. Par exemple, on peut trouver un raisonnement semblable, dans un des programmes les plus influents en sciences cognitives, la théorie de la vision de David Marr⁶⁷. Cependant, la théorie du réalisme représentationnel est contestée aujourd'hui par de plus en plus de chercheurs, à partir de matériel expérimental provenant à la fois de la psychologie du

67 MARR D., *Vision*, San Francisco, 1982.

développement, de la robotique et des neurosciences⁶⁸. L'interaction réussie avec des objets distants de l'environnement n'impose aucunement que nos représentations soient nécessairement véridiques. Au contraire, dans la plupart des situations pratiques, on construit seulement des représentations partielles, en raisonnant sur des données incomplètes. Ces représentations sont suffisantes pour nous guider dans nos actions au quotidien. De plus, la conception réaliste des représentations que nous avons esquissée se base sur la présupposition que le monde et le corps sont des sources d'entrées non signifiantes, neutres. Mais, comme nous le verrons, les systèmes cognitifs interviennent de façon active dans la définition des entrées et on pourra difficilement maintenir la thèse de la neutralité. L'image de la cognition qui est suggérée par ces expériences est celle d'une cognition dont les opérations sont relatives à un milieu de vie (*Umwelt*) particulier. En fonction du répertoire comportemental ou des intérêts d'action de l'organisme, le système cognitif définit à la fois les entrées et les opérations de transformation à effectuer pour construire les représentations appropriées.

L'étude des représentations dans un milieu de vie fera l'objet de la première section de ce paragraphe. Dans une deuxième section, nous verrons que l'on peut encore pousser plus loin la remise en question du computationnalisme. En effet, dans un contexte naturel, la cognition opère en interaction étroite avec le corps et le monde. Nous verrons que ceci modifie profondément la façon dont les systèmes cognitifs réalisent les fonctions cognitives. En particulier, l'environnement et le corps deviennent des partenaires actifs qui prennent une part de la résolution du problème à leur compte, limitant ainsi la charge computationnelle des systèmes cognitifs.

Pour donner une première idée de l'interaction des systèmes cognitifs avec les ressources dynamiques de leur environnement, on peut reprendre un exemple donné par Andy Clark dans son ouvrage *Being There*⁶⁹. Considérons le cas d'une tâche manuelle, où l'on doit emboîter des pièces finement ajustées. C'est une tâche que l'on essaie de simuler en robotique pour construire des lignes d'assemblage entièrement automatisées. D'après Clark, on peut envisager deux solutions très différentes à ce problème de simulation. La première solution suggérée par Clark est la solution mentaliste, qui ne tient pas compte de l'interaction avec l'environnement :

68 Voir, par exemple, pour un aperçu général des différents arguments, CHURCHLAND P.S., RAMACHANDRAN V.S. et SEJNOWSKI T.J., "A Critique of Pure Vision", in KOCH C. et DAVIS J. (eds.), *Large-scale Neuronal Theories of the Brain*, Cambridge (MA), 1994, pp. 23-60.

69 Cf. CLARK A., *Being There*, *op. cit.*, p. XII.

“Confronté à la question du contrôle d'un robot par un ordinateur pour réaliser une tâche d'emboîtement de pièces, on peut construire de multiples boucles de rétroaction. Ces boucles peuvent signaler à l'ordinateur si le robot n'a pas réussi à emboîter une pièce et lui permettre de recalculer le mouvement dans une direction légèrement différente”⁷⁰. Toutefois on peut également envisager une deuxième solution, qui tient compte des ressources dynamiques de l'environnement : “La cognition corporelle et contextuelle aborde le problème d'une façon différente. On peut par exemple attacher simplement le bras du robot à un joint en caoutchouc, qui laisse du jeu dans deux directions de l'espace. Dans cette nouvelle situation, l'ordinateur ne doit plus calculer les boucles de régulation fine du mouvement. Les pièces ‘se glissent et se faufilent’ en bonne position comme si de multiples ajustements de rétroaction étaient calculés continuellement”⁷¹. Dans cette deuxième solution, le joint en caoutchouc remplace le calcul des ajustements par l'ordinateur.

Cet exemple illustre qu'une conception passive de l'environnement tend à traiter la cognition comme un calcul purement mental et à faire abstraction du corps et du monde dans lequel le système fonctionne. Mais, comme nous le verrons, l'étude de la cognition dans des situations réelles suggère que les ressources utilisées par le système cognitif ne sont pas toutes représentées de façon effective. Souvent le système cognitif exploite directement les vertus dynamiques du corps et de l'environnement naturel. De cette façon la cognition est non seulement située dans un milieu de vie, mais aussi inscrite de façon effective dans son corps et dans le contexte. Cette inscription peut prendre la forme d'un simple retour d'information du contexte ou du corps, mais aussi une forme plus élaborée, comme le couplage dynamique, ou même la dilution de la pensée dans un système dynamique plus large qui émerge de l'interaction du corps, du système cognitif et du monde.

4.1. L'écologie subjective

La théorie du réalisme représentationnel considère les représentations comme étant des copies aussi fiables que possible des objets ou des événements de notre environnement. Dans cette section, nous voulons montrer qu'au lieu d'élaborations exhaustives de scènes ou d'événements dans le monde, les représentations sont plutôt des représentations partielles et véhiculent seulement l'information immédiatement pertinente pour le système. En plus cette information est souvent déformée. Elle est encodée

70 *Ibid.*, notre traduction.

71 *Ibid.* ; notre traduction.

sous un format qui reflète les intérêts contextuels ou les dispositions de l'organisme en question. La théorie alternative qui émerge à partir de ces considérations peut être qualifiée d'écologique. Elle considère que les représentations du système cognitif ne peuvent être étudiées qu'en se référant aux relations que le système cognitif entretient avec son milieu de vie.

Une façon d'illustrer la différence entre une telle théorie écologique et la théorie du réalisme représentationnel consiste à faire la distinction entre l'environnement d'un système cognitif et son milieu. Selon la théorie écologique, les représentations réfèrent seulement aux traits objectifs d'un milieu de vie, le fragment significatif de l'environnement, et non aux traits objectifs de l'environnement dans sa totalité. Nous ne nions évidemment pas la possibilité, dans certains contextes, de construire des représentations véridiques de l'environnement objectif. Toutefois, dans la théorie écologique, le rapport entre le milieu et l'environnement est inversé. De prime abord le système cognitif se rapporte à un milieu significatif, et seulement ensuite, en faisant abstraction des significations immédiates, il construit des représentations de l'environnement objectif, intersubjectivement partageables. L'environnement objectif est un cas particulier de milieu significatif. Dans la théorie du réalisme des représentations au contraire, le système cognitif se rapporte d'abord à un environnement neutre. Cet environnement fournit les entrées pour la construction de représentations véridiques, représentations auxquelles on peut alors accorder différentes significations. De ce point de vue, le milieu est un fragment de l'environnement objectif.

Pour mettre mieux en évidence cette distinction entre le milieu et l'environnement, prenons l'exemple d'un système cognitif extrêmement simple, celui de l'abeille⁷². On a pu montrer que les abeilles se posent de préférence sur les figures qui ont une forme ouverte, comme les étoiles et les croix, et évitent celles qui ont une forme fermée, comme le carré et le cercle. Le système perceptif de l'abeille possède donc une capacité de discrimination de formes, mais ces formes sont tout à fait spécifiques pour son espèce. La figure 1.7 montre à quoi pourrait ressembler le milieu de vie d'une abeille d'après les formes perceptives qu'elle peut discriminer. L'environnement est donné par une prairie en fleurs, dans laquelle alternent les fleurs écloses et les boutons. Le milieu est simplement donné par un ensemble de traits perceptifs, comme les étoiles et les cercles. Ces traits perceptifs spécifiques du milieu servent les intérêts biologiques de l'abeille : seules les fleurs (étoiles et croix) et non les boutons (cercles) ont une signification pour l'abeille.

72 Cf. UEXKÜLL J.V., *Mondes animaux et monde humain*, Paris, 1965, p. 52.

Figure 1.7. L'environnement (à gauche) et le milieu de vie (à droite) de l'abeille (source : UEXKÜLL J.V., 1965, planche 5).

Qu'en est-il des systèmes cognitifs plus complexes, comme celui de l'être humain ? Est-ce que la distinction entre le milieu et l'environnement y est toujours valable ? Nous voulons montrer que oui. D'abord, nous montrerons que du côté des entrées du système, l'information captée est une information partielle. Seul l'information pertinente est représentée. Les représentations de l'environnement sont des représentations partielles. Nous verrons ensuite, du côté de l'utilisation de cette information ou de sa signification, que la construction des représentations est relative au contexte d'action ou aux dispositions de comportement de l'organisme. Nous parlerons dans ces cas de représentations pratico-sociales.

4.1.1. Les représentations partielles

La théorie de la vision de David Marr se rapproche assez bien de la théorie du réalisme représentationnel que nous avons esquissée. Du point de vue d'une telle théorie de la représentation, la théorie de Marr peut être caractérisée de façon schématique par les deux hypothèses suivantes⁷³ :

- a. L'objectif spécifique du système visuel est de créer un modèle détaillé du monde. Typiquement, la réalisation de cet objectif consiste en la transformation de données à deux dimensions sur la rétine en une description tridimensionnelle du monde spatio-temporel.
- b. La transformation du signal se fait de façon hiérarchique. Le système extrait, dans des étapes successives de traitement, des traits

73 Cf. CHURCHLAND P.S. *et al.*, "A Critique of Pure Vision", *op. cit.*, p. 24.

de plus en plus élaborés et intègre les différents traits jusqu'à obtenir la représentation entièrement élaborée de la scène. Les niveaux supérieurs de traitement d'information n'ont pas d'influence sur les niveaux inférieurs.

Des données expérimentales venant de la psychologie de la perception, en particulier celles de V. S. Ramachandran et ses collaborateurs, suggèrent une image très différente de la vision⁷⁴. Selon ces expériences, le système visuel utilise des représentations partielles, au lieu de créer des modèles détaillés du monde, et ceci non de façon hiérarchique, mais de façon interactive : les niveaux supérieurs, comme la reconnaissance des formes ou l'attribution des significations, interagissent avec les niveaux inférieurs du processus de traitement, comme la segmentation de l'image ou la représentation du mouvement.

Considérons, par exemple, le dispositif présenté à la figure 1.8⁷⁵. Il s'agit d'un écran de télévision qui montre une matrice de points clignotants. On voit certains points illuminés (représentés sur la figure en noir) et d'autres éteints (représentés sur la figure en blanc). Tous les points clignotent à la même fréquence, quand les uns sont éteints les autres sont allumés, et les points sont disposés de façon à ce que les points s'alternent, un allumé, un éteint, etc. A vrai dire, ce que l'on observe quand on regarde cet écran, ce ne sont pas des points clignotants, mais un mouvement apparent des points. Le cerveau relie les points deux par deux et on les voit bouger de gauche à droite ou de haut en bas.

L'illusion de mouvement, créée à partir d'images fixes s'alternant, est tout à fait classique — que l'on pense au cinéma par exemple — et il s'agit d'un processus de traitement de bas niveau. Ce qui est étonnant dans cette expérience, c'est que le mouvement n'est pas traité de façon locale, mais que l'on perçoit le même mouvement d'oscillation pour l'ensemble des points. Plus précisément, il y a deux directions de mouvement possibles, de gauche à droite ou de haut en bas, mais dès que l'on fixe une direction de mouvement pour deux points particuliers, les autres points suivent automatiquement ce même mouvement. Si l'on change de direction à un endroit donné, l'ensemble des points change également. Des sujets normaux

74 Cf. par exemple RAMACHANDRAN V.S., "Interactions between Motion, Depth, Color and Form : the Utilitarian Theory of Perception", in BLAKEMORE C. (ed.), *Vision : Coding and Efficiency*, Cambridge, 1990. A côté des articles de V.S. Ramachandran et de P.S. Churchland *et al.*, on trouvera un aperçu intéressant d'expériences montrant le caractère partiel ou personnel — relatif à un milieu de vie — de nos représentations dans BALLARD D., "Animate Vision", in *Artificial Intelligence*, 48 (1991), pp. 57-86.

75 Cf. CHURCHLAND P.S. *et al.*, "A Critique of Pure Vision", *op. cit.*, p. 30.

ne voient jamais un mélange de certains mouvements horizontaux et certains mouvements verticaux. Soit ils voient un mouvement horizontal d'ensemble, soit un mouvement vertical d'ensemble.

Figure 1.8. Un écran de points clignotants de façon alternative. Un observateur voit l'ensemble des points bouger soit horizontalement, soit verticalement. Il ne les voit jamais bouger par la diagonale et jamais un mélange de mouvements horizontaux et verticaux (source CHURCHLAND P.S. *et al.*, 1994, p. 30).

L'illusion produite par un ensemble de points clignotant de façon synchrone est un cas particulier d'entraînement de mouvement (*motion capture*)⁷⁶. Elle illustre le fait que des considérations globales sur la scène visuelle influencent le traitement local du mouvement, même à un niveau très élémentaire de traitement. Une autre expérience illustrant ce même effet montre une matrice de six points disposés en deux colonnes, dont le point au milieu de la deuxième colonne est caché derrière un petit carré⁷⁷. Quand les points clignotent on voit un mouvement apparent de gauche à droite, le point au milieu de la colonne de gauche paraît disparaître à chaque clignotement derrière le carré disposé au milieu de la colonne de droite. Cependant, si l'on montre uniquement le point du milieu de la première colonne et le carré, l'illusion de mouvement disparaît. On voit seulement un carré immobile et un point clignotant sur place. Cette deuxième expérience montre de nouveau l'effet de l'entraînement d'un mouvement local par la disposition globale d'une scène. Mais, elle a un intérêt supplémentaire. Elle permet de montrer des effets d'interaction entre des modalités sensorielles, même à un niveau aussi bas du processus de traitement. En effet, on peut refaire une expérience sur base de la dernière disposition expérimentale. A partir du point

⁷⁶ Cf. *ibid.*, p. 29.

⁷⁷ Cf. *ibid.*, p. 30–31.

clignotant et du carré, on peut recréer une illusion de mouvement rien qu'en faisant entendre de façon alternée un son à la même fréquence que le clignotement, appliqué à l'oreille de gauche quand le point est allumé et appliqué à l'oreille de droite quand le point est éteint. Influencé par le son qui s'alterne de gauche à droite, on a de nouveau l'impression de voir un point en mouvement alterné.

Ces expériences contredisent les hypothèses de travail de la théorie représentationnelle classique. Classiquement les modalités sensorielles, comme la vision et l'audition, traitent l'information de façon indépendante et les représentations se combinent seulement à un stade de traitement plus élaboré. De même, à partir de l'effet d'entraînement de mouvement, l'explication classique séquentielle de la représentation du mouvement d'une scène est contredite. D'après l'explication classique, on calcule d'abord le mouvement des points individuels et, à un stade ultérieur, on combine les mouvements locaux de façon cohérente pour construire la représentation globale. Dans les expériences de Ramachandran exposées ci-dessus, le niveau de la construction de représentations globales, niveau ultérieur dans la théorie classique, intervient directement dans le calcul des mouvements locaux.

Au lieu de calculer de façon détaillée chaque élément de la scène, le système visuel utilise donc un court-circuit pour ranger l'ensemble des mouvements locaux dans un même mouvement global. L'on peut rencontrer dans d'autres situations l'utilisation de raccourcis ou de chemins de traverse semblables. Un masque creux, par exemple, présenté de façon concave (avec le nez pointant de l'autre côté) paraît convexe (avec le nez pointant vers nous) à partir d'une certaine distance d'éloignement. L'information sémantique de haut niveau — on reconnaît que le masque représente un visage — influence le processus de construction de formes tridimensionnelles à un niveau plus bas et on se représente le visage sous sa forme habituelle convexe⁷⁸.

Cependant, l'utilisation permanente de raccourcis ne rend-elle pas le système visuel extrêmement fragile ? En effet, tel raccourci peut être utile dans telle situation, mais peut nous induire en erreur dans telle autre. La réponse que donne Ramachandran à cette objection est double. D'abord, l'utilisation des raccourcis peut être justifiée à partir d'un argument écologique⁷⁹. Le calcul rapide fait par le système visuel, en fonction d'une information partielle, se base sur des contraintes qui sont habituellement (et

78 Cf. *ibid.*, p. 33–34.

79 Cf. RAMACHANDRAN V.S., *op. cit.*, pp. 348–349.

souvent trivialement) vérifiées dans le milieu de vie de l'organisme. Il n'y a simplement aucune raison de faire les calculs exhaustifs élaborés.

Supposons, par exemple, que l'on voie un léopard à la chasse. On peut se demander comment le système visuel fait correspondre les points individuels d'une image du léopard en mouvement, par exemple les taches sur son dos, aux points de l'image qui succède immédiatement. Il existe une solution computationnelle élégante à ce problème, proposée par Shimon Ullman, un collaborateur de David Marr⁸⁰. Ullman suggère que le système visuel calcule différentes combinaisons possibles des points et, parmi celles-ci, prend la combinaison qui produit la distance minimale globale entre les points des deux images⁸¹. Mais, ce calcul exhaustif du mouvement de chaque point n'est absolument pas nécessaire, vu que dans des situations normales, les taches du léopard ne vont pas bouger dans des sens différents. L'on peut supposer qu'il s'agit d'un mouvement d'ensemble et, à partir de la correspondance d'une seule tache d'une image à l'autre, extrapoler la correspondance directement aux autres points. C'est probablement ce que fait le système visuel, comme on peut s'en rendre compte à partir des expériences sur les effets d'entraînement de mouvement. Dans la perception du mouvement du léopard, le système visuel néglige des quantités énormes d'information de la scène visuelle, concernant par exemple le détail du mouvement des points individuels, et construit, en combinant des informations partielles, une représentation utile de la scène. Le raccourci de calcul ne risque pas de nous induire en erreur, si on applique cette stratégie de calcul dans des contextes appropriés, par exemple chaque fois que l'on rencontre des animaux tachetés en mouvement dans notre milieu de vie.

A côté de l'argument écologique, on peut aussi avancer un argument computationnel⁸². En effet, souvent, le système visuel combine différents raccourcis pour résoudre un problème computationnel particulier. Au lieu d'utiliser un seul algorithme, calculant de façon séquentielle les différentes étapes de la résolution, le système visuel obtient une réponse adéquate en combinant des trucs partiels, trucs qui sont autant de raccourcis circonstanciés. C'est cette caractéristique supplémentaire du système visuel, l'utilisation en parallèle de différents dispositifs de calculs rapides ou raccourcis, qui donne une certaine flexibilité au système visuel. Au lieu d'avoir un seul algorithme, mais extrêmement élaboré pour pouvoir faire

80 Cf. ULLMAN S., *The Interpretation of Visual Motion*, Cambridge (MA), 1979.

81 L'algorithme de reconnaissance des formes de Kohonen, que nous avons étudié dans le deuxième chapitre, est basé sur un principe semblable (cf. KOHONEN T., *op. cit.*). La proposition de Ullman peut donc être implémentée directement dans un réseau de neurones.

82 Cf. RAMACHANDRAN V.S., *op. cit.*, p. 359.

face de façon flexible à un grand nombre de situations, on aura une collection de calculs rigides, mais que l'on peut combiner de différentes façons selon les situations. Par exemple, l'extraction de formes tridimensionnelles à partir de points en mouvement peut se faire par un algorithme de calcul explicite⁸³. Mais, on peut montrer que ce même résultat peut être obtenu en combinant différentes stratégies partielles du même type que ceux que nous avons rencontrés plus haut⁸⁴.

Finalement, l'utilisation de différents raccourcis circonstanciés paraît plus plausible d'un point de vue biologique. Les capacités sophistiquées du système visuel sont le résultat d'une évolution naturelle. Comme l'évolution n'a pas de capacités d'anticipation, il semble plus facile d'obtenir une collection de trucs mis ensemble pour accomplir une certaine tâche, qu'un seul algorithme entièrement élaboré. L'utilisation de fragments matériels plus anciens, ici les raccourcis computationnels implémentés dans des circuits du système nerveux, est tout à fait courante dans l'évolution⁸⁵. Souvent, cela donne lieu à des solutions qui sont très éloignées du design optimal et élégant que l'on peut concevoir dans les systèmes artificiels de vision, mais il s'agit d'une inélégance qui produit des résultats remarquables.

4.1.2. Les représentations pratico-sociales

La considération des représentations partielles montre que la distinction entre le milieu et l'environnement est également valable pour des systèmes cognitifs plus complexes. Les systèmes cognitifs sélectionnent dans l'environnement seulement les paramètres qui ont une signification par rapport à leur comportement. De cette façon, les représentations du système cognitif définissent un monde opératoire ou — par analogie avec la biologie — une certaine niche sémantique. Cette approche écologique des représentations n'exclut pas l'encodage véridique des représentations. Mais, du point de vue de la théorie écologique, l'encodage véridique constitue seulement une niche particulière et pas le point de départ hypothétiquement neutre de la construction des représentations. Selon l'usage qui est fait des représentations, il sera plus économique d'encoder l'information sous une forme véridique ou avec un biais contextuel plus ou moins prononcé. Dans le deuxième cas, nous parlerons de représentations pratico-sociales. Remarquons que la distinction entre les deux catégories de représentations, véridiques et

83 Cf. ULLMAN S., *op. cit.*, p. 135.

84 RAMACHANDRAN V.S., *op. cit.*, pp. 357–359.

85 Cf., par exemple, GOULD S.J., *Le pouce du Panda. Les grandes énigmes de l'évolution*, trad. par J. Chabert, Paris, 1982.

pratico-sociales, n'est pas une distinction absolue, mais relative. Les représentations véridiques sont elles aussi relatives à un milieu de vie et dépendent donc ultimement du répertoire de comportement ou du style de vie de l'organisme. Seulement, elles sont encodées sous un format qui n'est pas spécifique pour telle ou telle action et peuvent donc être utilisées par des processus différents de manipulation de représentations.

Si d'un point de vue théorique les deux catégories de représentation sont possibles, des travaux récents suggèrent que l'encodage des représentations dans un format spécifique à l'action est omniprésent et plus fondamental. Alain Berthoz, dans ses travaux sur la représentation de l'espace, distingue pas moins de 4 catégories différentes de référentiels pour situer les objets dans l'espace et tous ces référentiels prennent comme point de référence un endroit du corps⁸⁶. D'abord, il y a le référentiel réтино-topique, qui permet de situer des objets par rapport à la position des yeux. Ensuite, il y a le référentiel vestibulaire au niveau de l'oreille interne, qui permet de situer les objets par rapport aux trois axes de mouvement de la tête et le référentiel postural qui, à partir des informations tactiles venant des pieds par exemple, situe les objets par rapport à l'axe vertical de notre corps. Finalement, on peut prendre comme référentiel certains membres du corps, comme les mains, les pieds, etc.

Selon l'action en cours, on utilisera l'un ou l'autre référentiel. Si l'on pointe un doigt vers une cible, par exemple, le cerveau simplifie le calcul en ne contrôlant que deux variables cinématiques : l'élévation et l'azimut par rapport à la verticale du lieu. Un point de rotation situé sur l'épaule est utilisé comme origine des axes de référence⁸⁷. Dans ce cas, le référentiel est lié à un membre particulier du corps. Pour maintenir la tête en position verticale stable en courant, on ne peut se référer au sol, puisque nos pieds ne touchent le sol que de temps en temps. Le cerveau utilise la détection de la gravité par le système vestibulaire pour stabiliser la tête et créer une plate-forme fictive au niveau de la tête comme référentiel⁸⁸. Le référentiel utilisé est situé au niveau du système vestibulaire.

A côté du choix de référentiel en fonction de l'action, le cerveau peut construire un référentiel nouveau approprié dans des contextes d'action nouveaux. Par exemple, des astronautes dans une navette spatiale ont beaucoup de difficultés, dues à l'absence de la gravité, à se maintenir en posture verticale par rapport au plancher de la station. En particulier, le référentiel postural, dans lequel l'information venant des muscles des pieds

86 Cf. BERTHOZ A., *Le sens du mouvement*, Paris, 1997, pp. 107–124.

87 Cf. *ibid.*, p. 118.

88 Cf. *ibid.*, p. 112.

est importante (si la cheville est tendue, c'est qu'on est penché en avant, par exemple), ne peut plus être utilisé. Néanmoins, après quelques jours, les astronautes savent reconstruire un nouveau référentiel postural, en fonction de la nouvelle signification qu'on peut donner à l'information venant des muscles des pieds dans la situation sans gravité, et ils peuvent se maintenir droit sans devoir vérifier leur position de façon visuelle⁸⁹. Ces expériences spectaculaires montrent non seulement le caractère spécifique de l'encodage spatial par rapport à une certaine action, mais en plus la flexibilité de cet encodage par rapport aux situations. Comme l'écrit Berthoz, "le cerveau cherche à élaborer une référence, il extrait du monde physique une grandeur pertinente qui simplifie le traitement neuronal des informations sensorielles et guide l'action. L'action est accrochée à un référentiel"⁹⁰.

A côté des référentiels qui prennent comme point de référence un endroit du corps, les référentiels égocentrés, Berthoz distingue également les référentiels allocentrés, liés à un objet de l'environnement ou à des relations entre des objets de mon environnement (je situe tous les objets dans cette pièce par rapport à la porte par exemple). L'usage de ces référentiels est de nouveau spécifique à telle ou telle action, mais cette fois-ci l'encodage ou le format de représentation n'est pas spécifique à l'action ou au mouvement de mon corps. L'encodage se fait par rapport aux objets extérieurs et est invariant par rapport à mon propre mouvement. Il s'agit d'un exemple d'encodage véridique des représentations.

Des expériences sur le rat et le singe ont permis de mettre en évidence l'utilisation du codage allocentrique chez les animaux. L'hippocampe, une structure neuronale d'une vieille partie du cortex qui joue un rôle important dans les processus émotionnels et dans la formation de mémoires nouvelles, joue probablement également un rôle important dans la construction de représentations allocentrées. On a pu montrer que certains neurones de l'hippocampe du rat, appelés cellules pyramidales, déchargent chaque fois que le rat passe par un endroit particulier d'une arène, quelle que soit son orientation ou le chemin parcouru entre deux passages⁹¹. Ces neurones couvrent un domaine spatial qui varie, mais qui ne dépasse pas quelques dizaines de centimètres de rayon. Chez le singe, on a retrouvé également dans l'hippocampe des neurones qui s'activent uniquement à partir de régions spécifiques de l'espace, mais avec la condition supplémentaire que la région spatiale soit significative pour l'animal (un lieu de protection, de fuite,

89 Cf. *ibid.*, p. 114.

90 *Ibid.*, p. 114.

91 Cf. *ibid.*, p. 139.

etc.)⁹². De nouveau, l'animal peut situer son mouvement dans un espace allocentré et l'encoder par rapport aux relations topographiques entre les repères de l'espace extérieur. Même si l'interprétation de ces expériences est très complexe et sujette à beaucoup de discussions, il semble clair que l'hippocampe est une structure d'intégration d'indices sensoriels multiples et qu'à partir d'une telle intégration, le cerveau peut construire des représentations indépendantes de la position du corps et du contexte d'action. Il se produit donc une abstraction progressive lorsqu'on franchit les niveaux successifs depuis le système vestibulaire jusqu'à l'hippocampe⁹³.

L'encodage allocentrique permet de construire des représentations de l'espace qui prennent comme point de référence des éléments de l'environnement et non des éléments du corps. Ce type d'encodage est apparu assez tardivement au cours de l'évolution, puisqu'il est lié au développement de structures d'intégration dans le cortex cérébral comme l'hippocampe. Il n'est pas aussi répandu que l'encodage égocentrique, l'encodage relatif aux mouvements de mon corps. Selon Berthoz, la plupart des animaux sont capables de réaliser un codage égocentrique, mais seuls les primates et l'homme sont vraiment capables d'utiliser le codage allocentrique⁹⁴.

Du point de vue du développement individuel également, l'encodage des représentations dans des formats spécifiques à l'action semble précéder l'encodage véridique. Considérons, par exemple, la réaction des enfants par rapport à des pentes dangereuses⁹⁵. Dans des expériences sur la représentation de l'espace chez l'enfant, on a placé des enfants (qui marchent et qui rampent) au-dessus de pentes ayant des degrés d'inclinaison différents. Les enfants qui marchent (quatorze mois dans l'expérience) étaient conscients du danger de pentes d'environ vingt degrés et plus. Confrontés à de telles pentes, soit ils refusaient de descendre, soit ils préféraient descendre en glissant. Les enfants rampeurs, par contre, ne reconnaissaient pas d'emblée les pentes dangereuses et devaient être rattrapés. A partir de là on a pu constater l'évolution intéressante suivante. Les enfants rampeurs peuvent, avec l'expérience, apprendre à éviter les pentes dangereuses de 20 degrés ou plus. Cependant, au point de transition, au moment où les enfants commencent à marcher, ce savoir semble tout d'un coup perdu. Les petits marcheurs doivent tout réapprendre sur les pentes dangereuses. Dans l'expérience, deux tiers des débutants marcheurs prenaient sans hésitation les

92 Cf. *ibid.*

93 Cf. *ibid.*, p. 140.

94 Cf. *ibid.*, p. 110.

95 Cf. THELEN E. et SMITH L., *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge (MA), 1994, p. 220.

pententes dangereuses, comme s'ils les rencontraient pour la première fois. La reconnaissance des pententes chez les enfants est donc relative au contexte de l'action et une représentation dans le contexte "rampeur" ne se transpose pas automatiquement au contexte "marcheur". La représentation de l'espace (ici les pententes) n'est donc pas une représentation de l'espace géométrique indépendant, mais une représentation relative aux contextes spécifiques d'action.

Les expériences de Berthoz et celles de la psychologie de l'enfant sur les représentations de l'espace suggèrent une image très différente de la représentation que l'image de la théorie représentationnelle classique. En effet, dans les exemples que nous avons étudiés, l'objectif spécifique du système représentationnel n'est pas de construire une image miroir ou une copie fiable de l'environnement, mais de préparer à l'action. Selon les cas, une telle représentation orientée vers l'action sera encodée dans un format plus ou moins général, utilisable dans des contextes divers d'action diverses, ou dans un format plus spécifique, valable uniquement pour telle ou telle action. A l'extrême, on trouve les situations où l'action et la représentation se confondent. Dans ces cas, la représentation de l'environnement est directement encodée sous la forme des opérations que le système peut exécuter sur cet environnement. Les représentations ne représentent rien d'autre que des possibilités d'action.

Pour illustrer ce cas extrême, à l'opposé des représentations véridiques, considérons le robot mobile construit par Maja Mataric au MIT⁹⁶. Ce robot implémente un modèle de fonctionnement du système de contrôle de mouvement du rat, d'inspiration neurobiologique. Un des éléments de fonctionnement de ce robot est un réseau de neurones dont les noeuds représentent une carte topographique de l'espace. Chaque noeud de ce réseau correspond à un endroit caractéristique de l'espace. Un tel endroit de l'espace est encodé sous la forme combinée du mouvement et des entrées sensorielles du robot à cet endroit (un couloir étroit, par exemple, sera encodé comme la combinaison d'un mouvement rectiligne et la perception de deux obstacles à faible distance latérale). La construction du réseau (par apprentissage associatif, par exemple) est telle que des noeuds adjacents sur le réseau correspondent à des endroits adjacents dans l'environnement. Supposons maintenant que le robot veuille se diriger vers un endroit particulier de l'environnement. La position courante du robot est indiquée par un noeud

96 Cf. MATARIC M., "Navigating with a Rat Brain : a Neurobiologically Inspired Model for Robot Spatial Representation", in MEYER J.-A. et WILSON S. (eds.), *From Animals to Animals I*, Cambridge (MA), 1991. On peut également trouver une description succincte du principe de fonctionnement de ce robot dans CLARK A., *Being There*, *op. cit.*, pp. 47-49.

activé et le noeud correspondant à l'endroit à atteindre s'active également. Le réseau calcule d'abord la plus courte distance entre la position courante et l'endroit à atteindre, en calculant la plus courte distance sur le réseau, entre les deux noeuds activés. Ensuite, comme les noeuds combinent de l'information sur le mouvement du robot et les entrées sensorielles à un certain endroit, le réseau peut lui-même coordonner le mouvement. Il suffit de suivre le chemin d'activation le plus court partant du noeud "position courante" au noeud "endroit à atteindre" et d'exécuter pour chaque noeud le mouvement encodé. Au lieu d'avoir, comme dans les systèmes classiques, une carte pour la représentation de l'environnement et un centre séparé de planification du mouvement, la carte peut elle-même agir comme engendrant l'action. Utiliser la carte de l'environnement et engendrer le mouvement se fait par le même processus.

Le cas des représentations directement opératoires sur l'environnement constitue la forme la plus simple de la boucle de traitement de l'information qui commence avec la sélection des paramètres pertinents dans l'environnement et aboutit à l'action. Dans cette boucle de traitement d'information, les représentations ne définissent pas une image du monde, comme dans la théorie du réalisme représentationnel, mais elles définissent une boucle d'action, qui va des représentations partielles jusqu'aux représentations opératoires, en passant éventuellement par des représentations plus générales ou non contextuelles. Une telle considération écologique des représentations permet de construire une image plus réaliste des entrées et des sorties des systèmes cognitifs que dans les modèles connexionnistes et les modèles cognitivistes classiques. Cependant, la théorie écologique des représentations que nous avons étudiée ici reste une écologie subjective. Le rôle de l'environnement est spécifié entièrement par l'interprétation de cet environnement à partir du milieu de vie de l'organisme, c'est-à-dire selon son comportement et son style de vie social. Mais l'environnement n'est pas seulement une ressource passive de données interprétables ou manipulables par un sujet, il connaît une organisation propre et des lois de transformation actives. Dans la section suivante, nous verrons comment la dynamique de l'environnement peut être utilisée dans la résolution de certaines tâches cognitives, limitant de cette façon encore davantage le poids des représentations et des computations explicites.

4.2. La détermination écologique

L'habileté et la précision de nage des grands animaux aquatiques comme les dauphins ou les thons est tout à fait étonnante. Elle surpasse de loin tout ce que l'on a pu réaliser dans nos sous-marins ou dans d'autres

systèmes nautiques artificiels. En plus, la vitesse de propulsion de ces animaux est restée longtemps un mystère. On estime que le dauphin, par exemple, n'a simplement pas la force de nager à la vitesse à laquelle on le voit nager. Pour expliquer ce fait étonnant deux chercheurs en dynamique des fluides, les frères Triantafyllou, ont proposé l'hypothèse que les animaux aquatiques exploitent de façon active la dynamique de leur environnement⁹⁷.

Figure 1.9. La propulsion d'un poisson par des tourbillons créés à partir de sa propre intervention. Un premier mouvement de queue crée un vortex dans un sens (1). Le retour de la queue crée un deuxième vortex dans le sens opposé (2,3). Un courant fort de propulsion vers l'avant et une force latérale résiduelle résulte de la rencontre des deux vortex (4) (source : TRIANTAFYLLOU et TRIANTAFYLLOU, 1995).

Pour accélérer, ces poissons profitent de courants rapides le long des rochers ou prennent leur élan à partir de tourbillons provoqués par le refoulement de l'eau au contact d'un obstacle. Toutefois, l'utilisation habile de la dynamique de l'environnement ne s'arrête pas là. Les grands animaux aquatiques interviennent eux-mêmes activement dans la création de cette dynamique. Ils créent des tourbillons ou des gradients de pression (par

⁹⁷ TRIANTAFYLLOU M. et TRIANTAFYLLOU G., "An Efficient Swimming Machine", in *Scientific American*, 272 (1995), pp. 64-71.

exemple par le mouvement de leur queue) et exploitent ensuite les courants d'entraînement créés par leur propre intervention (cf. la figure 1.9). Aidés de cette façon par un défilé continu de tourbillons ou de remous, ils sont capables de produire des départs rapides et des changements brusques de direction.

Cet exemple de la nage des animaux aquatiques, emprunté à Andy Clark⁹⁸, illustre bien l'interaction dynamique des systèmes biologiques avec l'environnement : ces systèmes profitent en permanence de la structure fine de leur environnement⁹⁹. Comme nous le verrons dans ce paragraphe, les systèmes cognitifs également, à l'image des animaux aquatiques, exploitent activement les ressources dynamiques de leur environnement. Pour le montrer, nous partirons de quelques thèmes développés par Andy Clark dans *Being There*, en les articulant à la problématique de l'inscription contextuelle et corporelle de la cognition qui nous préoccupe ici.

La théorie écologique des représentations du paragraphe précédent nous offre déjà un exemple d'un comportement interactif avec l'environnement. Au lieu de considérer seulement ce qui se passe à l'intérieur du système, comme c'est le cas dans les modèles connexionnistes et cognitivistes classiques, on s'intéresse à l'interaction du système avec son environnement. Dans ce cadre, nous avons vu que les systèmes cognitifs interviennent de façon active dans la sélection des entrées et dans la négociation, selon le contexte, des sorties du système. Cependant, dans la perspective de l'écologie subjective, l'environnement reste passif et n'intervient pas activement dans la résolution des problèmes. La prise en compte en plus du caractère dynamique de l'environnement permettra de simplifier la résolution de nombreux problèmes et de limiter la charge des représentations et des computations explicites. Au lieu d'un système dynamique, le système de traitement d'information et un environnement passif, on aura un couplage entre deux systèmes dynamiques en interaction.

Dans le cas des systèmes cognitifs humains, l'analyse interactive nous permettra d'aller encore un peu plus loin. Les systèmes cognitifs humains ont non seulement la capacité d'utiliser à bon escient les vertus dynamiques de leur environnement, mais ils peuvent également transformer de façon durable cet environnement. Par l'élaboration de configurations d'artefacts ou de *dispositifs* à partir de l'environnement, on peut diffuser des connaissances acquises et des savoir-faire pratiques. Ainsi, on peut situer les problèmes

98 Cf. CLARK A., *Being There*, *op. cit.*, p. 219.

99 Le domaine de l'écologie des milieux aquatiques est tout à fait passionnant de ce point de vue. Pour d'autres exemples d'exploitation dynamique des propriétés du monde aquatique, cf. VOGEL S., *Life in Moving Fluids*, Princeton, 1983.

cognitifs individuels dans des réseaux complexes de contraintes instituées. Finalement, à partir de l'interaction réciproque entre le système cognitif, le corps et l'environnement naturel et institué, peuvent émerger des entités globales d'un ordre supérieur. Comme nous le verrons, si les différentes composantes en interaction s'influencent de façon permanente, par des interactions réciproques, il devient difficile de dire quelle composante constitue "le système" et quelle composante constitue "l'environnement". Dans ces cas de dépendance forte entre sous-systèmes, il devient plus commode d'étudier l'évolution à partir de l'entité globale émergente.

Ainsi on peut distinguer différents niveaux d'intervention active de l'environnement dans la résolution des problèmes cognitifs. D'abord, il y a la possibilité du couplage entre systèmes dynamiques en interaction, puis la transformation active de l'environnement et, finalement, l'immersion complète du système cognitif dans son contexte opératoire. Pour désigner ces différents rapports où l'environnement joue un rôle déterminant dans la cognition, nous parlerons, à la différence de l'écologie subjective où le rapport à l'environnement est simplement un rapport d'*interprétation*, d'une *détermination* écologique¹⁰⁰.

4.2.1. Limiter la charge représentationnelle : le couplage dynamique

4.2.1.1. Les modèles cybernétiques

Pour décrire l'interaction avec un environnement dynamique, Clark utilise la métaphore parlante de l'envoi et du retour de balle (*catch and toss interactions*)¹⁰¹. "Le monde envoie des entrées vers le cerveau, qui les rattrape et renvoie des actions en retour. Les actions peuvent modifier ou simplifier les computations subséquentes, en forçant le monde à renvoyer des entrées plus faciles à l'usage et ainsi de suite"¹⁰². De nombreux exemples d'envois-retours donnés par Clark sont d'inspiration cybernétique. A l'instar des systèmes de régulation et de contrôle de Norbert Wiener¹⁰³, les systèmes cognitifs peuvent, par des boucles de rétroaction avec l'environnement, adapter leur comportement de façon continue pour réaliser un certain but. Pour attraper une balle, par exemple, on peut proposer deux stratégies. Soit on calcule, en utilisant une représentation interne de l'environnement, la

100 Nous devons l'usage de ce terme à une suggestion de Andy Clark dans *Being There* (cf. CLARK A., *Being There*, op. cit., p. 103).

101 Cf. *ibid.*, p. 105–106.

102 *Ibid.*, p. 106.

103 WIENER N., *Cybernetics or Control and Communication in the Animal and the Machine*, New York, 1948.

position d'atterrissage de la balle et on court ensuite vers cet endroit. Soit on regarde la balle en l'air et on court en contrôlant le changement de l'angle de la tangente d'élévation du point de mire vers la balle, pour maintenir ce changement constant¹⁰⁴. Un simple mécanisme de rétroaction permet d'ajuster ma vitesse de course au but poursuivi. Si l'angle d'élévation augmente trop vite il faut ralentir et s'il diminue il faut accélérer. Si l'on utilise ce simple truc, on attrapera la balle avant qu'elle ne tombe par terre. Il semble que les joueurs de base-ball, par exemple, utilisent ce type d'interaction directe avec l'environnement, plutôt que la simulation mentale de la trajectoire¹⁰⁵.

Cependant, les exemples d'envois-retours discutés par Clark ne se limitent pas aux simples boucles de régulation. L'on peut relever, dans le matériel expérimental présenté par Clark, une double originalité. D'abord, dans les systèmes cybernétiques classiques, les variables de contrôle se situent dans le système et l'environnement sert simplement de source d'informations pour ajuster le comportement par rapport au but. Toutefois, on peut inverser la donne. Comme nous le verrons, beaucoup de nos actions se font sous contrôle de l'environnement et ne supposent pas un contrôle centralisé au niveau du système cognitif. Ensuite, à côté du contrôle environnemental, Clark considère aussi des boucles de rétroaction élargies. En effet, au lieu d'avoir un renvoi immédiat d'information, on peut aussi construire des boucles de rétroaction plus larges, incluant la dynamique propre de l'environnement. De cette façon les envois-retours peuvent prendre la forme d'une interaction entre systèmes dynamiques qui interviennent conjointement dans la régulation de l'action.

On se demandera peut-être quelle peut être la pertinence des modèles cybernétiques pour les sciences cognitives. Est-ce qu'on n'a pas tout simplement changé d'objet d'étude, de la cognition vers la régulation ? Et, s'il n'y a plus de représentations, ni de computations, peut-on encore parler d'opérations *cognitives* ? Tout d'abord, les modèles cybernétiques permettent de mieux circonscrire l'usage — plus limité qu'on ne le croyait — des représentations. Parfois, comme dans le cas de la balle à attraper, un couplage dynamique avec l'environnement permet de résoudre un problème que classiquement on résout de façon computationnelle. Le choix entre une solution cybernétique et une solution représentationnelle, comme dans le cas du choix entre les représentations véridiques et les représentations partielles, dépendra du problème auquel on est confronté. Mais l'intérêt des modèles cybernétiques va plus loin. En effet dans la plupart des situations on

104 Cf. CLARK A., *Being There*, *op. cit.*, p. 27.

105 Cf. *ibid.*

rencontre un mélange de solutions cybernétiques et de solutions représentationnelles. Dans les systèmes cognitifs complexes, comme les nôtres, les interactions avec l'environnement et les computations sur des représentations coopèrent de façon permanente.

4.2.1.2. Les systèmes décentralisés sous contrôle de l'environnement

Retournons maintenant aux exemples de couplage dynamique avec l'environnement donnés par Clark. On peut se donner une première idée de la classe des systèmes cybernétiques sous contrôle environnemental en considérant de nouveau le petit robot à six pattes construit par Ken Espenschied¹⁰⁶, illustré à la figure 1.10.

Figure 1.10. Le robot à six pattes construit par Espenschied et son équipe. Chaque patte est une unité sensori-motrice autonome. La cohérence des six unités est assurée par l'interaction avec un environnement opératoire commun (source : CLARK A., 1997, p. 16).

Nous avons vu qu'il s'agit d'un petit robot qui peut marcher sur une surface plate et surmonter des obstacles fixes sur cette surface. Le principe de construction de ce robot est simple. Chaque patte est une unité sensori-motrice entièrement autonome, munie d'un petit réseau de neurones qui permet de représenter la fréquence de mouvement de la patte et la réponse (soit avancer, reculer ou lever la patte) de cette patte par rapport à l'information de l'environnement. Il n'y a pas de communication entre les pattes autre que celle de pouvoir inhiber ou encourager le mouvement des

¹⁰⁶ Cf. QUINN R. et ESPENSCHIED K., *op. cit.* On trouve également une discussion de cet exemple dans CLARK A., *Being There, op. cit.*, pp. 15–16.

pattes voisines et il n'y a pas non plus de contrôle central de mouvement. Au cours du mouvement, quand une certaine patte rencontre un obstacle, elle va tenter de le surmonter. Si elle échoue, elle passe la main à une autre patte jusqu'au moment où sa position a suffisamment changé et qu'elle décide de reprendre l'essai. Ce robot, en fait une collection de six sous-systèmes indépendants dont la cohérence est assurée par l'environnement opératoire commun, a été testé dans des conditions de mouvement réel. Il permet de reproduire des mouvements de marche qui ressemblent aux mouvements intelligents et adaptés des systèmes biologiques.

D'autres robots ont été construits sur base des mêmes principes. On peut illustrer ces principes à partir d'un robot extrêmement simple, construit par Rodney Brooks et dont la tâche est d'avancer sans se heurter aux objets de l'environnement¹⁰⁷. Au lieu d'être décomposé en un sous-système de perception et un sous-système de mouvement sous contrôle central, ce robot possède deux sous-systèmes sensori-moteurs indépendants (l'un pour déclencher le mouvement en absence d'obstacle, l'autre pour les arrêts d'urgence). La cohérence des sous-systèmes est de nouveau assurée par l'environnement. L'architecture de tels robots qui agissent sous contrôle environnemental peut être caractérisée, selon Brooks, de la façon suivante¹⁰⁸:

1. Décomposition du système en sous-systèmes produisant une activité. Chaque sous-système qui produit une activité connecte à sa manière la sensation à l'action.
2. Architecture décentralisée. L'interaction entre les sous-systèmes se limite à l'inhibition ou l'encouragement de l'activité d'un sous-système voisin. Il n'y a pas de transmission explicite d'information d'un système à l'autre.

107 Cf. VARELA F., THOMPSON E. et ROSCH E., *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, op. cit., p. 285. Remarquons que Varela *et al.* ne mettent pas l'accent dans leur discussion de la nouvelle robotique sur le contrôle par l'environnement. Ils considèrent surtout l'aspect pratico-social des représentations utilisées, comme dans l'écologie subjective. En négligeant le rôle de l'environnement dans l'établissement de la cohérence entre les différents sous-systèmes, ils ne parviennent pas à combiner l'écologie subjective avec la détermination écologique. Le problème est éludé en faisant figurer cette cohérence comme une simple émergence par auto-organisation : "A partir du chaos local de leurs interactions émerge, aux yeux d'un observateur, un schème cohérent de comportement" (*ibid.*, p. 285). Cependant, le schème d'action n'est pas seulement cohérent aux yeux d'un observateur (c'est-à-dire d'un hypothétique jugement réfléchissant), mais également de façon effective dans l'environnement (c'est-à-dire résultant de la détermination environnementale).

108 CLARK A., *Being There*, op. cit., p. 13.

La nouvelle robotique, telle qu'elle est décrite par Brooks, propose donc une image de la cognition profondément différente de l'image classique. Le recours à l'utilisation de représentations explicites est limité encore d'un cran en éliminant la nécessité d'un contrôle ou d'une planification centrale des sous-systèmes cognitifs. Selon Brooks, ce contrôle central rend le système trop lent. Il exige en effet une double conversion de données. D'abord, il y a la conversion de l'information sensorielle dans un code symbolique unique ou une représentation suffisamment générale pour que l'unité centrale de planification puisse l'utiliser. Ensuite, il faut convertir la sortie de l'unité centrale dans un format de représentation compatible avec les réponses motrices du système. Le contrôle central rend une réponse rapide et directe quasi impossible¹⁰⁹.

Les systèmes décentralisés sous contrôle de l'environnement nous donnent un premier exemple de la détermination écologique. L'interaction avec un même environnement d'un système cognitif composé de différents sous-systèmes indépendants assure la cohérence d'action de ce système sans avoir recours à des représentations internes. Le contrôle décentralisé n'exclut évidemment pas tout recours à la représentation. Il est approprié, comme nous l'avons vu, quand il faut produire des réponses rapides et quand le système doit accomplir des tâches spécifiques pour lesquelles on n'a pas besoin d'un modèle détaillé et centralisé du monde. Pour des comportements plus complexes il faudra combiner la capacité de réponse rapide en interaction directe avec l'environnement et des représentations explicites. Comme le dit Berthoz, opposant la solution cybernétique à une solution représentationnelle plus élaborée : « Pour devenir champion de ski, il ne suffit pas de traiter en permanence les informations des sens et corriger la trajectoire ; il faut dérouler la course dans son esprit, en prédire les étapes et l'état des capteurs sensoriels, entrevoir les solutions possibles de chaque erreur, faire des paris et prendre des décisions avant que le geste soit fait¹¹⁰. Le schème représentationnel plus élaboré auquel Berthoz fait référence dans ce passage est celui de la simulation de mouvement par le cerveau (hypothèses, prédictions, ...). Par exemple, en mémorisant la copie d'une commande motrice (la copie d'efférence), le cerveau est capable de prédire le signal de retour qui suivrait l'exécution de cette commande — pour un certain état de l'environnement — et d'adapter d'avance son comportement en tenant compte de cette prédiction¹¹¹. Si les signaux de rétroaction de

109 Cf. *ibid.*, p. 21.

110 BERTHOZ A., *op. cit.*, p. 7.

111 Cf. *ibid.*, pp. 12–13.

l'environnement sont trop complexes, un tel calcul par simulation anticipative peut offrir une meilleure solution.

4.2.1.3. *Les boucles de rétroaction élargies*

A côté des systèmes décentralisés sous contrôle de l'environnement, Clark considère une deuxième série de systèmes cybernétiques élargis, que nous avons regroupés sous le concept de systèmes à boucle de rétroaction élargie. Dans ces systèmes, le contrôle de l'action se situe de nouveau du côté du système cognitif, mais les sous-systèmes réalisant l'action ne se trouvent plus tous du côté mental, comme c'était le cas des agents autonomes de la nouvelle robotique. Ils sont distribués à travers le système cognitif, le corps et l'environnement. Un exemple de telles boucles élargies est donné par le bras de robot auquel nous avons fait référence dans l'introduction à ce paragraphe et dont la tâche est d'emboîter des pièces finement ajustées. Une solution mentaliste à ce problème calcule pour chaque essai, et de façon explicite, la position précise dans l'espace que le bras de robot doit atteindre. La solution distribuée combine l'élasticité d'un joint en caoutchouc, auquel on a attaché le bras, avec quelques calculs élémentaires de position spatiale. En poussant, les pièces se glissent et se faufilent en bonne position. L'ajustement par rétroaction de l'action — si nécessaire — ne se fait qu'après avoir exploité la dynamique de mouvement propre au bras de robot. L'envoi-retour n'est pas immédiat et c'est en ce sens que l'on peut parler d'une boucle de rétroaction élargie.

Des expériences en psychologie du développement suggèrent que pour la coordination du mouvement des membres du corps, le cerveau exploite également de telles boucles de rétroaction élargies. Par exemple, dans un modèle développé par Esther Thelen et Linda Smith¹¹², la coordination du mouvement des membres s'appuie sur des dynamiques corporelles intrinsèques, comme l'élasticité des muscles et la raideur des membres, et des éléments environnementaux, comme la force gravitationnelle. Comme pour le bras de robot, il n'y a pas une exécution de commandes motrices internes qui spécifient une trajectoire détaillée de mouvement, mais simplement la modulation de différents facteurs qui interviennent dans le mouvement. Si les dynamiques corporelles jouent un rôle si important, la modulation devrait être différente d'une personne à l'autre. Dans une expérimentation de Thelen et Smith sur le développement du comportement de préhension chez les enfants, on constate en effet une telle différence d'un enfant à l'autre. "Un enfant, Gabriel, était très actif par nature. Il générait des

112 Cf. THELEN E. et SMITH L., *op. cit.*, pp. 86–88.

mouvements rapides du bras. Pour lui, la difficulté était de convertir les battements en un mouvement dirigé. Pour ce faire, il devait apprendre à contracter les muscles au moment où son bras s'approchait d'un objet cible, de façon à ralentir le battement et à permettre un contact approprié. Hannah, au contraire, était calme par nature. Les mouvements qu'elle produisait étaient lents et peu larges. Son problème n'était pas de contrôler les battements, mais de générer suffisamment d'élan pour vaincre la gravité¹¹³. Dans ces expériences, chaque enfant partait d'un mélange de dynamiques corporelles intrinsèques différent. Même si la tâche — atteindre un objet — était la même pour tous les enfants, le problème de coordination à résoudre pour chaque enfant se présentait sous une forme différente. L'apprentissage des nouveaux comportements est donc autant conditionné par le développement de schèmes mentaux de contrôle que par les dynamiques corporelles intrinsèques.

Un autre exemple frappant de l'importance des facteurs corporels et environnementaux est fourni par les études de Thelen et Smith sur l'apprentissage de la marche chez le jeune enfant¹¹⁴. Ils ont montré que la disparition du réflexe de marche présent à la naissance ne tient pas à l'inhibition mentale de ce réflexe, comme on le supposait dans les théories représentationnelles classiques, mais tient principalement à l'augmentation du poids de la jambe ! Si on met par exemple un enfant de trois mois dans un bain, de façon à réduire le poids effectif de la jambe, le réflexe réapparaît. En fait, le réflexe ne disparaît jamais réellement au cours de la première année, mais doit se combiner avec des facteurs corporels, comme le poids de la jambe ; des facteurs mécaniques, comme l'élasticité des muscles ; des facteurs environnementaux, l'eau du bain par exemple ; et des éléments cognitifs, comme le développement de la capacité de modulation des différents facteurs. La coordination de la marche est également une coopération complexe de dynamiques corporelles et contextuelles, et d'éléments cognitifs.

Le couplage simple, la cognition décentralisée et les boucles de rétroaction élargies nous donnent trois types d'interaction d'un système cognitif avec un environnement dynamique. Au lieu de complexifier les tâches, cette interaction permet souvent de simplifier le système et de le limiter son rôle à celui de partenaire dans un processus à différents joueurs. De ce point de vue, les imperfections de l'environnement, le bruit, le frottement, le caractère approximatif des interactions, ne sont pas toujours des obstacles supplémentaires à surmonter par le système, mais coopèrent

113 CLARK A., *Being There*, *op. cit.*, p. 44.

114 Cf. THELEN E. et SMITH L., *op. cit.*, pp. 11–12.

dans la résolution du problème. Comme dans le cas du joint élastique du bras de robot, elles interviennent comme facteur d'atténuation des déviations par rapport à la position d'équilibre¹¹⁵.

L'élargissement des systèmes cybernétiques classiques à des tâches plus complexes nous a également permis d'illustrer deux façons différentes de limiter la charge représentationnelle des systèmes cognitifs. D'abord, dans le cas des robots autonomes, des sous-systèmes multiples sont coordonnés par l'interaction avec un même environnement opératoire. Le couplage dynamique des sous-systèmes avec un même environnement remplace l'agent de coordination central classique. Ensuite, nous avons vu que certains systèmes cognitifs s'appuient sur les vertus dynamiques de leur environnement, le corps et/ou le monde. Au lieu d'être le pôle centralisateur des sources informationnelles de l'environnement, les représentations interviennent comme partenaire dans le jeu des forces distribuées entre le système cognitif, le corps et le monde externe.

Dans les exemples d'interaction dynamique avec l'environnement considérés ici, l'environnement est l'environnement présent dans les conditions écologiques normales, que ce soit le corps de l'enfant ou l'environnement physique des agents autonomes, par exemple. L'on pourrait évidemment aussi penser à manipuler artificiellement cet environnement pour pouvoir réaliser des tâches cognitives particulièrement difficiles. En reprenant l'exemple du skieur de Berthoz, l'on peut se demander si, à côté de son choix mental des stratégies globales par la simulation interne, il ne s'appuie pas également en grande partie sur un environnement artificiellement préparé ? Que serait le champion de ski sans son recours aux propriétés dynamiques de son équipement de haute technologie ou l'interaction directe — à l'image des agents autonomes — avec les indices (couleurs, drapeaux, ...) de la piste qui lui permettent de décomposer sa tâche et de faire un contrôle décentralisé rapide de ses mouvements ? L'exploit du champion de ski s'appuie autant sur l'interaction dynamique avec un environnement artificiellement préparé, que sur les représentations explicites

115 Même dans les applications industrielles, les imperfections physiques peuvent jouer ce rôle positif. Ainsi, dans la régulation des machines à vapeur, les systèmes de régulation modernes sont tellement précis qu'ils produisent des oscillations de vitesse autour de la position d'équilibre. Les détecteurs électroniques réagissent de façon trop précise à des changements de vitesse et le système d'équilibrage surcorrigé en permanence. Dans les systèmes anciens plus rudimentaires, on avait pas ce problème. Le système de régulation était entièrement mécanique et le frottement entre les pièces stabilisait la réponse du système (cf. SMITHERS T., "Why Better Robots Make It Harder", in CLIFF D. *et al.* (eds), *From Animals to Animats III*, Cambridge (MA), 1994, pp. 64–66 ; cité dans CLARK A., *Being There*, *op. cit.*, p. 95).

du mouvement à effectuer. Comme nous le verrons dans la section suivante, la structuration active de l'environnement permet d'étendre les possibilités de couplage dynamique des systèmes cognitifs, du couplage avec l'environnement naturel et le corps, au couplage avec des dispositifs artificiellement préparés.

4.2.2. La structuration active de l'environnement : les dispositifs

De façon lapidaire, Clark résume les limitations des modèles connexionnistes par la phrase “bon en Frisbee, mauvais en logique”¹¹⁶. La considération des couplages interactifs nous laisse une impression similaire : bon en Base-ball, en ski, contrôle moteur, etc., mais mauvais en planification ou calcul séquentiel. Le couplage interactif permet de construire des systèmes à réponse rapide dans des environnements réels, mais qu'en est-il par exemple des tâches cognitives classiques de l'intelligence artificielle comme le calcul logique ou mathématique et la résolution séquentielle de problèmes ?

4.2.2.1. Les dispositifs

Un exemple tiré de la “bible” connexionniste, les deux tomes du *Parallel Distributed Processing*¹¹⁷, montre, de façon élégante et presque triviale, comment la résolution d'un problème de calcul numérique peut néanmoins se faire à partir d'un couplage simple avec l'environnement externe, en s'appuyant sur la manipulation simple de quelques dispositifs artificiels. D'abord, pour modéliser des calculs arithmétiques, comme la multiplication, à partir du modèle des réseaux de neurones, on peut partir de quelques calculs simples. Dans *Parallel Distributed Processing*, David Rumelhart et ses collaborateurs¹¹⁸ montrent comment la solution d'un problème simple de multiplication, comme $7 \times 7 = 49$, peut être produite par un système de catégorisation prototypique implémenté dans un réseau de neurones. Mais comment faire pour des opérations arithmétiques plus élaborées ? Pour résoudre la multiplication 7222×9422 , la plupart d'entre

116 CLARK A., *Being There*, op. cit., p. 60.

117 RUMELHART D. et MCCLELLAND J. (eds.), *Parallel Distributed Processing : Explorations in the Microstructure of Cognition. Volume 1 : Foundations, Volume 2 : Psychological and Biological Models*, Cambridge (MA), 1986. Cet exemple est cité dans CLARK A., *Being There*, op. cit., p. 60.

118 Cf. RUMELHART D., SMOLENSKY P., MCCLELLAND J. et HINTON G., “Schemata and Sequential Thought Processes”, in RUMELHART D. et MCCLELLAND J. (eds.), *Parallel Distributed Processing*, Cambridge (MA), 1986.

nous ont recours à des manipulations sur papier. De cette façon, on réduit l'opération complexe à quelques opérations simples, en commençant par 2×2 , tout en mémorisant les résultats des opérations intermédiaires sur un dispositif externe, le papier en l'occurrence. En couplant des résolutions habituelles de problèmes simples à la mémorisation externe des résultats intermédiaires, on arrive à la solution. De cette façon, comme le dit Clark, "l'environnement externe devient une extension cruciale de la cognition"¹¹⁹.

L'utilisation de ressources externes artificielles, illustrée dans cet exemple simple, aura des conséquences importantes. D'après Clark, la structuration active et l'utilisation massive de ces structures par les systèmes cognitifs humains permet de "dissiper le raisonnement" : "diffuser la connaissance acquise et le savoir-faire pratique à travers des structures sociales complexes et réduire la charge des cerveaux individuels en positionnant ces cerveaux dans des réseaux complexes de contraintes linguistiques, sociales, politiques et institutionnelles"¹²⁰. Clark désigne ces ressources artificielles par le terme de *scaffold* (échafaudage), terme que nous avons décidé de rendre en français par le terme de *dispositif*, terme qui suggère également la construction de moyens externes sur lesquels on peut s'appuyer dans la réalisation de certaines activités ou tâches cognitives¹²¹.

Une des sources de la recherche sur l'intervention des dispositifs dans la cognition citée par Clark est le travail du psychologue soviétique Lev Vygotski¹²². Vygotski s'est intéressé spécialement au rôle du langage dans la coordination de nos activités au quotidien. Plus généralement, Vygotski a souligné l'importance de structures externes dans notre compréhension et dans le traitement des données de l'environnement. Dans le développement de l'enfant, par exemple, on peut distinguer des périodes de développement dites proximales, périodes pendant lesquelles le développement dépend de façon cruciale de l'intervention externe des parents, par exemple quand

119 *Ibid.*, p. 46.

120 CLARK A., *Being There, op. cit.*, p. 180.

121 En utilisant le terme *dispositif* nous voulons également indiquer l'affinité des recherches en sciences cognitives sur l'inscription corporelle et contextuelle de la cognition avec les recherches en sciences humaines sur la médiation des savoirs. Cf. par exemple *Dispositifs et Médiation des Savoirs*, colloque tenu à Louvain-la-Neuve, le 24 et 25 avril 1998, département de communication, UCL (Les résultats de ce colloque ont été publiés sous la forme d'un numéro spécial de la revue française *Hermès*, cf. JACQUINOT-DELAUNAY G. et MONNOYER L. (coord.), "Le dispositif. Entre usage et concept", in *Hermès*, 25 (1999), pp. 5-244). Dans le cadre de ces recherches plus générales il faudrait parler ici de *dispositifs externes* de médiation des savoirs.

122 Cf. par exemple VYGOTSKI L., *Pensée et langage*, trad. du Russe par Françoise Sève, Paris, 1985 (éd. originale en 1934).

l'enfant apprend à marcher¹²³. D'après Vygotski, le langage peut lui aussi fonctionner comme appui externe au développement de l'enfant¹²⁴. Pour des tâches pratiques complexes, par exemple, on peut guider verbalement les mouvements de l'enfant. De cette façon, en s'appuyant sur l'expertise d'un adulte qui le guide pas à pas, l'enfant peut accomplir des tâches qui nécessitent beaucoup d'astuces, comme nouer ses lacets. Ensuite, quand l'enfant est seul, il peut répéter la séquence du dialogue intérieurement et l'utiliser pour guider son propre mouvement, focaliser son attention et se prémunir contre les erreurs habituelles. Dans cet exemple, le langage n'est pas simplement un outil pour véhiculer des contenus de pensée d'une personne à l'autre, mais intervient directement dans le contrôle de l'action¹²⁵.

Des expériences récentes en psychologie du développement soutiennent les thèses de Vygotski. Une équipe de psychologues a observé et enregistré les paroles d'un groupe d'enfants entre cinq et dix ans¹²⁶. Ils ont trouvé que la plupart des paroles privées (c'est-à-dire adressées uniquement à soi-même, vocalisées ou non) concernaient le contrôle de l'action et que l'incidence de telles paroles augmentait quand l'enfant était seul et tentait de résoudre quelque tâche difficile. A partir de ces expériences et d'autres, on a pu constater que le langage est un outil cognitif crucial, qui permet à l'enfant de se focaliser rapidement sur les aspects pertinents d'une situation et qui permet de l'accompagner dans l'exécution de certaines tâches difficiles.

Remarquons que dans les expériences sur le langage comme outil de contrôle de l'action, ce qui est premier, c'est le langage comme dispositif externe d'aide et pas le monologue intérieur. D'abord l'adulte explique à l'enfant les étapes à suivre, chaque fois qu'il accompagne l'enfant dans une certaine tâche. Ensuite seulement, l'enfant développe, en simulant le dialogue avec l'adulte, le langage intérieur. Dans le cas des dispositifs d'aide de calcul on observe un phénomène similaire. D'abord on apprend les calculs longs sur papier et seulement ensuite on internalise ces règles de calcul pour apprendre à faire les calculs longs de façon purement mentale. D'après ces exemples, la manipulation des symboles langagiers et la manipulation des structures mathématiques n'implique pas un substrat computationnel organisé également comme un langage symbolique. Tout ce qu'il faut, c'est une capacité de simulation de ressources extérieures, c'est-à-dire la

123 CLARK A., *Being There*, *op. cit.*, p. 47.

124 Cf. *ibid.*, p. 195.

125 Cf. spécialement le chapitre 7 de *Pensée et langage* (VYGOTSKI L., *op. cit.*, pp. 319–386).

126 Cf. BERK L. et GARVIN R., "Development of Private Speech among Low-income Appalachian Children", in *Developmental Psychology*, 20 (1984), pp. 271–286 ; cité dans CLARK A., *op. cit.*, p. 195.

possibilité d'internaliser des compétences cognitives qui trouvent leur origine dans des manipulations du monde extérieur.

La capacité de manipulation de symboles peut donc être compatible avec une architecture connexionniste. Par exemple, Rumelhart *et al.* argumentent que la simulation mentale ne suppose rien d'autre qu'un réseau de neurones capable de catégorisation et d'association par similarité, mais qui opère sur le domaine spécifique d'un type spécial de représentation externe¹²⁷. Ainsi l'utilisation effective de diagrammes de Venn peut optimiser un réseau de neurones spécialisé dans le calcul, ce qui permet ensuite de simuler, par le même réseau de neurones, des séquences de manipulation sur des diagrammes de Venn imaginés¹²⁸. Quoi qu'il en soit du mécanisme cognitif exact à la base de la manipulation des symboles par le cerveau, il semble clair que le rôle des dispositifs externes dans nos capacités linguistiques et mathématiques a longtemps été sous-estimé. Clark donne encore d'autres exemples du rôle des dispositifs externes dans nos capacités cognitives, comme le rôle des règles de l'environnement économique dans la prise de décision dans une entreprise ou, plus simplement, l'organisation physique d'un bureau. Dans certaines situations, le rôle des dispositifs peut même devenir tellement important que l'élaboration d'une solution d'un certain problème est presque entièrement guidée par ou sous contrôle de l'environnement. L'exemple des entreprises dans un environnement économique compétitif illustre bien une telle situation. Dans d'autres situations, comme le vote aux élections ou la planification de nos vacances, l'environnement externe intervient beaucoup moins et l'influence de nos opinions et de nos projets personnels pourra être plus grande dans le processus de délibération cognitive.

4.2.2.2. La conscience réflexive

La structuration active de notre environnement permet donc d'étendre nos capacités cognitives au-delà du contrôle efficace du mouvement ou de la catégorisation prototypique, vers l'intelligence symbolique, sociale et institutionnelle propre aux êtres humains. En fait, le milieu de vie de l'être humain est caractérisé par une véritable explosion de dispositifs artificiels. L'*homo faber* apparaît non seulement comme un bon fabricant d'outils d'aide aux tâches manuelles, mais surtout un bon fabricant d'outils de facilitation de tâches cognitives. Intuitivement on peut relier cette explosion de dispositifs en tout genre à la réflexivité de l'esprit humain. C'est parce qu'on peut

127 Cf. RUMELHART D. *et al.*, *op. cit.*, pp. 44-48.

128 *Ibid.*

réfléchir sur nos propres tâches cognitives que l'on peut structurer notre monde activement de façon à promouvoir, soutenir et étendre nos acquis cognitifs. Le passage par la conscience réflexive permet de projeter dans notre environnement de nouvelles médiations objectives pour la raison opératoire.

L'amélioration réflexive de nos capacités suppose évidemment des finalités ou des normes par rapport auxquelles on s'oriente pour l'évaluation des capacités. Ces finalités ou ces normes sont une projection de ce qui pourrait ou devrait être. Elles ne sont jamais entièrement de l'ordre de l'effectif, étant toujours partiellement réalisées et partiellement à l'état de projet. En abordant le thème de la conscience réflexive, nous sortons donc partiellement du domaine d'étude des sciences cognitives, qui est celui de l'étude de la pensée comme opération effective, et qui est précisément fondé sur l'abstraction de l'horizon intentionnel plus large des finalités et des normes qui peuvent motiver tel ou tel problème cognitif.

Néanmoins, se pose la question de savoir comment la conscience réflexive peut émerger à partir des opérations mentales effectives. Dans ce contexte, Clark formule la conjecture que l'émergence de la conscience réflexive s'appuie sur le développement des structures langagières¹²⁹ : "Dès qu'on formule une pensée sous forme langagière (sur papier ou de façon vocale), cela devient un objet pour nous-mêmes et les autres. Et comme objet, nous pouvons formuler des pensées sur cette pensée exprimée"¹³⁰. L'émergence de la réflexivité à partir du langage est rendue possible parce que le langage nous fournit des structures de signification relativement stables, qui peuvent à leur tour faire l'objet d'une pensée de second ordre. "Pour pouvoir fonctionner comme un instrument efficace de communication, le langage public a été structuré dans un code qui convient bien à l'échange intersubjectif dans lequel des idées sont présentées, affinées et critiquées"¹³¹. En formulant nos pensées sous forme langagière, soit dans un contexte de communication soit dans un contexte de contrôle de l'action, ces pensées deviennent également disponibles pour tout un ensemble d'opérations et de manipulations d'ordre supérieur, comme par exemple l'amélioration réflexive ou la critique intersubjective¹³².

129 Cf. CLARK A., *Being There*, *op. cit.*, p. 208.

130 *Ibid.*, p. 209.

131 *Ibid.*, p. 210.

132 Clark considère seulement la conscience réflexive qui intervient dans l'évaluation et l'amélioration de nos activités par rapport à des normes ou des finalités données. Pour l'illustration du rôle des dispositifs externes dans la cognition et l'émergence des facultés supérieures en s'appuyant sur ces dispositifs ceci peut nous suffire. Signalons cependant qu'à côté de ce premier niveau de réflexivité, on peut distinguer un deuxième niveau, où

La structuration active de notre environnement permet donc une augmentation considérable de nos compétences cognitives. Nous avons considéré respectivement la manipulation itérative de longues chaînes de symboles mathématiques, la diffusion des savoir-faire et des connaissances acquises dans des réseaux sociaux et institutionnels complexes, la communication verbale de séquences d'activités et finalement, dans la mesure où l'on suit la conjecture de Clark, l'émergence, à partir des structures langagières, de la capacité de réflexion sur nos propres tâches cognitives. Il est évidemment difficile de démêler l'ensemble de ces différents processus computationnels et les différents types de couplages chaque fois impliqués avec le corps, l'environnement naturel et l'environnement artificiel. Le point de vue adopté ici était, pour la clarté de l'exposition, génétique : de l'intelligence associative des réseaux de neurones à l'intelligence symbolique et l'intelligence des réseaux sociaux et institutionnels. Et ensuite de l'intelligence symbolique à la conscience réflexive. Dans la réalité les différentes capacités s'entremêlent et se renforcent pour produire une dynamique d'interaction entre l'esprit, le corps et l'environnement naturel et institué.

4.2.3. Les systèmes émergents

L'analyse interactive nous a permis d'aller au-delà des frontières étroites des systèmes cognitifs pour nous intéresser à la relation entre les contenus du système cognitif et l'environnement dans lequel il opère. L'image qui en ressort est un système cognitif étroitement couplé à un environnement à la fois naturel et institué. Cependant, parfois ce couplage peut devenir tellement fort qu'il devient difficile de dire quelle composante est 'le système' et quelle composante est 'l'environnement'. Dans ce cas, il faut compléter l'analyse interactive avec une analyse au niveau de l'entité globale émergente.

le travail de la critique porte non sur les activités, mais sur les normes et les finalités elles-mêmes. Ce travail critique de second ordre prendra la forme d'une interrogation en retour sur la teneur intentionnelle originelle de tel ou tel *telos* (norme, finalité) qui traverse notre monde vécu (cf. LADRIÈRE J., *L'éthique dans l'univers de la rationalité*, Namur, 1997, pp. 43-66). A côté de la conscience réflexive qui est liée au langage, on peut également, à un niveau infra-langagier ou anté-prédicatif, attribuer une certaine capacité réflexive au corps. Dans l'exemple de la spatialité du corps propre que nous avons discuté dans le deuxième paragraphe, cette réflexivité prend la forme d'une assomption intentionnelle des mouvements effectifs du corps dans un projet moteur d'ensemble (le partage de l'espace selon le haut, le bas, le lointain et le proche). Déjà au niveau du corps, il y a une projection intentionnelle de sens (*Sinngebung*) (cf. MERLEAU-PONTY M., *Phénoménologie de la perception*, op. cit., p. 166).

4.2.3.1. L'émergence directe et l'émergence indirecte

Il existe évidemment de multiples formes d'émergence d'un tout à partir des parties. Ici, nous nous intéresserons à deux formes particulières d'émergence, l'émergence directe d'une configuration globale à partir de l'interaction entre les parties d'un système et l'émergence indirecte d'une configuration globale à partir de l'interaction entre un système et son environnement.

Un exemple classique d'émergence directe est fourni par les boucles de convection dans un liquide¹³³. Quand on chauffe un liquide, on observe d'abord une simple diffusion de la chaleur du bas vers le haut. Si on continue à chauffer, au moment où le liquide dépasse une certaine température critique, il se réorganise et on observe un mouvement global d'ensemble circulaire, appelé mouvement de convection. A partir des interactions locales entre les molécules de liquide, émerge un ordre global qui n'est pas présent au niveau des parties.

Pour bien comprendre l'origine de ce type d'émergence, que l'on retrouve aussi dans les systèmes cognitifs, il est intéressant de regarder de plus près le fonctionnement des boucles de convection. Aux basses températures, on observe une simple diffusion de la chaleur. Le mécanisme physique de cette diffusion est simple. La chaleur appliquée dans le fond du récipient se traduit sur le plan microscopique par une plus grande vibration des molécules. Par les chocs, les molécules moins énergétiques, dans la couche de liquide immédiatement supérieure, se mettent à vibrer également et reçoivent une partie de l'énergie de la région plus chaude. Ainsi, de proche en proche, à partir des couches chaudes du fond jusqu'aux couches plus froides à la surface, la chaleur se diffuse à travers le récipient. Dans le cas de la diffusion, il n'y a pas de mouvement d'ensemble, mais simplement une transmission par interaction locale de l'énergie appliquée. Cependant, quand on dépasse un certain seuil de température, appelé température critique, la différence de densité entre les parties chaudes dans le fond (moins denses, plus "légères") et les parties moins chaudes à la surface du récipient (plus denses, plus "lourdes") devient tellement grande que le liquide commence à circuler dans le récipient. Les parties moins denses montent et les parties lourdes descendent de façon à produire un mouvement circulaire, les boucles de convection. Le nombre de telles boucles augmente avec la température et à la surface de contact entre les boucles se forment des hexagones réguliers¹³⁴. Les boucles de convection sont des formes collectives qui

133 Cf. par exemple PLATTEN J. et LEGROS J., *Convection in Liquids*, Berlin, 1984.

134 *Ibid.*

émergent sans que la forme de ces boucles, ni des hexagones réguliers à la surface, soit spécifiée au niveau des interactions locales entre les molécules. La seule variable de contrôle est la température et celle-ci ne contient aucune information ou prescription de la configuration émergente.

Malgré le fait que cela ne soit pas apparent dès la première approche, cette émergence est le résultat d'un couplage dynamique réciproque permanent entre deux sous-systèmes du système convectif, un système de diffusion thermique et un système mécanique de gradients de densité (et donc aussi de pression). On peut s'en rendre compte en considérant le comportement du système convectif quand un des deux sous-systèmes est absent. D'un côté, dans le système convectif en dessous du seuil critique de température, les effets de densité ne jouent pas et il n'y a pas de cohérence d'ensemble. On a la simple diffusion de proche en proche de la chaleur dans le récipient. De l'autre côté, dans un récipient où il n'y a pas de diffusion (en absence de chocs entre les molécules), quand on chauffe le récipient, l'énergie appliquée ne rencontre pas d'obstacles et se propage à la manière d'une onde mécanique à travers le récipient¹³⁵. La cohérence est totale et les interactions locales se propagent directement à travers l'ensemble du récipient. Dans la situation intermédiaire, qui est celle des boucles de convection dans le liquide, les deux sous-systèmes sont présents et s'influencent réciproquement. Les gradients de densité ou de pression sont suffisamment forts pour mettre le liquide en mouvement, mais, dès qu'il y a un mouvement, il est diffusé par les chocs thermiques et, au lieu d'avoir des ondes thermiques simples du fond vers la surface du récipient, les molécules sont entraînées dans le mouvement global circulaire convectif. Les boucles de convection sont donc le résultat du couplage interactif entre un effet de distribution statistique, la diffusion thermique, et un effet mécanique de plus longue portée, les ondes de densité ou de pression.

En considérant le couplage dynamique des deux sous-systèmes, on comprend mieux aussi pourquoi la variable de contrôle, la température, ne détermine pas entièrement la configuration globale émergente. La température contrôle l'énergie qui va aux deux sous-systèmes et, si on connaît les caractéristiques physiques du système (la densité, le poids moléculaire), on peut prédire, à partir de la température, le type de mouvement que l'on observera (ici le nombre de boucles de convection pour chaque température). Mais le mouvement précis que l'on observera, par

135 C'est le cas des ondes de chaleur que l'on observe à très basse température et sur des échantillons microscopiques (c'est à dire en absence de chocs). Pour un aperçu détaillé, cf. DREYER W. et STRUCHTRUP H., "Heat Pulse Experiments Revisited", in *Continuum Mechanics and Thermodynamics*, 5 (1993), pp. 3-50.

exemple des boucles qui tournent vers la gauche ou vers la droite, n'est pas contrôlé par la température, puisqu'elle est appliquée de façon uniforme sur le fond du récipient. La direction que prendront les boucles de convection dépend de la dynamique d'interaction locale entre le sous-système mécanique et le sous-système diffusif, dynamique qui dépend de l'histoire des influences réciproques et donc ultimement de fluctuations microscopiques qui ont initié cette dynamique et sur lesquelles on n'a pas de maîtrise.

L'exemple des boucles de convection illustre bien le concept d'émergence directe ou de l'émergence par auto-organisation. A partir des interactions locales sur lesquelles on n'a pas de maîtrise détaillée et des effets d'emballement entre deux sous-systèmes émergent des configurations d'ensemble. D'autres exemples d'émergence directe sont le vol en V d'un groupe d'oiseaux à partir des interactions locales entre paires d'oiseaux ou le comportement de fuite dans une certaine direction d'une foule en danger, sans que cette direction soit choisie à l'avance ou entièrement déterminée par les circonstances environnementales. Dans le cas des systèmes cognitifs, un exemple d'émergence directe est donné par l'émergence d'une représentation globale à partir de l'interaction locale entre des éléments individuels d'un système cognitif. Ce type d'émergence est présent dans certains réseaux de neurones que nous avons étudiés, les réseaux de Kohonen et de Héroult-Jutten, dont la programmation se fait par auto-organisation.

Cependant, dans tous ces exemples, l'interaction avec l'environnement n'intervient pas de façon fondamentale dans l'émergence. L'émergence est liée au couplage entre sous-systèmes internes et l'emballement de ces systèmes à partir de fluctuations qui ne sont pas sous le contrôle de variables environnementales. Dans le cas des systèmes cognitifs, en considérant le couplage interactif des systèmes cognitifs avec l'environnement, un autre type d'émergence pourra apparaître, l'émergence indirecte. Dans cette forme d'émergence, l'environnement n'est plus une source passive d'entrées du système, mais intervient activement dans l'auto-organisation. L'émergence est toujours déterminée par le couplage interactif réciproque entre sous-systèmes, mais cette fois-ci les sous-systèmes interagissent avec les données de l'environnement.

Considérons de nouveau un des exemples favoris de Clark, un robot dans un environnement opératoire réel, à l'image des agents autonomes de Brooks. Imaginons que le robot doive se positionner entre deux pôles pour pouvoir se recharger¹³⁶. Le poste de chargement est indiqué par une source lumineuse. Une solution non-émérgentiste serait d'équiper le robot avec des

136 Cf. STEELS L., "The Artificial Life Roots of Artificial Intelligence", in *Artificial Life*, 1 (1994), pp. 75-110 ; cité dans CLARK A., *Being There*, *op. cit.*, p. 108.

détecteurs qui mesurent sa distance par rapport aux pôles et d'un système computationnel qui permette de calculer une trajectoire entre les deux pôles. Une solution alternative, émergentiste, serait d'équiper le robot de deux routines simples indépendantes, une pour se détourner d'un objet quand il le touche et une pour s'approcher en zigzag de n'importe quelle source lumineuse. A partir de ces deux règles simples, le comportement global, se positionner entre les deux pôles, pourra progressivement émerger. Le robot est attiré par la lumière et avance en zigzag vers celle-ci. Quand il touche un pôle, il se retire et essaie de nouveau à partir d'une nouvelle position. Après quelques essais il trouvera une position d'équilibre entre les pôles, une position près de la source lumineuse, mais qui ne touche aucun des deux pôles. A partir de l'interaction dynamique entre deux sous-systèmes émerge donc une configuration globale qui n'est pas entièrement déterminée par les données de départ (les deux sous-routines), mais qui dépend également de l'histoire de l'interaction dynamique entre ces deux systèmes. L'émergence est indirecte en ce que l'interaction réciproque entre les deux sous-systèmes est médiatisée par leur interaction commune avec l'environnement¹³⁷.

4.2.3.2. Les variables globales

Le couplage réciproque entre systèmes ou sous-systèmes cognitifs, à travers leur interaction avec un environnement opératoire commun ou par un couplage direct, pourra faire émerger des comportements qui ne sont pas spécifiés au niveau des sous-systèmes, mais qui dépendent de leur histoire commune d'interaction. Dans ces cas, la description des sous-systèmes individuels ne suffira plus pour expliquer le comportement global émergent. Dans les situations de couplage fort, il faut définir une variable supplémentaire sur l'ensemble du système émergent, qui tient compte de l'interaction réciproque continue entre les deux sous-systèmes. Par exemple, dans le cas des boucles de convection, cette variable globale concernera le nombre ou la taille des boucles de convection pour différentes valeurs de la température appliquée au système. Dans le cas du robot, la variable globale déterminera les différents points d'équilibre entre les pôles, en fonction de la position des pôles et de la source lumineuse. De cette façon, en faisant abstraction de l'histoire détaillée des interactions réciproques, on obtient une

137 Le cas du petit robot à six pattes d'Espendschied que l'on a discuté dans ce paragraphe est également un exemple d'émergence indirecte. Plus précisément, selon l'histoire des interactions, deux positions d'équilibre peuvent émerger. Soit les pattes bougent de façon alternative, les unes après les autres en parcourant l'ensemble des pattes. Soit la patte arrière et la patte de devant d'un côté du robot bougent de façon synchrone avec la patte du milieu de l'autre côté (cf. *ibid.*, p. 17).

description du comportement global du système en fonction des valeurs des variables de contrôle (la température, la disposition de l'environnement, ...).

La description du système émergent en termes de variables globales fait abstraction des détails fins de l'évolution du système. En fait on ne considère que les positions d'équilibre qui résultent de l'interaction réciproque des sous-systèmes et l'on ne s'intéresse pas à la période transitoire qui mène à telle ou telle solution. Cette abstraction est justifiée, puisque les détails fins de l'évolution précise des sous-systèmes n'interviennent pas dans la détermination qualitative des configurations globales. Cependant, si l'on s'intéresse également au détail fin de la solution, par exemple si l'on veut savoir si la boucle de convection tournera vers la gauche ou vers la droite, il faudra combiner l'analyse au niveau du système émergent avec une analyse au niveau du comportement détaillé des sous-systèmes en interaction, à partir de leurs valeurs initiales. Dans le cas des boucles de convection, une telle analyse est impossible à cause du nombre astronomique de molécules en interaction. Cependant, dans des systèmes de taille intermédiaire une telle analyse combinée pourrait prendre tout son sens.

Un exemple d'un tel système de taille intermédiaire est fourni par les analyses de l'anthropologue et psychologue cognitiviste Edwin Hutchins de la navigation d'un voilier en pleine mer¹³⁸. Dans le modèle développé par Hutchins, l'émergence à partir de règles locales d'interaction et l'adaptation contextuelle locale coopèrent pour maîtriser l'art complexe de navigation du voilier. Dans ses recherches, Hutchins a constaté que, malgré la présence d'un but et des objectifs globaux à poursuivre, les membres d'une équipe de navigation ne suivent pas un plan ou un scénario détaillé de comportement. Si jamais un tel plan formel existe, il s'avère très incomplet ou, souvent, impraticable. Plutôt qu'un scénario précis, les membres de l'équipage suivent seulement quelques règles locales habituelles d'interaction avec d'autres membres, selon la fonction qu'ils occupent, et c'est l'histoire concrète de la dynamique des interactions qui détermine le comportement global émergent. On peut se faire une idée de ces règles locales à partir des exemples donnés par Hutchins : "Chaque membre de l'équipage doit seulement savoir ce qu'il doit faire quand certaines conditions se produisent dans l'environnement. Un examen des tâches de l'équipage montre que beaucoup sont spécifiées sous la forme "fais X quand Y". Voici quelques exemples de procédures :

- a. "Mesure la profondeur de l'eau chaque fois que le pont le demande.

138 Cf. HUTCHINS E., *Cognition in the Wild*, Cambridge (MA), 1995 ; cité dans CLARK A., *Being There*, *op. cit.*, p. 76.

- b. Enregistre l'heure et la profondeur mesurée chaque fois qu'une telle mesure est envoyée au pont.
- c. Mesure et rapporte les directions et les orientations du bateau chaque fois que celui qui tient le livre de loch le demande¹³⁹.

Les règles locales, après apprentissage, peuvent s'effectuer par de simples mécanismes d'association prototypique. Ensuite, par ajustement réciproque, les réponses des membres individuels sont rendues plus adéquates par rapport aux tâches à effectuer. Par exemple, quand un membre de l'équipage constate une lacune, il le communique à une personne compétente. Celle-ci exécute une tâche de correction, ce qui a des conséquences plus loin dans la chaîne des interactions. De cette façon, l'ensemble des interactions produit l'adaptation désirée¹⁴⁰.

En fonction des variables de contrôle, les instructions du capitaine et l'environnement physique de navigation, différentes séquences d'action et différentes configurations de traitement d'information par l'équipage peuvent émerger. À côté de la décomposition fonctionnelle de la tâche en règles ou mécanismes, il est donc utile d'étudier une variable globale qui reflète la dynamique des interactions. De cette façon, le capitaine pourra prévoir les effets de ses instructions au niveau de la configuration d'ensemble qui émerge de tel ou de tel enchaînement d'interactions. L'analyse interactive locale du traitement d'information effectuée par chaque membre, qui nous permet de réfléchir sur les ajustements locaux, doit donc être complétée par une analyse au niveau du comportement global émergent.

L'émergence indirecte de configurations globales d'un système ou de différents systèmes cognitifs en interaction avec leur environnement montre les limites de l'analyse interactive. Dans ces cas de couplage fort entre systèmes ou sous-systèmes, il faut définir des variables globales supplémentaires pour décrire l'évolution d'ensemble. L'intérêt d'une telle description holistique dépend évidemment du point de vue auquel on veut se placer pour étudier le système. Parfois, les considérations globales ne sont pas pertinentes et il est plus intéressant d'étudier l'adaptation contextuelle ou les couplages locaux des sous-systèmes, comme dans l'analyse interactive. Dans d'autres cas, l'analyse interactive ne pourra même pas suffire et il faut "ouvrir la boîte noire" pour décrire le mécanisme interne de telle ou de telle opération de traitement d'information, s'intéresser aux processus d'apprentissage ou aux imperfections d'un sous-système par exemple. Ces différents niveaux d'analyse correspondent au parcours que nous avons effectué dans ce

139 HUTCHINS E., *op. cit.*, p. 199.

140 CLARK A., *Being There, op. cit.*, p. 78.

chapitre, de l'analyse computationnelle interne, en passant par l'analyse interactive aux systèmes émergents. En considérant l'interdépendance de ces différents niveaux, comme dans le cas de la navigation du bateau, émerge une image de la cognition transformée par les différents élargissements conceptuels que nous avons considérés et où les représentations, le corps et l'environnement sont profondément entrelacés.

Chapitre 2

Les conséquences méthodologiques Du fonctionnalisme à la modélisation contextuelle

1. Introduction

Dans le premier chapitre nous avons donné un aperçu des modèles théoriques en sciences cognitives, à partir du fil conducteur de l'opérationnalisation de l'intentionnalité. Cet aperçu nous a conduit des modèles cognitivistes classiques jusqu'aux évolutions récentes qui remettent à l'avant-plan la problématique du contexte des modèles computationnels, à la fois dans sa dimension sémantique et fonctionnelle. En un premier temps, nous sommes partis d'une définition préliminaire du concept de computation et nous avons étudié plusieurs modèles théoriques qui donnent un contenu plus spécifique à ce concept. Ceci nous a amené à caractériser différentes *formes* d'intelligence, l'intelligence symbolique des modèles cognitivistes classiques, l'intelligence habituelle des modèles connexionnistes et, finalement, l'intelligence exploratoire des modèles mixtes. En un deuxième temps, nous avons vu que la simple considération des mécanismes computationnels ne permet pas de rendre compte adéquatement de l'intentionnalité. Pour modéliser les représentations et les fonctions, il faut croiser l'étude des mécanismes computationnels avec l'étude du contexte sémantique de fonctionnement (l'écologie subjective) et l'étude des ressources dynamiques de ce contexte (la détermination écologique).

L'étude des processus cognitifs par des méthodes similaires à celles utilisées dans les sciences de la nature est manifestement très féconde. Toutefois, une telle approche naturalisée de la cognition ne risque-t-elle pas de remplacer l'étude des activités cognitives par un simple mécanisme de l'esprit ? De plus, même si la modélisation de l'intentionnalité garde une certaine spécificité au sein du champ des sciences cognitives — par l'intérêt pour le contexte sémantique et fonctionnel —, on peut se demander si la naturalisation ne réduit pas également les opérations de contextualisation à de simples mécanismes causaux. Ces questions ont déjà été abordées de façon indirecte au premier chapitre, en argumentant en faveur d'une théorie

de la cognition qui tiennent compte à la fois de la dimension des mécanismes, des fonctions et des représentations. Toutefois, de nombreux auteurs contestent le bien-fondé d'une telle approche *large* de la cognition. C'est pourquoi nous aborderons, dans ce deuxième chapitre, la question du statut méthodologique de l'opération de modélisation envisagée. Plus particulièrement, nous étudierons la possibilité d'une méthode spécifique de naturalisation qui convient à la modélisation de l'intentionnalité.

En un premier temps, nous regarderons en quoi les sciences cognitives en général s'opposent à une reprise mécaniste de la naturalisation. Plus particulièrement, les réflexions méthodologiques de David Marr, dans le cadre de sa théorie de la vision, nous semblent très éclairantes dans ce débat. Les réflexions de Marr permettent de préciser la spécificité de l'approche computationnelle de la cognition, à la fois par rapport à une approche purement syntaxique ou algorithmique et par rapport à une réduction biologique, à partir des neurosciences. Toutefois, comme nous le verrons, dans la théorie de Marr la question de la nature spécifique du niveau computationnel est évacuée par une théorie fonctionnaliste des rapports entre l'esprit et le cerveau. C'est pourquoi nous serons amenés à approfondir, en un deuxième temps, les perspectives ouvertes par la théorie de Marr, à partir de la question de l'articulation entre le niveau computationnel et le niveau de l'organisation biologique. Cette question nous permettra de caractériser de façon plus adéquate le niveau de détermination causale spécifique qui intervient dans l'étude de la cognition, en particulier en étudiant de plus près le concept de système téléonomique.

Ensuite, en prenant acte de la critique du mécanisme qui caractérise l'entreprise des sciences cognitives, nous retournerons dans le deuxième paragraphe au problème particulier de la modélisation de l'intentionnalité, afin d'étudier la possibilité d'une méthode de modélisation spécifique, qui prenne également en compte l'anticipation du contexte d'effectuation des opérations cognitives à l'intérieur des modèles. Nous confronterons cette perspective de modélisation avec deux critiques importantes de la modélisation de l'intentionnalité en sciences cognitives : la critique fonctionnaliste d'une part et la critique de la clôture opérationnelle de l'autre. Ces critiques portent respectivement sur le problème de la modélisation des représentations et des fonctions. La première critique, la critique fonctionnaliste, considère que la question de la définition des représentations n'introduit pas de dimension supplémentaire d'analyse par rapport à la méthodologie computationnelle définie par Marr. Comme nous le verrons, le fonctionnalisme, en s'appuyant sur une hypothèse de modularité des unités de traitement d'information dans le cerveau, évacue la possibilité d'une prise

en compte de la définition par le système de son propre contexte de significations. La deuxième critique, la critique de la clôture opérationnelle, prend acte de l'importance de la dimension représentationnelle, mais propose une hypothèse d'architecture du système nerveux qui exclut tout couplage fonctionnel avec l'environnement. Une telle perspective réduit l'anticipation contextuelle à une autostabilisation et ne peut rendre compte de la stabilisation du pouvoir d'anticipation en interaction avec l'environnement. Nous verrons qu'une architecture cognitive ouverte à l'interaction avec le contexte, mais néanmoins capable d'une anticipation contextuelle à l'intérieur de cette interaction, permet d'envisager une pratique de modélisation qui prend en compte la dimension représentationnelle et fonctionnelle de l'intentionnalité.

2. Le défi de la cognition : modéliser les computations

Dans le premier chapitre de ce travail, nous sommes partis d'une définition préliminaire du concept de computation qui nous a permis de donner une caractérisation des systèmes cognitifs dans les termes du concept de traitement d'information. Dans ce deuxième chapitre, dans le contexte des discussions autour du mécanisme et de la naturalisation, nous voulons approfondir la caractérisation de la méthodologie des sciences cognitives et indiquer en premier lieu en quoi cette méthodologie se distingue d'autres approches, comme les neurosciences ou l'intelligence artificielle.

Pour étudier le rapport entre les différentes disciplines qui interviennent dans l'étude de la cognition, nous partirons d'abord du traitement de cette question dans le cadre de la théorie de la vision de David Marr. Les réflexions méthodologiques de Marr ont été très influentes dans le développement des sciences cognitives et nous donneront une première définition des concepts clefs en sciences cognitives, que l'on approfondira ensuite. Dans un deuxième temps, nous dépasserons le point de vue de Marr pour nous intéresser à la question de l'organisation biologique.

2.1. La théorie computationnelle de David Marr et ses critiques

Pour caractériser la méthodologie spécifique des sciences cognitives, nous partirons des distinctions entre computation, algorithme et implémentation introduites par David Marr dans sa théorie de la vision¹. Pour bien comprendre ces distinctions, nous devons d'abord les situer par

1 MARR D., *Vision*, San Francisco, 1982.

rapport au contexte d'élaboration de la théorie de Marr. Ensuite nous présenterons le détail des concepts introduits par Marr et discuterons quelques critiques de l'approche computationnelle, critiques qui pourraient miner l'entreprise des sciences cognitives. Finalement nous reprendrons le concept de computation comme concept central des sciences cognitives, tout en élargissant le point de vue de Marr, encore trop lié à une hypothèse d'optimum ou de perfection du système cognitif. Pour la présentation de la théorie de Marr, nous nous baserons sur le compte rendu critique de sa théorie dans l'article de Patricia Kitcher *Marr's computational theory of vision*².

2.1.1. La computation comme traitement d'information

En introduisant les distinctions entre computation, algorithme et implémentation, Marr entendait critiquer aussi bien les approches purement neurophysiologiques de la vision (niveau de l'implémentation) que les approches purement syntaxiques (niveau algorithmique). Plus particulièrement, Marr voulait se distancier par rapport à la méthodologie du programme de recherche naissant de la vision artificielle³. Dans ce programme on supposait que la vision procède en deux étapes : d'abord la segmentation de l'image en traits (*features*) et ensuite le regroupement des traits dans des catégories en vue de la reconnaissance de l'objet. Ce scénario plausible se basait, entre autres, sur les fameuses expériences de Barlow, et de Hubel et Wiesel, d'enregistrement isolé des activités neuronales. Par ces expériences, on avait découvert des neurones "détecteurs" au niveau du cortex visuel, c'est-à-dire des neurones s'activant uniquement quand un trait spécifique (couleur rouge, courbure anguleuse, forme quadratique, etc.) se présente dans le champ visuel⁴. Cependant, malgré le succès initial de ce programme, on n'a jamais découvert de détecteurs de traits de plus haut niveau, correspondant au regroupement hypothétique des traits en un tout signifiant. En plus, l'implémentation sur ordinateur d'un programme de segmentation d'images en traits s'est révélée très difficile à cause de la complexité des images réelles⁵.

2 Cf. KITCHER P., "Marr's Computational Theory of Vision", in *Philosophy of Science*, 55 (1988), pp. 1–24.

3 Cf. KITCHER P., *op. cit.*, p. 2.

4 Cf. MARR D., *op. cit.*, pp. 12–14.

5 Comme le suggère le réseau de reconnaissance des visages de Cottrell que nous avons étudié au premier chapitre, il n'y a probablement pas une segmentation de l'image en traits, mais une segmentation en composantes holistiques ou prototypes (cf. CHURCHLAND P.M., *The Engine of Reason, the Seat of the Soul*, *op. cit.*, p. 47).

A côté des problèmes empiriques, le programme de recherche de la vision artificielle se confrontait également à de sérieux problèmes d'ordre théorique. Marr soulève deux grandes objections aux présuppositions théoriques de ce programme de recherche, la première portant sur l'hypothèse de détection de traits par des neurones isolés et la deuxième sur la recherche d'un algorithme de reconnaissance. L'étude de ces deux objections nous permettra de mieux comprendre la nécessité d'introduire un niveau d'étude spécifique par rapport au niveau neurophysiologique (la détection des traits) et par rapport au niveau algorithmique (l'algorithme de reconnaissance), niveau qui sera le niveau du traitement d'information, appelé par Marr le niveau computationnel.

La première objection de Marr peut être illustrée à partir du problème appelé dans la littérature le problème du "neurone grand-mère". Supposons que la vision procède effectivement par segmentation de l'information visuelle en traits et que, à partir de là, le système visuel recompose l'image en vue de la reconnaissance de l'objet. Alors, hypothétiquement, lors de la reconnaissance, le neurone ayant comme entrée l'ensemble des traits correspondants à l'objet devrait décharger. Le "neurone grand-mère", c'est le neurone qui, selon cette théorie, décharge quand je reconnais ma grand-mère⁶. Supposons que l'on ait trouvé le "neurone grand-mère". Supposons même que l'on puisse retracer la chaîne des stimuli nerveux de la rétine jusqu'à ce neurone. Que saurait-on alors de la façon dont nous reconnaissons des grand-mères ? Quelle information le système visuel utilise-t-il ? Comment combine-t-il cette information ? On pourrait voir le chemin des stimuli nerveux, mais on ne saurait pas pourquoi les neurones empruntent ce chemin. En bref, on ne saurait pas ce que le système visuel est en train de faire quand nous reconnaissons notre grand-mère⁷.

Une objection similaire peut être soulevée par rapport à la recherche de l'algorithme de reconnaissance. Comme nous avons vu, les premiers programmes de vision artificielle se basaient sur une théorie de la vision en deux étapes : d'abord, la détection des traits et, ensuite, un algorithme de reconnaissance. Le but de l'algorithme est de simuler les performances du système visuel humain. En partant d'une approximation raisonnable de l'information disponible sur la rétine, le programme doit fournir en sortie approximativement les mêmes classifications perceptives que l'être humain pour un certain domaine d'application⁸. Mais un tel algorithme ne nous apprend de nouveau rien sur ce que le système visuel est en train de faire. Il

6 Cf. KITCHER P., *op. cit.*, pp. 2-3.

7 Cf. *ibid.*, p. 3.

8 Cf. *ibid.*, p. 3.

fournit seulement la structure logique ou syntaxique reliant l'entrée à la sortie du système visuel. De plus, on peut construire de nombreux algorithmes qui produisent des opérations de transformation entrée-sortie équivalentes, sans savoir lequel est effectivement utilisé par le système visuel.

Ni la détection des neurones qui s'activent lors d'une reconnaissance, ni la syntaxe des opérations du système visuel ne permettent de comprendre ce qu'est la vision. Ce qu'il faut comprendre, c'est la fonction de traitement d'information réalisée par le système visuel. On doit pouvoir spécifier la fonction de transformation des informations de l'environnement présentes au niveau de la rétine jusqu'aux représentations de l'environnement objectivement visibles pour nous⁹. Le concept d'algorithme est donc à relier avec le concept de fonction de traitement d'information. A la différence des approches syntaxiques — qui utilisent aussi une forme de computation, mais sur des symboles formels — les contenus traités par le système cognitif ne sont pas simplement syntaxiques, mais ont une signification par rapport à la *fonction* computationnelle.

Le concept de computation, entendu comme fonction de transformation d'informations, permet de donner une première définition du niveau de détermination propre à la pensée. A côté du niveau computationnel, la théorie de Marr donne aussi une définition des deux autres niveaux, le niveau algorithmique et le niveau neurophysiologique. Selon Marr, une théorie de la vision doit pouvoir rendre compte des trois niveaux et de leur interdépendance.

Le niveau computationnel caractérise le traitement d'information qui est effectué par chaque sous-système du système visuel. Au niveau algorithmique, on choisit le format de représentation des entrées et des sorties d'un sous-système particulier et l'on construit un algorithme qui calcule la fonction de transformation entrée - sortie¹⁰. La théorie computationnelle précise quelle fonction est calculée. Par exemple, pour calculer une fonction de reconnaissance de formes, on peut calculer les angles et les courbures des objets dans le champ visuel, et l'algorithme fournit un moyen pour réaliser ce calcul, par exemple la méthode de Newton d'approximation de la tangente d'une courbure. Le point important à noter dans le rapport entre computation et algorithme est qu'une même fonction peut toujours être calculée par différents algorithmes¹¹. Le niveau computationnel ne détermine donc pas

9 Cf. *ibid.*, p. 4.

10 Cf. *ibid.*, p. 10.

11 On a par exemple le choix entre le calcul exact d'une fonction ou le calcul d'une valeur approximative (suffisamment proche de la valeur exacte). Cette dernière peut être

complètement la théorie au niveau algorithmique. Néanmoins, la construction d'algorithmes peut fournir des indices intéressants sur la nature de la computation que le système est en train d'effectuer.

Entre le niveau de l'algorithme et le niveau de l'implémentation on rencontre une relation semblable. Un algorithme peut être implémenté dans différents matériels (*hardwares*), différents câblages neuro-anatomiques, et ne détermine donc pas complètement le niveau neurophysiologique¹². Mais, de l'autre côté, la théorie neurophysiologique peut donner des indications intéressantes sur les algorithmes physiologiquement possibles ou impossibles¹³. En résumant l'argumentation de Marr, on peut dire que le rapport entre les niveaux est un rapport de conditionnement réciproque. Aucun niveau ne détermine tous les autres, mais l'étude d'un niveau peut fournir des éléments intéressants pour construire la théorie à un autre niveau.

A partir de ces réflexions méthodologiques sur les niveaux d'étude du système visuel, nous pouvons tirer au moins deux enseignements. Premièrement, il y a un double niveau d'analyse des opérations cognitives, le niveau syntaxique (l'opération computationnelle) et le niveau des fonctions de traitement d'information (la fonction computationnelle). Deuxièmement, le rapport entre les opérations cognitives et l'implémentation neurophysiologique n'est pas univoque. A une opération donnée peuvent correspondre différentes structures neurophysiologiques qui implémentent cette opération.

2.1.2. Les critiques de l'approche computationnelle

Les concepts et la méthodologie de recherche introduits par Marr ont été extrêmement influents. Evidemment, sa théorie a été remaniée après lui et elle a été étendue à d'autres champs d'étude comme le langage, la mémorisation ou le raisonnement. Néanmoins les distinctions fondamentales entre fonction computationnelle, algorithme et implémentation sont toujours utilisées et permettent de circonscrire de façon relativement claire la méthodologie spécifique à l'oeuvre dans les sciences cognitives.

Certains auteurs contestent la nécessité de s'intéresser au niveau computationnel. Parmi ces auteurs, il y en a d'abord qui considèrent que c'est uniquement au niveau de la neurophysiologie que l'on pourra construire une

obtenue en calculant les premiers termes d'une série de Taylor ou encore par une méthode itérative de descente de gradient.

12 Ce point tout à fait familier pour l'informaticien peut être illustré en considérant une simple équivalence logique entre deux opérations : (p ou q) équivaut à (non [non p et non q]). Si on a un circuit capable de construire les opérateurs 'ou', 'et' et 'non' alors on pourra construire deux câblages équivalents.

13 Cf. KITCHER P., *op. cit.*, p. 10.

véritable science de la cognition. Les Churchland, par exemple, considèrent que sur base du progrès en neurosciences les concepts trop flous des sciences cognitives seront amenés à disparaître et seront remplacés par les concepts plus rigoureux des neurosciences¹⁴. Un autre opposant de l'entreprise computationnelle, J. Searle, considère que seule une naturalisation neuro-physiologique de l'intentionnalité nous permettra de comprendre de quelle façon la pensée peut émerger du cerveau¹⁵. Malgré le fait que les neurosciences sont tout à fait précieuses dans les recherches en sciences cognitives, comme nous l'avons déjà affirmé ci-dessus, il est clair qu'une étude des structures neuronales ne sera jamais suffisante pour connaître les fonctions computationnelles réalisées par le système cognitif. Celles-ci n'apparaissent qu'au niveau de l'étude du traitement de l'information.

Un deuxième groupe d'auteurs, dans lesquels on peut ranger par exemple Jerry Fodor ou Daniel Dennett, considèrent que l'étude du traitement de l'information doit se limiter au niveau syntaxique. La méthodologie prônée par ces auteurs est une forme de reconstruction rationnelle¹⁶. Elle s'inspire de la linguistique formelle de Chomsky, qui, dans

14 Cf. par exemple CHURCHLAND P.M. et CHURCHLAND P.S., "Stalking the Wild Epistemic Engine", *Noûs*, 17 (1983), pp. 5–18.

15 Cf. SEARLE J., *L'intentionnalité*, *op. cit.*, p. 322. Searle distingue néanmoins un niveau d'analyse au-dessus de l'approche neuroscientifique, l'analyse logique de l'intentionnalité du point de vue de la première personne. Cependant, il nous semble qu'on ne peut pas "naturaliser" d'emblée à partir des résultats d'une analyse purement logique vers l'implémentation neurophysiologique. L'analyse logique peut prescrire ce qui est possible (non-contradictoire), mais ne dit pas ce qui, parmi ce qui est possible, se réalise effectivement. De ce point de vue, on peut voir l'entreprise des sciences cognitives comme proposant un niveau intermédiaire d'analyse, entre l'analyse logique des opérations intentionnelles et leur réalisation matérielle, le niveau des opérations intentionnelles effectives. On pourrait objecter à cette proposition que le concept de l'arrière-fond (*background*) joue déjà chez Searle le même rôle d'intermédiaire. Ce concept a été introduit par Searle pour désigner les savoir-faire et les précompréhensions implicites rendant possible l'intentionnalité (cf. *ibid.*, pp. 143–144). Cependant, il s'agit de nouveau d'un concept qui relève de l'analyse logique — et qui à ce niveau-là est tout à fait pertinent — et qui ne peut être transposé comme tel à l'analyse des opérations pré-intentionnelles effectives, opérations qui sont désignées en sciences cognitives par le concept de connaissance implicite (*tacit knowledge*). Pour la définition du concept de connaissance implicite en sciences cognitives, voir l'ouvrage de Pascal Engel *Philosophie et Psychologie* (cf. ENGEL P., *Philosophie et psychologie*, Paris, 1996, pp. 222–283).

16 On emprunte le concept de reconstruction rationnelle à Habermas qui l'utilise pour caractériser la méthode de la linguistique de Chomsky et l'étendre à d'autres domaines d'étude (cf. HABERMAS J., "What is Universal Pragmatics ?", in ID., *Communication and the Evolution of Society*, trad. par Th. McCarthy, London, 1979, pp. 8–14). Comme l'analyse formelle ou syntaxique de la cognition de Fodor est également inspirée des idées de Chomsky, ce concept nous semble ici tout à fait adéquat. Les critiques que nous adressons à l'approche purement syntaxique s'adressent donc par transposition également

les années 60, a donné son coup d'envoi à la première génération des sciences cognitives¹⁷. Pour Chomsky, le but de la méthode de reconstruction rationnelle en linguistique est d'élaborer une grammaire de chaque langue étudiée. Par grammaire, il entend un procédé permettant d'engendrer toutes les phrases correctes d'une langue et qui sont utilisées par un locuteur compétent. La reconstruction rationnelle part du fonctionnement effectif des langues naturelles et, à partir de là, construit un système formel, constitué de quelques éléments de base et de règles d'engendrement élémentaires, dans le cadre duquel on peut obtenir l'ensemble des phrases de la langue¹⁸.

Cette méthode correspond exactement au niveau algorithmique de Marr. Dans la recherche d'un algorithme pour la vision, on part des classifications perceptives dont est capable une personne compétente et on construit les règles formelles qui nous permettent d'engendrer ces classifications à partir de quelques informations visuelles de base. On retrouve cette même méthode chez Fodor et Dennett. Au lieu d'algorithme, Fodor parle d'une reconstruction des capacités computationnelles de l'organisme (entendues comme des computations purement syntaxiques sur des symboles formels)¹⁹, capacités qui forment une grammaire de la pensée de façon analogue à la grammaire de la langue de Chomsky²⁰. Dennett parle d'une reconstruction du *design* d'un système²¹. Chez Dennett, la reconstruction rationnelle se place donc du point de vue d'un concepteur potentiel qui pourrait effectivement, et pas seulement théoriquement, engendrer le système²².

Pour mieux comprendre le choix de ces auteurs en faveur d'une approche purement syntaxique, reprenons brièvement la critique que Fodor adresse à l'approche de Marr. Fodor critique une psychologie qui veut aller au-delà d'une étude syntaxique et s'intéresser non seulement à la forme, mais

à Habermas, en particulier à l'utilisation de la reconstruction rationnelle dans sa théorie formelle de la pragmatique linguistique (cf. *ibid.*, pp. 15–20) et dans sa réforme de l'analyse transcendantale (cf. *ibid.*, pp. 21–25).

17 Cf. ENGEL P. *Philosophie et psychologie*, *op. cit.*, p. 50.

18 Cf. HABERMAS, J., *What is Universal Pragmatics ?*, *op. cit.*, pp. 8–14.

19 Cf. FODOR J., *The Language of Thought*, Cambridge (MA), 1979, p. 51. L'usage du concept de computation chez Fodor est donc différent de celui de Marr, dans la mesure où Fodor se limite au seul aspect syntaxique ou opératoire de la computation.

20 "Nothing can be an adequate representation of a message unless it can serve as input to a device capable of computing the structural description of those sentences which express that message ; 'structural description' is here taken in its technical linguistic sense." (*ibid.*, p. 111).

21 Cf. KITCHER P., *op. cit.*, p. 14.

22 Cette position méthodologique se base sur un argument évolutionniste, selon lequel le cerveau est le produit d'une histoire naturelle d'adaptation optimale à l'environnement (cf. DENNETT D., *The Intentional Stance*, Cambridge (MA), 1987, pp. 33–35).

aussi aux contenus des représentations traitées par le système visuel. Sa critique concerne la difficulté de prendre en compte les interactions sémantiques entre l'organisme et l'environnement. L'étude des contenus implique en effet que l'on s'intéresse à la relation entre l'environnement et l'organisme qui représente cet environnement ou qui représente les actions possibles dans cet environnement. Selon Fodor, une telle entreprise est vouée à l'échec parce qu'elle doit soit attendre une science complète de la nature pour pouvoir utiliser des descriptions précises de l'environnement, soit se satisfaire des descriptions du langage ordinaire qui ne sont pas suffisamment précises pour construire des lois exactes²³.

La critique de Fodor paraît à première vue un peu étrange. En effet, on voit mal pourquoi les descriptions de la physique seraient insuffisantes pour décrire les propriétés des objets dans l'environnement dont le système visuel tente de construire une représentation. Une description exacte paraît seulement nécessaire dans le cas où l'on suppose que le système visuel tente de construire une représentation véridique, entièrement adéquate de l'objet. Dans ce cas, seulement à partir d'une connaissance exacte de l'objet dans l'environnement, on peut évaluer si le système visuel "représente" effectivement cet objet. L'objection de Fodor cache donc une présupposition qui porte sur le concept de représentation. Dans la théorie de Fodor, la représentation doit non seulement contenir une information de l'environnement, mais cette information doit également être véridique. Comme le précise Fodor, "les capacités computationnelles [*syntaxiques*] de l'organisme doivent constituer une solution à de tels problèmes [*de représentation*] dans la mesure où les jugements perceptifs produits sont (a) médiatisés par de l'information sensorielle et (b) vrais"²⁴. De plus, "dans la mesure où les mécanismes sensoriels encodent une information de l'environnement, c'est l'état physique de l'environnement qui est ainsi encodé"²⁵. Ce type de réalisme représentationnel défendu par Fodor est tout à fait typique des premiers modèles cognitivistes. Fodor critique à juste titre la difficulté de construire une psychologie des contenus sur cette base.

Cependant, rien ne nous oblige à restreindre le concept de représentation à la représentation véridique du monde. Comme nous l'avons vu dans le chapitre précédent, toute représentation est relative à un milieu de vie et ne sera donc jamais vraie dans un sens absolu. De plus, de nombreuses représentations sont partielles, encodant seulement les informations relatives aux tâches poursuivies. L'entreprise d'une psychologie portant sur les

23 Cf. KITCHER P., *op. cit.*, p. 13.

24 FODOR J., *op. cit.*, p. 51 ; notre traduction et nos crochets.

25 *Ibid.*, p. 47, note 14 ; notre traduction.

contenus des représentations cognitives n'est donc pas nécessairement dépourvue de scientificité. Tout ce qu'il faut, c'est une connaissance suffisamment détaillée de l'environnement pour spécifier physiquement le fragment significatif de cet environnement pour le système cognitif en question. C'est seulement dans le cas particulier où l'on suppose un réalisme inférentiel des contenus, comme Fodor, que l'approche sémantique paraît vouée à l'échec.

Si on reconnaît le caractère bien-fondé d'une psychologie computationnelle, il semble que cette psychologie pourra même rendre des services appréciables au projet purement syntaxique de Fodor. En fait, le lien entre une telle étude purement syntaxique et l'étude des computations sur les représentations est fourni par l'interprétation des symboles formels de la représentation syntaxique. Cette interprétation est donnée sous forme de règles de correspondance (par exemple, "à tous les x correspond la couleur rouge"). Ces règles de correspondance sont établies lors de la définition formelle du problème à résoudre, quand on choisit un format de représentation formelle des entrées et des sorties. Aussi bien pour la définition de la tâche à accomplir (la définition des entrées et des sorties), que pour l'interprétation des symboles formels (le choix du format de représentation), l'analyse syntaxique des opérations mentales dépendra du niveau computationnel²⁶. De plus, comme nous l'avons déjà remarqué plus haut, on peut construire de nombreux algorithmes qui produisent des opérations de transformation entrée-sortie équivalentes, sans savoir lequel est utilisé de fait par le système visuel. Une étude au niveau des fonctions de traitement d'information du système cognitif permet donc également de mieux circonscrire l'espace des solutions.

A partir de ce bref parcours des arguments qui plaident pour et contre l'entreprise des sciences cognitives, émerge la perspective d'une intense collaboration entre les neurosciences cognitives, l'intelligence artificielle et la psychologie cognitive. Les neurosciences étudient la structure matérielle, l'intelligence artificielle et la psychologie cognitive respectivement la structure syntaxique et fonctionnelle des opérations intentionnelles. Aucun niveau ne peut faire valoir la suprématie. Les neurosciences dépendent de la psychologie cognitive et de l'intelligence artificielle pour la définition des opérations et des fonctions mentales dont elles étudient la structure

26 Par exemple, "il se pourrait parfaitement que certaines caractéristiques des réseaux neuronaux leur permettent en fait d'accomplir des processus qui n'apparaissent comme différents (reconnaître un phonème, discriminer une forme, extraire un prototype) qu'au niveau de l'analyse fonctionnelle" (cf. SERON X., "La neuropsychologie cognitive", in FRANCK R. (ed.), *Faut-il chercher aux causes une raison ? L'explication causale dans les sciences humaines*, Paris/Lyon, 1994, pp. 344-366, p. 361).

matérielle, l'analyse syntaxique dépend d'une analyse au niveau du contenu traité pour la définition du problème ou du rôle fonctionnel de tel ou tel algorithme. En sens inverse, l'espace de résolution d'un problème donné est délimité par les opérations logiques possibles du système cognitif et les opérations possibles dépendent à leur tour des dispositifs biologiques et matériels disponibles pour les effectuer.

2.1.3. Le problème du fonctionnalisme

Les distinctions introduites par Marr nous ont permis d'examiner deux méthodes d'analyse des opérations cognitives : la méthode algorithmique étudiant la syntaxe des opérations cognitives et la méthode fonctionnelle précisant le traitement de l'information effectué. Ces deux approches concernent la fonction computationnelle réalisée par le système cognitif. D'autre part, cette fonction est implémentée au niveau du système nerveux central. C'est à ce niveau que l'on rencontre une troisième approche de la cognition, l'approche par les neurosciences. Toutefois, dans la théorie de Marr, la question de l'implémentation n'intervient pas en tant que telle dans l'étude des opérations et des fonctions cognitives. Comme nous l'avons vu, le niveau neurophysiologique intervient uniquement en ce qu'il impose des contraintes générales sur l'espace des computations possibles qui peuvent être implémentées dans un système nerveux.

Marr propose d'étudier les opérations cognitives à partir du niveau des fonctions computationnelles. Cependant, pour décomposer les tâches du système visuel il adopte une hypothèse d'optimum. En effet, il suppose que le système visuel utilise les solutions computationnelles les plus économiques pour transformer et encoder les contenus de représentation²⁷. Même si cette hypothèse paraît plausible du point de vue d'un ingénieur ou de quelqu'un qui veut concevoir un système visuel, il se pourrait que l'évolution naturelle n'ait pas imposé de contrainte aussi forte. En fait, les ressources dont un être vivant dispose pour s'adapter à son environnement sont souvent limitées et les solutions adoptées sont parfois du bricolage à partir d'éléments existants.

Pour illustrer les limites d'une telle approche fonctionnaliste, nous pouvons partir d'une analogie avec l'approche des fonctions en biologie. Dans le contexte des discussions autour du fonctionnalisme en biologie, un exemple souvent cité est l'exemple de l'étude de la structure de la main du Panda. Dans une approche fonctionnaliste, si l'on veut comprendre la structure de préhension de la main du Panda, on part d'une description

27 Cf. KITCHER P., *op. cit.*, p. 21.

fonctionnelle de l'organe dont le Panda aurait par exemple besoin pour rassembler de la nourriture dans son milieu de vie habituel. Ensuite, on étudie comment, au cours de l'histoire de la sélection naturelle, cet organe a évolué pour accomplir de façon optimale cette fonction. En réalité cependant, la description fonctionnelle ne nous apprend pas grand chose sur la main réelle du Panda. Celle-ci est en fait une forme qui a évolué à partir d'une patte de carnivore, mais avec un os du poignet tellement démesuré que cette patte peut être utilisée pour coincer des objets, comme des morceaux de bambou, sa principale source de nutrition²⁸. La plupart des détails sur la fonction de préhension se trouvent ici au niveau de l'implémentation. Un autre exemple frappant du caractère *ad hoc* des structures, qui sont le résultat de l'évolution naturelle, est donné par la structure de l'oreille interne des mammifères²⁹. Au niveau de l'oreille moyenne des mammifères se situent trois petits os, qui ont pour fonction d'amplifier les sons captés par l'oreille. Chez les reptiles, prédécesseurs des mammifères, ces mêmes os faisaient partie du mécanisme d'accrochage de la mâchoire inférieure et avaient comme fonction de mâcher la nourriture. Ce qui à l'origine faisait partie de la mâchoire inférieure a été incorporé dans l'oreille moyenne pour améliorer l'audition. La solution ne semble peut-être pas très élégante, mais la raison pour laquelle elle a été adoptée est que les ressources (les os) étaient disponibles et que cela marchait.

De même, en sciences cognitives, rien ne dit que les capacités computationnelles du système nerveux sont des capacités optimales. Peut-être que beaucoup de nos solutions computationnelles sont également des solutions *ad hoc*, qui sont viables sans être des solutions optimales³⁰. C'est précisément ce qui est suggéré par les expériences de psycho-physique de Ramachandran que nous avons étudiées dans le premier chapitre. Comme nous l'avons souligné à cet endroit, la plupart de nos représentations sont des représentations partielles, relatives à un milieu de vie et à des intérêts contextuels. Les capacités de traitement d'information du système visuel, dont Marr espérait rendre compte à partir de l'hypothèse de l'optimum, seraient donc plutôt obtenus par un bricolage circonstancié de solutions partielles.

Ce que nous retenons de ces critiques du fonctionnalisme, c'est le fait que l'organisation fine du système cognitif en sous-systèmes fonctionnels ne

28 Cf. GOULD S.J., *Le pouce du Panda*, *op. cit.*, p. 20.

29 Cf. RAMACHANDRAN V.S., "Interactions between Motion, Depth, Color and Form : the Utilitarian Theory of Perception", *op. cit.*, pp. 346–360.

30 Pour une critique plus étendue de l'hypothèse de l'optimum de Marr, cf. CHURCHLAND P.S., RAMACHANDRAN V.S. et SEJNOWSKI T.J., "A Critique of Pure Vision", *op. cit.*

correspond pas nécessairement à une organisation optimale, mais peut également résulter d'une combinaison *ad hoc* de ressources biologiques disponibles. Des hypothèses spécifiques — à côté des hypothèses générales — sur le fonctionnement du système nerveux seront donc importantes dans l'étude du processus effectif de traitement d'information.

L'absence d'une contrainte d'optimum dans l'évolution naturelle fournit une deuxième raison pour laquelle le chercheur en sciences cognitives devrait s'intéresser au fonctionnement du système nerveux. La première, qu'on a déjà rencontrée chez Marr, est que la neurophysiologie permet de circonscrire l'espace des computations possibles qui peuvent être implémentées dans un système nerveux. Ici on constate qu'on ne doit pas seulement connaître l'espace des opérations possibles, mais également avoir une idée des ressources effectivement disponibles. Une étude de la structure effective du cerveau sera donc importante pour pouvoir construire une théorie des opérations effectives dans les systèmes cognitifs naturels. Si l'on tient compte de l'importance du niveau de l'implémentation, ainsi que du fait que la relation entre structure cérébrale et fonction cognitive n'est pas univoque — une même structure cérébrale peut implémenter différentes fonctions, selon le processus dans lequel elle est engagée —, l'étude des fonctions cognitives demandera un va-et-vient entre le niveau neurophysiologique, spécifiant les structures, et le niveau computationnel, spécifiant les systèmes de traitement d'information dans lesquels elles peuvent intervenir³¹.

2.2. L'articulation entre la computation et l'implémentation

La perspective ouverte par la théorie de Marr sur l'articulation d'un niveau opératoire (algorithmique) et un niveau fonctionnel dans l'étude de la cognition se trouve limitée par le fait que sa théorie évacue la prise en compte de la question de l'implémentation dans l'étude des fonctions. Par là, le concept de fonction est réduit à une simple fonction *mathématique* de traitement d'information et ne permet pas de rendre compte de l'aspect finalisé (orientation vers un but) de la fonction cognitive. Comme nous l'avons vu, dans la théorie de Marr la problématique du caractère biologique des systèmes cognitifs naturels est traitée comme une question périphérique

31 Pour une argumentation plus détaillée de cette position, voir ZAWIDZKI T. et BECHTEL W., "Gall's Legacy Revisited : Decomposition and Localization in Cognitive Neuroscience", in ERNELING C.E. et JOHNSON D.M. (eds.), *Mind as a Scientific Object : Between Brain and Culture*, Oxford, sous presse. Ce texte est disponible sur l'internet à l'adresse [http : // www.artsci.wustl.edu /~wbechtel](http://www.artsci.wustl.edu/~wbechtel).

par rapport à l'étude computationnelle et n'intervient qu'au niveau des contraintes globales de son approche.

Dans cette deuxième section, nous souhaitons approfondir la question de l'articulation entre le niveau computationnel et le niveau de l'implémentation biologique. En un premier temps, nous étudierons la spécificité du concept de fonction en biologie. Les débats autour du concept de fonction en biologie visent à clarifier les rapports entre les principes de la biologie et ceux de la physique. Nous partirons également de la problématique du rapport entre la biologie et la physique, en nous posant plus particulièrement la question du rapport entre l'étude fonctionnelle du système nerveux, à partir des modèles biologiques, et l'étude des interactions moléculaires dans le cerveau, à partir des lois de la physico-chimie. Ensuite, en un deuxième temps, nous reprendrons la question du rapport entre les fonctions computationnelles et l'implémentation neurophysiologique, en essayant d'articuler l'analyse de l'organisation fonctionnelle du système nerveux et l'analyse computationnelle de Marr.

2.2.1. L'explication fonctionnelle en biologie

Une première façon d'aborder la question de l'organisation fonctionnelle du système nerveux est de partir du fait que le fonctionnement du système nerveux central peut lui-même s'étudier à plusieurs niveaux différents : réseau neuronal, cellule isolée ou système moléculaire³². Ces différents niveaux sont organisés hiérarchiquement entre eux et le rapport entre deux niveaux consécutifs est chaque fois un rapport structure - fonction, le niveau inférieur fournissant la description structurale du système dont on étudie la fonction au niveau supérieur³³. Par exemple, au niveau des cellules isolées, les neurones peuvent fonctionnellement être définis comme des systèmes transmetteurs de potentiels d'action, tandis qu'au niveau de la structure

32 Cf. SERON X., *op. cit.*, p. 361.

33 En toute rigueur la description structurale contient deux composantes : la description des éléments et de leur relations, et la spécification du fonctionnement du système. Il faudrait donc parler à chaque niveau de structure - fonctionnement - fonction au lieu de parler simplement de structure - fonction (cf. PAILLARD J., "Réflexions sur l'usage du concept de plasticité en neurobiologie", in *Journal de Psychologie Normale et Pathologique*, 73 (1976), pp. 33-47). De plus, il est assez trivial de constater de nouveau que ces différents niveaux sont variables indépendamment, une même structure peut donner lieu à des dynamiques de fonctionnement différentes, par exemple selon le chemin temporel emprunté ou la façon dont le système est sollicité de l'extérieur (cf. *ibid.*, pp. 44-45). Cependant, pour ne pas alourdir l'exposé (et l'inflation des niveaux d'étude) nous faisons abstraction de cette distinction supplémentaire et parlons tantôt de structure, tantôt de fonctionnement quand nous désignons la description structurale.

moléculaire, ils seront décrits en termes de flux ionique et de perméabilité membranaire. A un niveau supérieur, les réseaux de neurones seront fonctionnellement décrits par leur connectivité, le signe des activations et les transformations dans les structures d'interconnexion. La neurophysiologie du neurone isolé occupe ici le niveau structural. Dans la perspective d'une analyse psychologique finalement, les réseaux de neurones seront considérés comme l'implémentation structurale des fonctions psychologiques. Les caractéristiques de fonctionnement d'un réseau de neurones peuvent par exemple expliquer comment le système réalise une opération de discrimination de formes³⁴.

Ce qui nous intéresse ici n'est cependant pas tant de montrer comment l'on peut, à partir du niveau physique, atteindre progressivement le niveau des réseaux de neurones, en spécifiant les relations complexes de type structure - fonction pour chaque niveau. Ce que l'on veut comprendre, c'est l'existence même des relations structure - fonction. En effet, ce qui pose problème pour le philosophe des sciences, c'est le statut du concept de fonction, en ce qu'il semble faire appel à des causes finales, et son rapport à la causalité physique efficiente. Ce qu'il faut donc clarifier, c'est le saut qualitatif entre la détermination simplement physique et la détermination biologique ou organisationnelle.

La hiérarchie des rapports structure - fonction que nous avons schématiquement décrite ci-dessus pourrait suggérer que, par réduction successive des niveaux, on peut réduire l'explication fonctionnelle psychologique à l'explication physicaliste du niveau le plus bas. Pour montrer qu'une telle réduction n'est pas possible, nous procédons en deux étapes. D'abord à partir d'un exemple tiré de la biologie moléculaire, le niveau le plus bas de notre hiérarchie, nous montrons comment l'explication en biologie demeure nécessairement incomplète si elle se limite aux seules causes efficientes physiques. Ensuite, nous montrons comment on peut articuler de façon cohérente l'explication fonctionnelle et la causalité physique en utilisant le concept d'organisation téléonomique.

Considérons ce qui a été découvert au sujet de la structure du matériel génétique, l'ADN et l'ARN³⁵. L'ADN est une longue molécule composée de nucléotides. Chaque nucléotide porte une des 4 bases organiques du code génétique, soit l'adénine (A), la guanine (G), la cytosine (C) ou la thymine (T). La structure en double hélice de l'ADN est le résultat de l'association complémentaire des bases entre elles : l'adénine s'associe à la thymine et la cytosine à la guanine pour former des couples de nucléotides A - T, T - A,

34 Cf. SERON X., *op. cit.*, p. 361.

35 Cf. ROSENBERG A., *The Structure of Biological Science*, Cambridge, 1985, pp. 38-42.

C - G, G - C. Le code génétique de l'ADN est contenu dans les chaînes de nucléotides, par exemple la chaîne 3' - TTAATCCCACAG - 5' et la copie complémentaire 5' - AATTAGGGTGTC - 3' (cf. la figure 2.1.). Ce code contient l'information nécessaire pour la construction des protéines par une double opération : d'abord la transcription du code de l'ADN sur l'ARN et ensuite la traduction du code de l'ARN en chaîne de molécules protéiques. L'ARN joue le rôle de messenger entre l'ADN dans le noyau cellulaire et la construction de protéines au niveau des ribosomes.

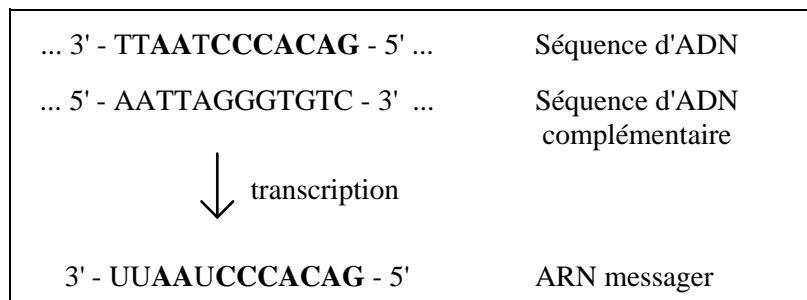


Figure 2.1. Un petit segment du gène de l'opéron tryptophane. La chaîne ARN est identique à la séquence d'ADN de départ, à part la thymine (T) qui a été remplacée par l'uracil (U). Comme la thymine, l'uracil est une base complémentaire de l'adénine (A). Les bases sont représentées par les symboles A = adénine, C = cytosine, T = thymine, U = uracil, G = guanine. Les chiffres 3', 5' indiquent la direction de la chaîne de nucléotides (figure adaptée de ROSENBERG A., 1985, p. 38).

La transcription du code de l'ADN vers l'ARN se fait par un mécanisme de bases complémentaires. Cependant l'ARN diffère légèrement de l'ADN. Les bases dont il est composé incluent l'adénine, la guanine et la cytosine, mais au lieu de la thymine l'ARN contient l'uracil (U). Comme la thymine, l'uracil est complémentaire à l'adénine. De cette façon l'on obtiendra à partir de la chaîne ADN complémentaire 5' - AATTAGGGTGTC - 3' la chaîne ARN 3' - UUAUCCACAG (cf. la figure 2.1.). La seule différence avec la chaîne d'ADN de départ, la chaîne 3' - TTAATCCCACAG - 5' est que dans la copie ARN la thymine est remplacée par l'uracil. Cette différence entre l'ADN et l'ARN est restée longtemps un mystère. Pourquoi l'ADN contient la thymine et l'ARN l'uracil ? En effet, on n'a pas besoin de la thymine pour assurer la transcription de l'information génétique. Tout ce qui est fait par la thymine dans l'ADN pourrait être fait par l'uracil. Aussi bien la stabilité de la double hélice que la transcription ADN - ARN pourraient être assurées par l'uracil. En plus, le coût énergétique pour produire une molécule de thymine

est beaucoup plus élevé que celui requis pour produire une molécule d'uracil et la présence de la thymine viole donc la loi de minimisation de l'énergie³⁶.

La présence de la thymine s'explique si on considère l'extrême instabilité d'une autre base de l'ADN, la cytosine. La structure de la cytosine est très semblable à celle de l'uracil et, par une simple réaction de désamination, la cytosine se transforme en uracil. Pour éviter une telle mutation et assurer la fiabilité du système génétique un mécanisme de régulation permet de corriger la mutation cytosine - uracil³⁷. Cependant, pour qu'un tel mécanisme puisse fonctionner, il faut pouvoir reconnaître la mutation et donc pouvoir distinguer la cytosine transformée en uracil des autres bases non mutées. Ceci ne serait évidemment pas possible si au lieu de la thymine l'ADN contenait déjà l'uracil (cf. figure 2.2.). La présence de la thymine s'explique donc par rapport à la fonction spécifique de l'ADN qui est d'assurer une réplication et une traduction fiable du code génétique. Au niveau de l'ARN une telle fiabilité n'est pas requise. Un même gène produit beaucoup de molécules d'ARN et si quelques-uns sont mal transcrits ou désaminés avant la traduction en protéines, le dommage causé ne sera que mineur. Le supplément d'énergie, pour transformer l'uracil en thymine, est donc justifié dans le cas de l'ADN, mais pas dans le cas de l'ARN.

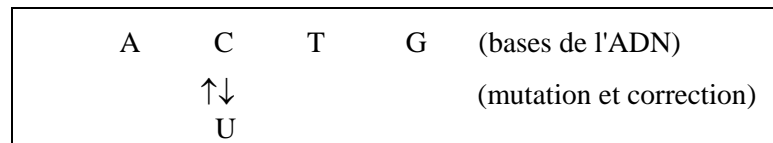


Figure 2.2. La désamination de la cytosine (C) en uracil (U). La reconnaissance de la mutation ne serait pas possible si l'ADN contenait déjà l'uracil au lieu de la thymine (T).

Cette explication de la présence de la thymine dans l'ADN a plusieurs traits remarquables, par quoi on peut la distinguer de l'explication physicaliste. Un trait important, c'est l'appel au rôle fonctionnel de la thymine et de l'uracil. Leur présence s'explique par rapport à la fonction spécifique de l'ADN et de l'ARN dans le système cellulaire. Par conséquent, la différence entre les bases organiques de l'ADN et l'ARN est expliquée par *l'effet* de cette différence et non par les *causes*. C'est en faisant appel au rôle fonctionnel des composantes que l'explication en biologie fait appel à des

36 Cf. *ibid.*, p. 39.

37 Cf. *ibid.*, p. 40.

causes finales, l'explication d'un phénomène par les effets qu'il pourra produire.

Le recours aux fonctions biologiques peut être contrasté avec l'explication physicaliste la plus proche de la biologie moléculaire, l'explication par la physico-chimie. Là aussi on trouve une explication de la présence de la thymine dans l'ADN et son absence au niveau de l'ARN. L'ADN et l'ARN sont produits dans une réaction de polymérisation. La différence entre les deux s'explique par l'absence du sucre de l'uracil et la présence du sucre de la thymine lors de la polymérisation de l'ADN et inversement pour l'ARN³⁸. La chimie organique peut fournir l'ensemble des schémas réactionnels nécessaires pour ces synthèses. Elle peut même expliquer l'ensemble des effets de cette différence : le détail des réactions impliquées dans la désamination, dans la transformation inverse de l'uracil en cytosine, et les effets de la thymine dans ces réactions. En un sens, la chimie organique explique parfaitement la différence entre l'ADN et l'ARN. Elle nous fournit la chaîne causale des réactions chimiques dont résulte la présence de la thymine dans l'ADN et son absence au niveau de l'ARN. Cependant, elle n'explique pas la raison d'être de cette différence. Du point de vue de la chimie, cette différence est un fait entièrement accidentel de l'univers. Ce n'est qu'à partir d'une compréhension des *fonctions* spécifiques des molécules d'ADN et d'ARN que la raison de leur différence de structure s'explique.

2.2.2. Le concept de système téléonomique

En essayant de caractériser la spécificité de l'organisation fonctionnelle des systèmes biologiques, nous avons pu mettre en évidence deux éléments : le caractère hiérarchique d'organisation en différents sous-systèmes, allant par exemple de l'organisation moléculaire jusqu'à l'organisation des réseaux de neurones, et le caractère orienté ou fonctionnel des sous-systèmes. Maintenant, il faut encore montrer comment on peut articuler de façon cohérente le caractère fonctionnel — l'explication par les effets — et la causalité physique efficiente — l'explication par les causes.

Pour rendre compte du caractère orienté des systèmes vivants, Jacques Monod a introduit le concept de téléonomie³⁹. La raison pour introduire ce nouveau concept, à côté du concept plus familier de téléologie, est que le caractère orienté des êtres vivants est souvent attribué à l'observateur qui

38 Cf. *ibid.*, p. 41.

39 Cf. MONOD J., *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Paris, 1970, p. 25.

étudie le système et pas au système lui-même. Il s'agirait d'une projection de l'observateur⁴⁰. C'est pour souligner l'existence de lois (*nomos*) objectives de la finalité, indépendantes de toute projection, que Monod préfère parler de téléonomie plutôt que de téléologie. En suivant cet usage nous utiliserons dans ce travail le concept de téléologie quand il s'agit de la projection d'une finalité par un sujet, et le concept de téléonomie pour désigner les finalités objectives d'un objet naturel.

L'étude des systèmes finalisés est l'objet de la cybernétique. En effet, la cybernétique étudie les systèmes régulés, et les systèmes finalisés sont un cas particulier de tels systèmes⁴¹. Dans les systèmes finalisés, il s'agit d'accomplir une certaine tâche ou de maintenir un certain équilibre en s'adaptant aux situations. Ces systèmes supposent des mécanismes de contrôle du comportement ou des mécanismes d'ajustement par rapport à un but à atteindre, mécanismes qui relèvent de la cybernétique.

La forme générale d'un système cybernétique est donnée par un programme et des mécanismes de régulation. Le programme indique les opérations à effectuer et les mécanismes de régulation le contrôle des opérations⁴². Un bon exemple d'un tel système est donné par un thermostat. Les instructions sont simples : quand la température dépasse ou s'abaisse en dessous d'une température de référence il faut respectivement diminuer ou augmenter la température. Le contrôle du système est basé sur la rétroaction. A chaque moment l'information de sortie du système (la valeur de la température produite) est transférée à un organe de commande de l'entrée qui permet d'ajuster le fonctionnement de l'appareil pour garder la température constante.

D'autres exemples de systèmes de régulation bien connus, analogues au thermostat, sont le système de régulation de la température du corps ou le système de régulation du niveau de glucose dans le sang par le foie⁴³. En

40 Selon Kant l'attribution d'une finalité interne à un objet relève du jugement réfléchissant et pas du jugement déterminant de la science. On attribue une finalité aux êtres vivants par analogie avec le caractère finalisé des artefacts, objets construits par l'homme en vertu d'une certaine fin (KANT E., *Critique de la faculté de juger*, trad. par A. Philonenko, Paris, 1989, p. 201). Le mouvement de la première cybernétique, en proposant le schème de la rétroaction, s'est clairement opposé à ce point de vue. Ce n'est que dans la deuxième cybernétique, avec l'introduction des systèmes auto-référentiels et la problématique de la connaissance de tels systèmes, que la position kantienne sera réhabilitée (cf. VAN DE VIJVER G., *Van cybernetica naar connectionisme. Een epistemologische studie van doelgerichtheid*, Gent, 1991, p. 2 et p. 19).

41 Cf. LADRIÈRE, J., "La cybernétique", in *Encyclopaedia Universalis*, Vol. 5, Paris, 1970, pp. 256–258 et pp. 274–278, p. 257.

42 Cf. *ibid.*.

43 Pour une analyse détaillée de ce système dans un cadre cybernétique, cf. ROSENBERG A.,

biologie moléculaire, également, de multiples systèmes de régulation ont été mis en évidence. Par exemple les nombreux mécanismes de régulation qui assurent le bon déroulement des opérations de réplication, de transcription et de traduction, comme le mécanisme de réparation de la mutation de la cytosine au niveau de l'ADN que l'on a étudié⁴⁴. A côté des dispositifs stabilisateurs, comme les thermostats, on distingue encore, dans la classe des systèmes finalisés, les dispositifs capables d'accomplir une certaine tâche⁴⁵. Ici la régulation se situe au niveau de la rétroaction des signaux analysant la situation extérieure à laquelle il faut s'adapter. Un exemple, tiré de la robotique, sont les "insectes" de R. Brooks, petits automates capables d'apprendre à marcher en ajustant les paramètres "musculaires" par des boucles de rétroaction simples avec l'environnement⁴⁶.

Les mécanismes de régulation et le programme de contrôle associé sont les premières caractéristiques des systèmes téléonomiques. Toutefois elles ne sont pas propres au système téléonomique, mais appartiennent à tout système cybernétique en général, dont des systèmes de transmission d'information ou des systèmes de calcul sont également des formes particulières. Les caractéristiques spécifiques de l'organisation des systèmes téléonomiques peuvent être données, en suivant les analyses de Nagel⁴⁷, par un ensemble de conditions sur l'organisation du système cybernétique. Ces conditions concernent le rapport du système aux sous-systèmes qui le composent.

Soit un système téléonomique S, son comportement C et son environnement opératoire. Alors le système génère un comportement finalisé si nous pouvons spécifier un ensemble de sous-systèmes de S, reliés de façon causale de la manière suivante :

1. L'état des sous-systèmes à un temps t cause, *d'une manière purement physique*, la réalisation du but B à un temps ultérieur t + p.
2. L'état de chacun des sous-systèmes au temps t est indépendant de l'état des autres au même temps t.
3. Chaque sous-système n'a qu'un nombre limité d'états possibles.
4. Si l'état d'un des sous-systèmes change suffisamment au temps t, alors, en l'absence de changement dans les autres sous-systèmes, au temps t + p le système n'atteindra pas le but B.

op. cit., pp. 53–57.

44 Pour une étude détaillée de la cybernétique moléculaire, cf. MONOD J., *op. cit.*, pp. 85–107.

45 Cf. LADRIÈRE J., "La cybernétique", *op. cit.*, p. 257.

46 Cf. CLARK A., *Being There*, *op. cit.*, p. 12–14.

47 Cf. NAGEL E., *The Structure of Science*, Indianapolis, 1979, pp. 411–418.

Mais

5. Les sous-systèmes interagissent d'une façon telle que, chaque fois qu'un grand changement apparaît dans un des sous-systèmes au temps t , un changement supplémentaire dans les autres sous-systèmes à un temps $t + t_1$ permet au système d'atteindre le but B au temps $t + t_2$.

Les deux dernières conditions concernent les mécanismes de régulation rencontrés ci-dessus. La première condition spécifie que les systèmes téléonomiques sont des systèmes physiques, fonctionnant d'après les lois de la causalité efficiente. La réalisation du but doit donc pouvoir s'appuyer sur un mécanisme purement physique. Mais il y a une différence essentielle avec les systèmes physiques auxquels on ne peut pas attribuer de fonctions ou de buts. Considérons par exemple un fluide qui conserve son état gazeux par l'interaction de la pression, de la température et du volume. Même si ce système tend vers un certain état d'équilibre nous ne dirons pas qu'il s'agit d'un système téléonomique orienté vers le maintien de l'état gazeux. Ce qui manque dans ce système pour parler d'organisation téléonomique, c'est l'indépendance des sous-systèmes : à chaque instant la pression, le volume et la température sont fixés de façon simultanée par des influences réciproques. L'on peut donc difficilement parler d'une organisation en vue d'un but.

En résumant l'analyse de Nagel et celle du système cybernétique en général, nous pouvons dire que c'est l'indépendance des sous-systèmes, combinée avec la rétroaction, qui permet l'apparition de comportements finalisés à partir d'interactions purement physiques⁴⁸. L'explication fonctionnelle

48 Il y a aussi des systèmes qui ne satisfont pas à toutes les conditions mais qui sont néanmoins orientés vers un certain état d'équilibre d'après un principe d'optimum. Nous pensons par exemple au cas des systèmes thermodynamiques. L'irréversibilité des systèmes thermodynamiques implique l'apparition d'un caractère orienté au niveau du système dans sa totalité sans que pour autant on puisse parler de téléonomie. Dans le cas des systèmes thermodynamiques, la condition d'autonomie des sous-systèmes est remplie (les seules interactions sont des chocs entre particules voisines), mais il n'y a pas de rétroaction (conditions 4 et 5). La condition d'autonomie (ici la non-corrélation des particules à distance), combinée avec une hypothèse probabiliste sur l'état initial des particules, suffit pour expliquer l'apparition de l'irréversibilité (cf. par exemple HUANG K., *Statistical Mechanics*, New Delhi, 1988, pp. 84–88.). Il n'est donc pas nécessaire, pour expliquer le caractère orienté des systèmes thermodynamiques, d'introduire l'irréversibilité au niveau des particules, comme le fait Prigogine (cf. PRIGOGINE I. et STENGERS I., *La nouvelle alliance*, Paris, 1986, pp. 345–346), ou d'identifier qualitativement l'organisation du système thermodynamique à l'organisation de l'être

n'est donc ni superflue, ni entièrement réductible aux propriétés purement physiques des composantes. Les fonctions sont des propriétés objectives de systèmes pris dans leur totalité, résultant de l'organisation de sous-systèmes indépendants en système téléonomique par des mécanismes de rétroaction. Paradoxalement le caractère organique d'un système vivant ne tient donc pas à la cohésion plus forte qui existerait entre des composants physiques, comme on pourrait éventuellement le penser intuitivement, mais à la capacité inverse de créer des systèmes avec une forte indépendance. C'est la capacité du vivant à créer des frontières entre des unités élémentaires qui fait émerger les conditions d'une réelle interaction entre systèmes qui ont chacun leur mot à dire.

Pour terminer ces précisions sur le statut de la détermination organisationnelle ou biologique, revenons à l'exemple du début de ce paragraphe, la différence entre l'ADN et l'ARN. Nous sommes maintenant en mesure de formuler d'une façon plus précise l'origine de cette différence. La présence de la thymine dans l'ADN s'explique par son rôle fonctionnel dans un système organisé plus large. La thymine a comme fonction de minimiser les mutations parce qu'elle fait partie d'un système téléonomique qui produit le comportement de réplication fiable de l'ADN. Les composants individuels de ce système, la thymine, l'ADN ou les enzymes de reconnaissance de l'uracil, ne manifestent pas les propriétés de correction, c'est seulement au niveau du système d'ensemble qu'elles se manifestent.

2.2.3. L'émergence des capacités de traitement d'information à partir du cerveau

Au début de cette section, nous avons posé la question du rapport entre les fonctions computationnelles et l'implémentation neurophysiologique de ces fonctions. A partir de notre analyse de l'organisation fonctionnelle des systèmes biologiques, on peut maintenant aborder le problème de l'émergence des opérations cognitives à partir du cerveau. Tout d'abord, si la cognition dépend de l'organisation fonctionnelle du système nerveux, alors la décomposition des fonctions computationnelles en fonctions élémentaires de traitement d'information ne pourra plus se faire d'après une hypothèse d'optimum — comme dans la théorie de Marr —, mais devra être guidée par une analyse des opérations de transformation d'information réalisées par le système nerveux et par une analyse des unités d'information représentées dans le cerveau. La question de l'émergence des opérations cognitives devient alors la question de l'émergence de la capacité d'encoder des

vivant, comme le fait Monod (cf. MONOD J., *op. cit.*, p. 28–29).

représentations dans le système nerveux et de réaliser des opérations logiques de transformation de ces représentations.

L'encodage des représentations dans le cerveau peut se faire au niveau de l'activité des neurones et les opérations logiques peuvent être réalisées par la transmission de configurations d'activité d'un ensemble neuronal à un autre, par exemple par l'addition de différents signaux ou par une inhibition sélective⁴⁹. Evidemment, seul un système nerveux suffisamment complexe, comprenant plus que de simples boucles réflexes ou de boucles de régulation, aura ces capacités. Le point exact à partir duquel on peut commencer à parler de représentation et de computation, et où l'on quitte le domaine des réflexes et des simples adaptations automatiques, est difficile à établir. Pour avoir un aperçu détaillé de ce problème et des différentes définitions du concept de représentation, en fonction de la généralité du schème d'encodage, on peut se rapporter par exemple aux analyses de John Haugeland⁵⁰ et d'Andy Clark⁵¹. Sans entrer dans ce débat extrêmement technique, nous reprenons ci-dessous de façon schématique l'analyse du concept de représentation par Clark.

La définition que donne Clark du concept de représentation est très générale. Elle vise à inclure les représentations de bas niveau, comme les réseaux de neurones qui implémentent directement une information sur la position des membres du corps ou une information sur certains traits élémentaires de l'environnement, aussi bien que les représentations de haut niveau, qui peuvent être manipulées en l'absence des objets auxquels elles réfèrent, comme des représentations abstraites ou des représentations d'objets absents. Alors, en cherchant un dénominateur commun de ces différents niveaux de généralité, un système sera appelé représentationnel si :

- (a) l'on peut identifier des états internes qui ont comme fonction spécifique de véhiculer une information sur des états de choses corporels ou environnementaux,
- (b) la relation de représentation entre les états internes et les états de choses est systématique, elle définit un schème d'encodage, et
- (c) le schème d'encodage est suffisamment général ou complexe, portant sur un ensemble large de stimuli venant de l'environnement ou du corps.

49 Cf. par exemple le modèle du neurone formel de McCulloch et Pitts que nous avons discuté au premier chapitre (cf. MCCULLOCH W. S. et PITTS W., *op. cit.*).

50 Cf. HAUGELAND J., "Representational genera", in RAMSEY W. *et al.* (eds.), *Philosophy and Connectionist Theory*, Hillsdale, 1991, pp. 61–89, p. 62.

51 Cf. CLARK A., *Being There, op. cit.*, pp. 143–147.

Un état interne sera donc une représentation s'il a comme fonction de véhiculer une information suffisamment générale sur ce qui se passe dans l'environnement. La simple présence d'un schème d'encodage (condition b), c'est-à-dire d'une corrélation causale systématique entre l'état interne et un trait du corps ou de l'environnement, ne suffit pas pour parler d'une représentation. Comme le souligne Clark, "il existe une corrélation causale entre les marées et la lune. Cependant nous ne dirons pas que la mer représente la lune ou inversement. En effet, il ne nous paraît pas plausible que les marées (par exemple) aient été sélectionnées, conçues ou aient évolué dans le but de véhiculer une information sur la position de la lune"⁵². Ce qui importe, en plus de la corrélation causale, c'est le rôle que joue l'état interne dans le système dont il fait partie. Il faut que ce rôle soit précisément de véhiculer de l'information et que cette information soit effectivement utilisée par d'autres états, c'est-à-dire qu'il y ait d'autres états qui "consomment" l'information (condition a). Toutefois, même si le système a été conçu ou a évolué pour véhiculer l'information de son environnement, nous n'avons pas encore une condition suffisamment spécifique pour parler d'une représentation. Un tournesol qui se tourne en fonction de la position du soleil est corrélé de façon causale avec la position du soleil et cette corrélation a une fonction adaptative pour le tournesol. Nous ne dirons toutefois pas que le tournesol représente la position du soleil. De même, un simple robot conçu pour réagir à une source lumineuse et se mouvoir vers cette source de façon autonome détecte la lumière mais ne construit pas pour autant des représentations. Il faut de plus que la corrélation causale soit suffisamment complexe, la troisième condition donnée par Clark (condition c) : "Parler en termes de représentation commence à faire sens si l'on est en présence d'états internes qui sont corrélés de façon systématique à tout un ensemble de contingences de l'environnement"⁵³.

Pour illustrer cette définition sur un exemple concret, revenons au cas des représentations de bas niveau, comme les représentations de la position des membres du corps. Chez le rat, les neurones du cortex pariétal postérieur représentent l'information au sujet de la direction de la tête et ils la catégorise en gauche, droite et tout droit⁵⁴. Malgré une corrélation causale assez directe avec les positions du corps, il s'agit bien d'une représentation : nous avons un schéma d'encodage général (chaque catégorie de direction couvre différents stimuli) et le système nerveux du rat "consomme" l'information sur la direction pour aider le rat à courir, comme on peut l'observer à partir

52 *Ibid.*, p. 146 ; notre traduction.

53 *Ibid.*, p. 147 ; notre traduction.

54 Cf. *ibid.*, p. 144.

d'expériences d'enregistrement électrique de l'activité des neurones du cortex du rat⁵⁵.

Cette définition du concept de représentation permet d'éclairer le lien étroit qui existe entre le concept de représentation et les concepts d'opération computationnelle et de fonction computationnelle. D'abord, la représentation suppose un schème d'encodage général transformant des stimuli venant des états de chose du corps ou de l'environnement en états représentationnels internes (conditions b et c). Elle suppose donc nécessairement des opérations logiques de transformation des stimuli, c'est-à-dire des opérations computationnelles. Par conséquent l'on n'aura pas de représentation sans opération computationnelle. Ensuite, le concept de représentation apparaît seulement en lien avec l'organisation fonctionnelle d'un système de traitement d'information (condition a). Un état représentationnel est un état interne dont le rôle dans le système d'ensemble est de réaliser une certaine fonction computationnelle.

A partir de cette définition du concept de représentation, nous pouvons maintenant préciser la question de l'articulation entre les fonctions computationnelles et leur implémentation. Rappelons d'abord brièvement les résultats de notre analyse. Nous avons caractérisé les systèmes biologiques et les systèmes cognitifs respectivement à partir du concept de système téléonomique et du concept de système computationnel. D'un côté, les systèmes téléonomiques sont définis par une organisation en sous-systèmes indépendants et des boucles de régulation entre les sous-systèmes. Comme exemple de tels systèmes, nous avons étudié l'organisation du génome en sous-systèmes spécifiques régulés par l'action des enzymes. De l'autre, les systèmes computationnels peuvent être définis comme étant des systèmes dont les états internes ont pour fonction spécifique de véhiculer des informations générales sur l'environnement et le corps. Leur fonctionnement implique la présence de schèmes systématiques d'encodage et des opérations de traitement de l'information, schèmes et opérations que nous avons désignés par le terme générique d'opérations computationnelles.

Dans le cas particulier de l'étude des systèmes cognitifs naturels, il faudra combiner les principes venant de l'analyse de systèmes téléonomiques et des systèmes computationnels. La différence entre les systèmes cognitifs naturels et les systèmes cognitifs artificiels tient à l'implémentation. Les mécanismes cognitifs artificiels peuvent être implémentés dans des dispositifs électroniques diverses, tandis que dans le cas des systèmes naturels ils sont implémentés dans le système nerveux des organismes biologiques supérieurs. Les capacités cognitives des êtres vivants trouvent

55 Cf. *ibid.*, p. 146.

leur origine dans l'organisation complexe du système nerveux — initialement seulement capable de régulation interne — dans un système téléonomique capable de construire des relations sémantiques complexes entre l'organisme et son environnement.

3. Le défi de la modélisation contextuelle : modéliser les représentations et les fonctions

Dans le premier paragraphe de ce chapitre, nous avons étudié l'opération générale de modélisation en sciences cognitives, à partir des concepts d'opération computationnelle et de fonction computationnelle. En étudiant de plus près ces deux concepts, nous avons argumenté contre une reprise mécaniste du projet de l'étude scientifique de la cognition. Premièrement, à partir de la théorie de la vision de Marr, nous avons vu que l'objet spécifique des sciences cognitives n'est pas tant les mécanismes opératoires que les fonctions de traitement d'information réalisées par les systèmes cognitifs. Deuxièmement, les fonctions computationnelles ne sont pas uniquement des fonctions *mathématiques* de traitement d'information, mais sont des propriétés téléonomiques qui émergent de l'organisation fonctionnelle du système nerveux.

Dans ce deuxième paragraphe, nous retenons cette critique du mécanisme qui caractérise l'opération générale de modélisation de la cognition. Toutefois, dans cette problématique générale nous avons présupposé que la définition des fonctions et des représentations était donnée d'avance. Une telle présupposition n'est plus justifiée dans la modélisation de l'intentionnalité, vu que l'on s'intéresse également à la création d'un contexte sémantique par le système cognitif et à la définition des fonctions, en rapport avec la dynamique de ce contexte. Dans ce paragraphe, nous verrons que l'approche computationnelle de Marr, même quand elle est articulée à la dimension de l'organisation, évacue la prise en compte du contexte sémantique et fonctionnel des opérations intentionnelles.

3.1. L'interaction entre les représentations et la critique fonctionnaliste

Dans l'approche computationnelle de Marr, les représentations manipulées par les systèmes cognitifs n'ont de signification que par rapport à la fonction de traitement d'information dans laquelle elles interviennent. Ainsi, en reprenant l'exemple du programme de reconnaissance de formes, les représentations manipulées par un tel programme — les angles, les courbures, les surfaces, etc. — n'ont une signification que par rapport à la

fonction de reconnaissance de formes. De même, dans les réseaux de neurones étudiés au premier chapitre, l'algorithme de la catégorisation de données implémenté dans ces réseaux réalise des processus de traitement (reconnaître un visage, discriminer des phonèmes, reconnaître des formes, ...) qui n'apparaissent comme différents qu'au niveau de l'analyse fonctionnelle.

Dans l'approche de Marr, c'est la fonction qui détermine les significations, ou encore c'est la fonction qui définit le contexte sémantique. Toutefois, la prise en compte de la dimension de l'implémentation dans l'étude des opérations cognitives indique un autre rapport possible entre les représentations et les fonctions. En effet, comme nous l'avons vu, du point de vue de l'implémentation la représentation n'est pas seulement définie dans son rapport au traitement de l'information, mais également par un schème d'encodage dans le cerveau. En termes techniques, la représentation suppose un schème d'encodage général qui définit des opérations de transformation logiques des stimuli captés par le système nerveux en état représentationnel interne (la condition b et c de l'analyse de Clark). Du point de vue de l'implémentation, c'est donc l'encodage qui crée le contexte sémantique et pas la fonction. C'est bien cette indépendance du contexte sémantique par rapport à la fonction que nous avons pu étudier dans la section du premier chapitre consacrée à l'écologie subjective. On se souvient par exemple des expériences sur la représentation de l'espace chez l'enfant. Dans ces expériences, une même fonction de reconnaissance de pentes donne lieu à un traitement d'information différent selon différentes représentations de l'espace de locomotion.

Le fonctionnalisme suppose que les représentations sont des données neutres qui reçoivent leur signification uniquement de l'interprétation fonctionnelle. Comme le remarque également Kitcher, cette position objectiviste s'appuie sur une hypothèse particulière d'architecture fonctionnelle des systèmes cognitifs, qui est l'hypothèse de la modularité⁵⁶. Cette hypothèse suppose la présence, dans le système cognitif, de modules fonctionnels indépendants et la collaboration de tels modules dans la réalisation de la tâche cognitive. Selon cette hypothèse, il n'y a pas d'interférence entre les modules et la construction des représentations n'est pas influencée par une dynamique autonome d'interaction entre les modules. C'est finalement cette hypothèse qui justifie la priorité de la fonction (le but) sur les représentations (un champ sans autonomie propre) dans le fonctionnalisme.

Dans cette section, nous voulons présenter quelques arguments qui remettent en question l'hypothèse de la modularité. Comme la problématique de la modularité porte sur l'architecture fonctionnelle des systèmes cognitifs,

56 Cf. KITCHER P., *op. cit.*, p. 19.

nous aborderons la discussion par le biais des neurosciences cognitives, qui étudient l'organisation fonctionnelle fine des systèmes cognitifs en rapport aux mécanismes d'implémentation. Comme les neurosciences cognitives tentent d'élaborer des ponts entre les fonctions et les mécanismes, on trouvera, dans ce champ de recherche, aussi bien des disciplines étudiant la structure cérébrale, comme la neuro-anatomie ou la neurophysiologie, des disciplines qui étudient directement les rapports structure - fonction, comme la neuropsychologie clinique ou l'imagerie fonctionnelle, et des approches fonctionnelles, proposant des schémas de décomposition d'une fonction complexe en fonctions élémentaires, comme la psychologie cognitive.

Les premiers arguments sur lesquels nous nous baserons portent sur la difficulté de l'interprétation des expériences conduites en neurosciences cognitives. Si une expérience contredit l'indépendance de deux modules fonctionnels, deux interprétations équivalentes peuvent être proposées : soit on sauvegarde l'hypothèse de modularité en postulant une décomposition plus fine du système en modules indépendants, soit on postule qu'on est en présence de deux modules interactifs. Nous verrons dans cette section, à partir d'une discussion des méthodes utilisées en imagerie fonctionnelle et en neuropsychologie clinique, qu'il n'est pas toujours possible de trancher entre ces deux interprétations. Ensuite, dans la section suivante, nous donnerons un aperçu des données expérimentales qui montrent l'existence de modules interactifs et ceci même au niveau de processus élémentaires de traitement de l'information visuelle. Ces expériences suggèrent l'image d'un système cognitif qui crée de façon active, à travers l'interaction de différents sous-systèmes représentationnels, son propre contexte de significations. Une prise en compte de cette interaction dans la modélisation, permettra alors d'articuler l'étude des mécanismes opératoires dans chacun des sous-systèmes et l'étude de la définition par le système d'un contexte d'effectuation ces opérations.

3.1.1. La décomposition modulaire dans les études structure - fonction

3.1.1.1. La méthode de la dissociation en neuropsychologie clinique

Regardons d'abord le cas des études de la neuropsychologie clinique sur les lésions cérébrales. Historiquement, l'accumulation de matériel expérimental sur les lésions cérébrales a joué un rôle important dans l'établissement de l'hypothèse de la modularité⁵⁷. Un premier exemple de

57 Cf. COLTHEART M. et DAVIES M., "Le concept de modularité à l'épreuve de la neuropsychologie", in ANDLER D. (ed.), *Introduction aux sciences cognitives*, op. cit.,

découverte de systèmes modulaires remonte aux travaux fondateurs sur les troubles du langage de Broca et de Wernicke⁵⁸. Par ses observations de 1861, Broca avait établi l'existence d'une lésion cérébrale qui cause une perte de la capacité de production de la parole. Néanmoins, malgré ce déficit, le patient avait maintenu sa capacité de compréhension du langage parlé. A partir de ces observations, on pourrait être tenté de conclure à l'indépendance de la capacité de compréhension et de la capacité de production du langage parlé. Cependant, il se pourrait que ces fonctions soient réalisées par un système unique et que la capacité de production dépende par exemple de la capacité de compréhension. Les travaux de Wernicke, publiés en 1874, montrent toutefois que les deux capacités sont bien indépendantes. Wernicke décrit le cas d'un trouble de la parole qui est l'inverse de celui étudié par Broca : à cause d'une lésion cérébrale dans le lobe temporal, un patient présente un trouble au niveau de la compréhension, mais ses capacités d'expression sont restées intactes. Wernicke conclut à l'existence de deux régions fonctionnelles distinctes : une région, dans le lobe frontal, siège des représentations motrices associées à la production des sons et l'autre, dans le lobe temporal, siège des représentations de la signification associée aux sons.

La méthode utilisée par Wernicke pour établir l'indépendance des deux modules de traitement du langage, un module pour la production et un pour la compréhension des sons, est aujourd'hui connue sous le nom de double dissociation clinique. Dans un article sur l'hypothèse de la modularité en neuropsychologie, Max Coltheart et Martin Davies précisent les critères qui permettent d'établir la dissociation entre des modules de traitement. Ils distinguent la double dissociation en tant que telle et des dissociations sélectives d'un module donné. D'après Coltheart et Davies, ces deux types peuvent être définis comme suit :

“(a) On soutient qu'un système X est un module parce qu'on a observé qu'une lésion cérébrale pouvait altérer son fonctionnement sans modifier le comportement normal de tous les autres systèmes, ces autres systèmes pouvant en revanche être altérés chez des patients pour lesquels X fonctionne normalement.

(b) On soutient qu'un système X se compose de deux modules X1 et X2 parce qu'on a observé que : (i) chaque fois que le fonctionnement de X est altéré par une lésion cérébrale, X1 et X2 (ou l'un ou l'autre) sont altérés ; (ii) X1 peut être altéré tandis que X2 est toujours intact et (iii) X2 peut être altéré tandis que X1 est toujours intact”⁵⁹.

pp. 109–130.

58 Cf. MISSA J.-N., *L'esprit-cerveau. La philosophie de l'esprit à la lumière des neurosciences*, Paris, 1993, pp. 142–143. On trouve également, dans ce passage, un aperçu bibliographique des travaux contemporains sur l'aphasie.

59 COLTHEART M. et DAVIES M., *op. cit.*, p. 119.

De nombreuses décompositions modulaires ont été établies à partir de cette méthodologie de la dissociation. En passant en revue la littérature récente sur l'étude des troubles du langage, Coltheart et Davies ne distinguent pas moins de cinq modules de traitement du langage différents qui existent séparément (un module sémantique, syntaxique, morphologique, phonologique et orthographique)⁶⁰. De même, dans le domaine de l'étude des différents troubles associés à la reconnaissance visuelle, des expériences de double dissociation ont pu établir l'existence d'un module séparé pour la catégorisation d'un objet et un module séparé pour la représentation spatiale de cet objet⁶¹. Ou encore, dans le domaine de la reconnaissance des visages, on a pu observer l'existence d'un module séparé de reconnaissance de l'émotion associée à un visage et un module de reconnaissance de l'identité de la personne associée à ce visage⁶².

La méthode de la dissociation semble donc fournir des arguments convaincants en faveur de l'organisation modulaire des systèmes cognitifs. Cependant, une étude détaillée de la méthodologie de la dissociation montre que même à partir de cas de double dissociation on ne peut pas encore inférer la modularité, malgré ce que prétendent Coltheart et Davies. Dans un article sur la méthodologie de recherche en neuropsychologie cognitive, Clark Glymour montre que les conclusions tirées à partir d'expériences de double dissociation sont seulement valables si certaines hypothèses auxiliaires sont satisfaites⁶³. Ceci apparaît manifestement si on considère des expériences de double dissociation sur des modules avec une certaine complexité interne⁶⁴. Nous pouvons illustrer les problèmes rencontrés dans le cas de modules complexes à partir d'un système de deux modules de reconnaissance de formes (d'estimation de surface, de détection des angles, etc. ...), dont le premier traite plus facilement des formes rectangulaires, tandis qu'un deuxième module traite plus facilement des formes ovales. Considérons par exemple que le premier module de ce système soit lésé partiellement, de façon à ce qu'il ne soit plus capable de traitement des ovales, mais que le traitement des rectangles soit conservé. A cause de cette lésion partielle le système d'ensemble ne sera plus capable de reconnaître les

60 Cf. *ibid.*, p. 128.

61 Cf. *ibid.*, p. 122.

62 Cf. SERON X., *op. cit.*, p. 353.

63 Cf. GLYMOUR C., "On the Methods of Cognitive Neuropsychology", in *The British Journal for the Philosophy of Science*, 45 (1994), pp. 815-835.

64 On rencontre cette situation par exemple dans le cas d'une architecture semi-connexionniste, où la structure interne de chaque module est un réseau de neurones artificiels (cf. *ibid.*, p. 830 et SHALLICE T., *From Neuropsychology to Mental Structure*, Cambridge, 1988).

formes ovales. Supposons également que, dans un autre déficit, le second module soit lésé partiellement, mais que cette fois-ci la lésion atteigne uniquement la capacité de traitement des rectangles. On obtiendrait alors une expérience de double dissociation telle qu'on l'a représentée à la figure 2.3.

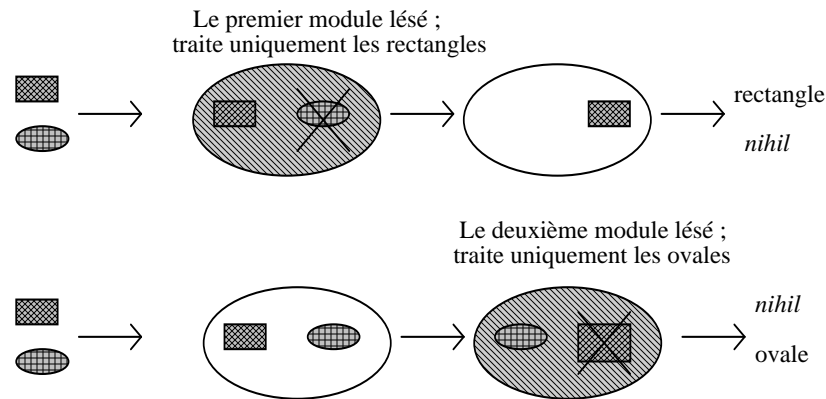


Figure 2.3. Une double dissociation à partir d'une lésion partielle des modules de traitement. Dans le premier cas, le système ne sait plus reconnaître les ovales, dans le deuxième cas il ne reconnaît plus les rectangles (d'après GLYMOUR C., 1994, pp. 831-832).

Quelle conclusion pourrait-on tirer d'une telle expérience de double dissociation ? On peut au moins conclure qu'il existe un module $m(A)$ impliqué dans le traitement de la donnée A, les rectangles, et un module $m(B)$ impliqué dans le traitement de la donnée B, les ovales. Mais, la double dissociation ne nous permet pas de conclure que le module $m(A)$ n'est pas impliqué dans le traitement de la donnée B ou que le module $m(B)$ n'est pas nécessaire dans le traitement de la donnée A. La double dissociation ne permet donc pas de conclure à l'existence de deux modules indépendants, l'un dédié spécifiquement au traitement de A et l'autre au traitement de B, comme on le veut dans l'hypothèse forte de la modularité.

L'interprétation de l'expérience de double dissociation dépendra d'hypothèses auxiliaires. Si on a des informations supplémentaires sur le processus de traitement de l'information dans lequel les modules sont engagés, on pourra trancher entre l'existence de deux modules complexes ou de deux modules simples. Cependant, en l'absence de telles hypothèses auxiliaires, il y a différentes architectures qui peuvent être postulées à partir d'une observation de double dissociation. Faut-il en conclure que les travaux sur la double dissociation, comme ceux de Broca et de Wernicke sur le

traitement du langage, soient mal fondés ? Aucunement. Ce que Glymour tente de mettre en évidence, c'est qu'on ne peut pas conclure à la modularité rien qu'à partir des expériences de double dissociation. L'interprétation de ces expériences reste ouverte tant qu'on ne dispose pas d'hypothèses auxiliaires sur le rôle des modules dans le traitement des données. Dans le système hypothétique de reconnaissance de figures géométriques que l'on a discuté, l'expérience de double dissociation combinée avec des hypothèses auxiliaires sur la structure interne des modules permet de conclure à l'existence de modules séparés, mais pas à un module de traitement de rectangles et un module de traitement d'ovales. Cette dernière interprétation s'imposerait, si on supposait en plus que le premier module traite uniquement les formes rectangulaires et le deuxième uniquement les formes ovales.

L'interprétation des expériences de dissociation en neuropsychologie cognitive dépendra donc d'hypothèses auxiliaires sur la structure interne ou la complexité du traitement de certaines données. Cette information supplémentaire sur le système étudié peut venir soit de disciplines étudiant la structure fine du système cognitif, comme la neurophysiologie, soit de disciplines s'intéressant à la décomposition fonctionnelle des tâches cognitives en tâches élémentaires, comme la psychologie cognitive ou la simulation des processus cognitifs sur ordinateur.

3.1.1.2. La méthode de la soustraction en imagerie cérébrale

On rencontre des problèmes d'interprétation analogues à ceux rencontrés en neuropsychologie dans d'autres disciplines s'intéressant aux rapports structure - fonction, comme l'imagerie fonctionnelle ou l'enregistrement direct de l'activité électrique de neurones isolés. Pour donner une illustration supplémentaire de ces problèmes d'interprétation dans l'étude des rapports structure - fonction, regardons de plus près une méthodologie expérimentale en imagerie fonctionnelle où l'hypothèse de modularité joue un rôle important, l'observation de l'activité du cerveau à l'aide du PET-scan.

Le PET-scan (*Positron Emission Tomography*) permet d'observer les régions du cerveau où le flux sanguin augmente lors de l'exécution d'une certaine tâche. Lors d'une expérience PET-scan, on demande aux sujets d'exécuter différentes tâches et pour chaque tâche on prend une image de l'activité du cerveau. Si les tâches sont choisies de façon à différer seulement dans une ou quelques opérations simples, alors on peut, par soustraction, isoler la tâche élémentaire (l'opération ou les opérations simples) qui est présente uniquement dans une tâche et pas dans l'autre. Pour identifier la région du cerveau où l'on peut localiser cette tâche élémentaire, on soustrait

les images correspondantes obtenues à l'aide du PET-scan et l'activité cérébrale montrée dans l'image résultante indique la localisation recherchée.

<i>Expériences d'imagerie cérébrale</i>	<i>Localisation dans le cerveau : soustraction</i>
expérience 1 (tâche de contrôle) : le sujet doit regarder un point noir sur un écran (1)	
expérience 2 : le sujet doit regarder passivement des mots qui apparaissent sur un écran (2)	région spécifique de lecture de mots : (2) moins (1)
expérience 3 : le sujet doit prononcer des mots qui apparaissent sur un écran (3)	région spécifique de prononciation verbale : (3) moins (2)
expérience 4 : le sujet doit prononcer un verbe qu'il associe à un substantif qui apparaît sur un écran (4)	région spécifique de traitement de la signification : (4) moins (3)

Tableau 2.1. Schéma de l'expérience de Petersen *et al.* pour identifier le module de traitement des significations associées à des mots simples.

Un bon exemple de cette méthodologie de soustraction est donné par une étude de Petersen *et al.* sur la localisation du module de traitement des significations associées à des mots simples⁶⁵. Pour identifier la région impliquée dans ce traitement, l'équipe de Petersen a procédé à sept expériences différentes, trois à partir de mots entendus, trois à partir de mots vus et une tâche de contrôle commune. Dans une tâche d'association de significations, les sujets devaient produire à haute voix un verbe qui pourrait correspondre à un mot simple, comme dire *habiter* à partir du mot *maison* ou *lire* à partir du mot *livre*. Pour obtenir la région spécifique impliquée dans le traitement des significations, il fallait ensuite soustraire de l'image PET obtenue à partir de cette première tâche, d'autres images obtenues dans des tâches de prononciation ou des tâches de lecture seule. On a synthétisé les différentes étapes de cette expérience dans le tableau 2.1., en reprenant, pour simplifier, uniquement les expériences à partir de mots lus.

Dans un article sur la méthodologie du PET-scan, Robert Stufflebeam et William Bechtel commentent l'expérience de Petersen⁶⁶. Le point crucial

65 Cf. PETERSEN S.E., FOX P.T., POSNER M.I., MINTUN M. et RAICHLE M.E., "Positron Emission Tomographic Studies of the Processing of Single Words", in *Journal of Cognitive Neurosciences*, 1 (1989), pp. 153-170.

66 Cf. STUFFLEBEAM R.S. et BECHTEL W., "PET : Exploring the Myth and the Method", in *Philosophy of Science*, supplement (1997). Ce texte est disponible sur l'internet à l'adresse <http://www.artsci.wustl.edu/~wbechtel/>.

dans cette expérience est la soustraction des images et donc l'isolation de processus additifs, ici la lecture, la prononciation et l'association de signification. Ces processus doivent intervenir dans la tâche d'ensemble, l'assignation de significations à des mots lus, sans influences réciproques. Comme le précisent bien Stufflebeam et Bechtel, une telle hypothèse ne peut être vérifiée en se basant uniquement sur les expériences PET-scan, mais doit se baser sur une concordance avec des informations venant d'études de la structure, d'autres études structure - fonction et les études de décomposition fonctionnelle. Le point fort de l'expérience de Petersen est précisément une sélection judicieuse de tâches indépendantes, basées sur des informations venant d'autres disciplines. Comme dans le cas de la neuropsychologie cognitive, la recherche des modules s'appuie sur un ensemble d'hypothèses auxiliaires venant d'autres disciplines.

Souvent la modularité des tâches étudiées lors d'une expérience PET-scan n'est pas aussi bien établie que dans l'expérience de Petersen *et al.* Dans ce cas, plusieurs interprétations d'une même expérience sont possibles. Un exemple d'une telle interprétation alternative d'une même expérience PET-scan est donnée par les travaux de Cohen *et al.* sur les processus attentionnels⁶⁷. A partir d'un modèle connexionniste, Cohen *et al.* montrent que les résultats obtenus à l'aide du PET-scan et qui postulent l'existence de modules attentionnels spécifiques (comme un module de fixation de l'attention et un module de transfert de l'attention d'un objet à l'autre) peuvent être interprétés également à partir d'un réseau de modules interactifs dont les propriétés attentionnelles ne sont pas localisées dans tel ou tel module. En présentant ce modèle alternatif, ils montrent que le fonctionnement de réseaux interactifs simples, au niveau de la microstructure, peut donner l'apparence, au niveau de la macrostructure, de processus séquentiels et dissociables. Pour pouvoir trancher entre ces interprétations différentes, les expériences PET-scan devront être combinées avec des hypothèses supplémentaires sur la structure interne des modules (étude des structures) ou sur le rôle des modules dans le traitement de l'information (étude des fonctions).

3.1.2. Les modules interactifs

Les recherches sur les rapports structure - fonction fournissent des exemples frappants de processus cognitifs dissociables, à partir de l'étude

67 Cf. COHEN J.D., ROMERO R.D., SERVAN-SCHREIBER D. et FARAH M.J., "Mechanisms of Spatial Attention : The Relation of Macrostructure to Microstructure in Parietal Neglect", in *Journal of Cognitive Neuroscience*, 6 (1994), pp. 377-387.

des lésions, et de processus cognitifs additifs, à partir de l'imagerie cérébrale. Cependant, l'interprétation de ces expériences dépend d'hypothèses auxiliaires sur les fonctions cognitives observées et sur la structure interne des modules. En l'absence de telles hypothèses auxiliaires, l'observation de dissociations ou de processus additifs ne permet pas de conclure à la modularité, c'est-à-dire à l'indépendance des modules du système cognitif.

Dans cette section, nous voudrions présenter des expériences qui montrent que l'interactivité des composantes des systèmes cognitifs a été largement sous-estimée. Il s'agit d'expériences de psycho-physique et d'enregistrement de l'activité électrique de neurones sur des primates non-humains. En combinant les résultats de ces expériences avec les résultats précédents sur les dissociations et les processus additifs, on doit conclure à une organisation du système cognitif en modules interactifs. Dans une telle organisation, les processus d'ensemble sont autant déterminés par les tâches spécifiques de ces modules que par la dynamique de leur interaction.

3.1.2.1. L'interaction des modules dans le système visuel

Une première série d'expériences montrant l'interactivité sont les expériences de psycho-physique de V. S. Ramachandran que l'on a étudiées dans le premier chapitre. Ici, nous rappelons brièvement les résultats de ces expériences, en indiquant en plus leur pertinence par rapport à la discussion de l'hypothèse de la modularité. Ensuite, nous voudrions approfondir la position de Ramachandran à partir d'une discussion des avantages computationnels d'une théorie interactive de la vision.

Les expériences de Ramachandran concernent le traitement de stimuli visuels extrêmement simples, un processus dont on suppose traditionnellement qu'il est hautement modulaire. Cependant, de multiples interactions entre modules ont été mises en évidence. On se souvient par exemple des expériences sur l'effet d'entraînement de mouvement. Dans ces expériences, où l'on présente à un sujet un écran de points clignotants, un module de traitement de haut niveau, la construction d'une image globale de la scène, interagit avec un module de bas niveau, le traitement local du mouvement des points. Dans une autre expérience, on a observé une interaction entre deux modules appartenant à des modalités sensorielles différentes. Quand on applique un bruit de façon alternative aux deux oreilles et que l'on présente sur un écran un point clignotant et un carré, on a l'impression de voir le point disparaître à chaque clignotement derrière le carré. L'association entre le bruit et la scène visuelle crée l'illusion d'un point qui bouge de gauche à droite. L'information sur la signification de l'image visuelle peut également influencer des processus de traitement de bas niveau. Ainsi, dans une

expérience où l'on présente un masque concave (avec le nez pointant de l'autre côté) on perçoit néanmoins un masque convexe (avec le nez pointant vers nous) à partir d'une certaine distance d'éloignement. Ici la signification de l'image (je vois un visage) interfère avec la construction de la forme tridimensionnelle (un visage réel se présente toujours de façon convexe).

Les expériences de Ramachandran montrent de façon convaincante que le traitement de l'information visuelle ne se fait pas de façon séquentielle, dans une série où chaque module effectue une partie différente de la tâche globale, mais de façon interactive. On peut montrer qu'une telle organisation du système cognitif présente de nombreux avantages computationnels. Par exemple, comme nous l'avons vu dans le premier chapitre, l'interactivité permet de produire des réponses rapides au lieu de devoir construire des représentations détaillées et véridiques de la scène visuelle.

Dans une discussion sur les avantages computationnels de la vision interactive⁶⁸, Patricia Churchland, Ramachandran et Terrence Sejnowski montrent qu'il y a au moins quatre raisons pour lesquelles une organisation du système visuel en modules interactifs est plus avantageuse qu'une organisation hiérarchique en modules indépendants. Deux raisons concernent la simplification d'opérations particulièrement complexes comme la segmentation d'une image visuelle en figure - fond et la construction de l'image tridimensionnelle d'une scène. La première opération, la segmentation, est facilitée si on dispose d'informations sémantiques sur les objets présents dans la scène. Si on reconnaît les objets, on pourra plus facilement regrouper correctement les stimuli. Mais, comment peut-on reconnaître les objets si on n'a pas au préalable segmenté l'image, c'est-à-dire regroupé les traits en un tout cohérent ? Ce paradoxe est un des problèmes classiques d'une théorie de la segmentation qui suppose que la segmentation précède la reconnaissance. A partir d'un cadre interactif cependant, on peut supposer une collaboration entre le système de segmentation et le système de reconnaissance. Une segmentation partielle informe le système de reconnaissance et une reconnaissance partielle est envoyée en retour au système de segmentation qui procède à un regroupement plus fin des stimuli. Des problèmes complexes de segmentation peuvent alors être résolus de façon itérative en exploitant une collaboration entre le système de segmentation et le système de reconnaissance.

Les deux raisons suivantes invoquées par Churchland *et al.* pour montrer l'avantage de l'organisation interactive concernent le développement du système visuel. Aussi bien pendant la période d'établissement des

68 Cf. CHURCHLAND P.S., RAMACHANDRAN V.S. et SEJNOWSKI T.J., "A Critique of Pure Vision", *op. cit.*

connexions que pendant les périodes d'apprentissage par conditionnement opérant, des interactions entre modules de traitement jouent un rôle important.

L'établissement des connexions du système visuel ne se fait pas selon un processus préprogrammé, mais dépend également de l'activité du système visuel pendant le développement. Le mécanisme de développement supposé est évolutif : le cerveau crée d'abord un ensemble redondant de connexions et sélectionne ensuite les connexions utiles pour la poursuite de ses activités. Cette sélection peut se faire sans interaction notable entre les différents modules de traitement. Cependant, si on suppose que le même mécanisme de développement intervient dans la construction des processus visuels de haut niveau, Churchland *et al.* montrent qu'on doit postuler des mécanismes d'interaction réciproque entre ces processus et les processus de bas niveau.

Ensuite, pendant les périodes d'apprentissage par conditionnement opérant, une interaction entre modules semble également nécessaire. En effet, dans les expériences de conditionnement opérant, le système reçoit une récompense s'il associe la réponse correcte à un ensemble de stimuli dans son champ visuel. Mais pour savoir quels sont les stimuli qui, dans son champ visuel, sont pertinents par rapport à l'expérience, il doit déjà disposer d'une certaine interprétation de la scène visuelle. Des modules de détection de stimuli et des modules de traitement des significations doivent donc interagir au cours des expériences de conditionnement⁶⁹.

Les expériences de psycho-physique de Ramachandran et les arguments computationnels de Churchland *et al.* nous indiquent l'importance et l'utilité de l'interaction entre modules dans le système visuel. La dynamique d'interaction permet d'offrir des solutions globales à des problèmes cognitifs et éventuellement d'améliorer ces solutions en réitérant l'interaction entre les modules.

3.1.2.2. Les solutions globales de coordination du mouvement

L'image des modules interactifs que l'on obtient à partir des expériences sur le système visuel nous permet d'aller encore un pas plus loin. En effet, si les solutions globales offrent des avantages computationnels par rapport à une solution modulaire, peut-être les solutions globales sont-elles adoptées par le système avant la décomposition de l'information en parties distinctes ? C'est précisément ce que l'on peut mettre en évidence à partir d'expériences d'enregistrement de l'activité électrique de neurones isolés ou de groupes de

⁶⁹ Pour des références plus détaillées sur l'ensemble de ces arguments, voir l'article de Churchland *et al.* (*ibid.*, pp. 49–54).

neurones sur des primates non-humains. En étudiant de plus près ces expériences, nous voulons indiquer le rôle essentiel que joue l'interactivité dans l'organisation fonctionnelle du système nerveux.

Dans *Being There*, Andy Clark résume quelques-unes de ces expériences d'enregistrement d'activité électrique de neurones isolés. Il s'agit d'expériences faites par deux équipes de neuroscientifiques de la Washington University School of Medicine et qui portent sur la motricité et la perception visuelle du singe⁷⁰. Dans ce type d'expériences on implante — avec les précautions nécessaires — une électrode dans le cerveau d'un animal et on observe l'activité électrique à cet endroit. En utilisant des électrodes extrêmement fines on peut mesurer l'activité d'un groupe de quelques neurones ou même d'un neurone isolé.

Considérons d'abord la question suivante : comment est-ce que le cerveau du singe coordonne le mouvement des doigts ? Longtemps on a répondu à cette question à partir d'un modèle simple et intuitif. On supposait que dans une certaine région du cortex, l'aire 4 de Brodmann⁷¹, se trouvait une carte des mouvements : l'activation d'un groupe de neurones spécifiques sur la carte donne lieu au mouvement d'un doigt, l'activation d'un deuxième groupe au mouvement d'un autre, etc. Pour bouger plusieurs doigts à la fois ou pour bouger la main entière, il faudrait activer les différents groupes de neurones à la fois.

Ce modèle simple correspond à l'hypothèse de l'organisation modulaire du cortex moteur et il suppose que pour chaque composante élémentaire de mouvement, le mouvement d'un doigt, du coude, de la lèvre etc., on peut identifier un groupe de neurones qui contrôle ce mouvement. Cette hypothèse a donné lieu à la représentation homunculaire (de *homunculus* : petit homme) du cortex moteur, comme dans le célèbre dessin de Penfield que l'on a repris à la figure 2.4. Dans ce dessin on représente les fonctions motrices associées à chaque groupe de neurones de l'aire 4, de façon proportionnelle à l'importance de la région impliquée dans cette fonction.

Le modèle homunculaire du cortex moteur a été obtenu à partir de stimulations électriques localisées dans le cortex. Constatant un mouvement spécifique en fonction de l'endroit de cette stimulation, on a conclu à l'organisation modulaire. Cependant, comme le remarquent Boisacq-Schepens et Marc Crommelinck, la question de savoir si les neurones

70 Cf. CLARK A., *Being There*, *op. cit.*, pp. 130–135.

71 Pour la description des différentes zones du cerveau, nous utilisons la nomenclature de Boisacq-Schepens et Crommelinck, qui est celle de la cytoarchitectonie de Brodmann, cf. BOISACQ-SCHEPENS N. et CROMMELINCK M., *Neuro-psycho-physiologie*, Paris, 1994, t. 1, p. 40. L'aire 4 du cortex moteur est parfois appelée aussi l'aire motrice 1 ou M1, comme dans la discussion de Andy Clark.

stimulés contrôlent l'activité des muscles ou le mouvement reste une question ouverte. En effet, on ne peut pas savoir sur base de ces seules expériences si la représentation motrice modulaire est de nature anatomique (représentations de muscles) ou de nature fonctionnelle (les fonctions motrices)⁷². En fait, une autre série d'expériences dont nous donnerons le détail ci-dessous montrent qu'une même fonction motrice est contrôlée, non à partir d'un groupe de neurones, mais à partir de différents groupes de neurones qui s'activent de façon conjointe et qui se chevauchent partiellement⁷³. L'activation d'un neurone isolé correspond donc plutôt à une subdivision anatomique que fonctionnelle.

Figure 2.4. Modèle modulaire du cortex moteur (l'aire 4 de Brodmann) (source : PENFIELD W. et RASMUSSEN T., *The cerebral cortex of man*, Macmillan, 1950).

Regardons maintenant les expériences discutées par Andy Clark. D'après le modèle homunculaire, le mouvement de chaque doigt est contrôlé par un groupe spécifique de neurones. Une conséquence de ce modèle est qu'un mouvement coordonné de plusieurs doigts demande l'activation de plusieurs groupes de neurones en même temps, c'est-à-dire d'une région plus étendue du cortex moteur. Cependant cette prédiction n'est pas vérifiée. Schieber et Hibbard ont observé⁷⁴, à partir d'expériences sur la motricité chez

72 Cf. *ibid.*, p. 153.

73 Cf. *ibid.*, p. 153.

74 Cf. SCHIEBER M., "How Might the Motor Cortex Individuate Movements ?", in *Trends in*

le singe, que le mouvement de doigts isolés exige plus d'activité du cortex moteur que des mouvements élémentaires de la main entière. On observe donc exactement l'inverse de ce qu'on aurait attendu à partir du modèle homunculaire. D'après Schieber et Hibbard, on doit plutôt considérer le mouvement des doigts isolés comme le cas complexe et des mouvements d'ensemble de base, comme ouvrir et fermer la main entière, comme étant le cas simple. Ceci concorde mieux aussi avec les besoins naturels du singe, pour qui prendre des objets ou agripper des branches sont des activités primordiales de survie.

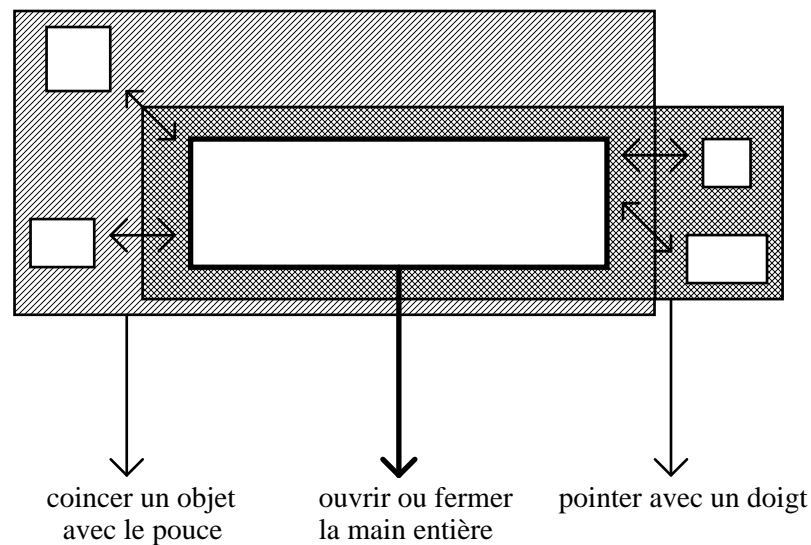


Figure 2.5. Utilisation d'un module de base impliquant le mouvement de tous les doigts (ouvrir ou fermer la main) par deux autres modules qui suppriment ou activent sélectivement des parties du module de base (figure de l'auteur).

Des expériences subséquentes ont montré que dans le cas des mouvements isolés, certains neurones du cortex moteur sont impliqués dans la prévention du mouvement des autres doigts. Schieber suggère qu'au cours de l'évolution naturelle se sont développées d'abord des structures neuronales coordonnant des mouvements d'ensemble et que seulement après se sont créés des circuits supplémentaires pour réaliser des mouvements plus fins. Ces circuits peuvent inhiber des parties du circuit contrôlant le mouvement

d'ensemble et en renforcer d'autres. Le modèle de l'organisation du contrôle moteur que l'on obtient est donc un modèle formé de quelques circuits de base pour le contrôle des mouvements d'ensemble, comme ouvrir ou fermer la main, et des circuits qui chevauchent partiellement ces circuits de base pour le contrôle des mouvements fins. On a représenté schématiquement cette organisation à la figure 2.5.

Les expériences sur la motricité de la main du singe nous montrent donc la présence de modules correspondant à l'encodage global des mouvements et à la décomposition de ce mouvement global en mouvements plus détaillés par intégration de ce module dans un circuit plus large. De cette façon, "la représentation motrice serait constituée par une mosaïque de foyers musculaires préférentiels se chevauchant partiellement et contrôlant, à la limite, une même articulation"⁷⁵.

Le chevauchement entre modules est également observé dans une deuxième série d'expériences discutée par Clark. Il s'agit des travaux de l'équipe de Van Essen sur le système visuel du singe. On sait que les neurones des aires visuelles du singe répondent de façon préférentielle à certains types de stimuli bien définis. Ainsi les neurones de l'aire V4 sont particulièrement sensibles aux couleurs et aux formes complexes, comme des formes en spirale ou en étoile, et les neurones de l'aire V5 (appelée aussi MT) aux directions de mouvement. Pour cette raison, les neurones de ces aires, qui correspondent à des traitements visuels de bas niveaux, ont été appelés des neurones détecteurs de traits (*feature detectors*).

Cependant des observations plus fines ont permis de montrer que ces réponses ou sensibilités maximales d'un neurone pour un certain stimulus ne spécifient pas entièrement le rôle de ce neurone dans le traitement de l'information. "Bien qu'une cellule puisse répondre de façon maximale à une structure en spirale, la même cellule répondra aussi à différentes autres structures"⁷⁶. En effet, Van Essen et Gallant ont observé que les neurones de l'aire V4 qui répondent de façon préférentielle à des formes en spirale ont une certaine sensibilité également pour des formes hyperboliques (croix et hyperboles) et pour quelques formes cartésiennes (droites et points), et que d'autres, qui répondent de préférence à des formes hyperboliques, conservent une certaine sensibilité aux formes en spirales et aux formes cartésiennes simples⁷⁷. Ces observations ont amené les chercheurs à traiter les cellules isolées non comme des détecteurs de traits, mais plutôt comme des filtres qui

75 BOISACQ-SCHEPENS N. et CROMMELINCK M., *op. cit.*, p. 153.

76 CLARK A., *Being There*, *op. cit.*, p. 135 ; notre traduction.

77 Cf. VAN ESSEN D. et GALLANT J., "Neural Mechanisms of Form and Motion Processing in the Primate Visual System", in *Neuron*, 13 (1994), pp. 1-10.

laissent passer sélectivement des informations. En répondant, par des degrés différents d'activation, un même neurone peut encoder plusieurs types d'information et participer à différents circuits de traitement d'information. Dans le cas des neurones isolés, on observe donc également un certain chevauchement entre les circuits de traitement de l'information. Cette image est encore renforcée par l'observation que les réponses de certains neurones peuvent être modulées en fonction de l'attention et que même les caractéristiques de réponse des neurones de V1 (la première aire visuelle dans le cortex visuel) peuvent dépendre de modifications du contexte⁷⁸.

Les expériences sur la motricité et le système visuel du singe nous donnent des exemples supplémentaires d'interactivité dans l'organisation des systèmes cognitifs. Pour terminer cette section sur l'interactivité dans l'organisation fonctionnelle des systèmes cognitifs, résumons brièvement les différents arguments avancés. A partir des expériences de Ramachandran on a vu que l'interaction produit des illusions motrices ou des formes tridimensionnelles. Ces illusions s'observent dans des situations expérimentales artificielles, quand on isole le système cognitif de son milieu de fonctionnement naturel. Dans les situations habituelles, l'interaction entre les modules produit une image approximativement correcte de la réalité, qui permet de guider efficacement le comportement. Ensuite, à partir des arguments de Churchland *et al.*, on a exploré quelques avantages computationnels de l'interactivité. A côté de la capacité de réponse rapide, comme dans les exemples de Ramachandran, l'interactivité permet aussi de simplifier les calculs par un mécanisme itératif. Ainsi, on a vu que la segmentation d'une image visuelle était facilitée si on utilise une information sur la signification de la scène visuelle, information qui doit elle-même se baser sur une première idée de la segmentation. En itérant entre le module de segmentation et le module de reconnaissance des objets, on peut affiner

78 D'après les expériences discutées ici, un groupe de neurones d'une aire visuelle (expériences de Van Essen et Gallant) ou d'un centre moteur (expériences de Schieber et Hibbard) peut être utilisé dans différentes fonctions perceptives ou motrices. Ces observations ont amené certains chercheurs à postuler l'existence dans le cerveau de régions d'activité, appelées zones de convergence, dont le seul rôle est d'envoyer des signaux à ces groupes de neurones pour "organiser le travail" : par des signaux d'activation, de suppression et de modulation de l'activité de ces groupes de neurones, ceux-ci pourront intervenir tantôt dans telle, tantôt dans telle autre fonction de traitement de l'information (cf. VAN ESSEN D., ANDERSON C. et OLSHAUSEN B., "Dynamic Routing Strategies in Sensory, Motor and Cognitive Processing", in KOCH C. et DAVIS J. (eds.), *Large-scale Neuronal Theories of the Brain*, Cambridge (MA), 1994 ; DAMASIO A. et DAMASIO H., "Cortical Systems for Retrieval of Concrete Knowledge : The Convergence Zone Framework", in KOCH C. et DAVIS J. (eds.), *Large-scale Neuronal Theories of the Brain*, Cambridge (MA), 1994).

progressivement la réponse. D'après Churchland *et al.*, ces deux mécanismes d'interactivité, l'interaction simple pour les réponses rapides et l'interaction itérative pour des calculs complexes, se retrouvent également dans les processus de développement et d'apprentissage du système visuel. Finalement les expériences de stimulation et d'enregistrement d'activité électrique de neurones nous ont fourni des informations sur l'organisation fine et locale des modules. On a observé à ce niveau l'encodage de réponses globales de façon distribuée dans des modules complexes et le chevauchement entre modules qui se greffent sur ces modules de base.

3.2. Le couplage fonctionnel et la critique par les théories de la clôture opérationnelle

Dans les expériences que nous avons discutées dans la section précédente, nous avons toujours rencontré une combinaison de deux éléments : l'interactivité et la présence de modules de traitement de l'information. D'un côté, l'observation des dissociations en neuropsychologie et l'isolation des processus additifs en imagerie cérébrale indiquent une certaine modularisation, mais n'excluent pas l'interactivité. De l'autre, les analyses interactives de Ramachandran, Schieder et Van Essen montrent l'importance de l'interactivité dans le cerveau, mais elles s'appuient également sur l'existence de modules séparés de traitement et sont donc compatibles avec un certain degré de modularisation. A partir de ces expériences, nous pouvons conclure à une organisation fonctionnelle des systèmes cognitifs en modules interactifs.

Cette image interactive a évidemment une conséquence sur la position fonctionnaliste. Contrairement à l'analyse de Marr, on ne peut plus dire que la fonction détermine entièrement la signification des représentations. Le champ des représentations a son autonomie propre à cause de la dynamique autonome d'interaction entre les modules, et la signification de ce champ ne se réduit donc pas au rôle fonctionnel des représentations par rapport à une fonction computationnelle d'ensemble. Dans l'étude des représentations il faudra donc croiser la perspective fonctionnelle avec une étude de la cohérence ou de la topologie interne du champ des significations. Ce croisement permet d'envisager une méthode de modélisation qui convient aux propriétés de l'intentionnalité telles qu'on les a mises en évidence à partir de la théorie écologique.

La relative indépendance du contexte sémantique par rapport aux fonctions aura des conséquences en retour sur la modélisation des fonctions dans les processus cognitifs. En effet, dans l'image interactive, l'état final du système sera autant déterminé par la fonction à réaliser que par la dynamique

d'interaction qui détermine le chemin particulier qui sera suivi pour réaliser la fonction. Au lieu de fournir le programme détaillé du processus, comme dans l'approche computationnelle de Marr, la fonction aura donc davantage un rôle de stabilisation d'un processus qui a sa dynamique propre.

Dans cette deuxième section, nous analyserons une deuxième perspective qui remet en question la possibilité d'une méthode de modélisation contextuelle de l'intentionnalité, la perspective de la clôture opérationnelle. Cette perspective accepte l'importance du contexte sémantique dans l'étude des opérations cognitives, mais pousse l'autonomie des significations jusqu'à éliminer tout lien fonctionnel avec l'environnement. Elle part de l'hypothèse que les interactions dans les systèmes cognitifs sont tellement fortes que la construction des représentations est uniquement déterminée par la dynamique de ces interactions. La stabilisation du système se fait par un processus d'auto-organisation et on ne doit plus faire appel au rôle stabilisateur de l'environnement. Cependant, comme nous le verrons dans cette section, l'hypothèse de la clôture n'est vérifiée que localement. Plutôt que de fermer la porte à toute perspective interactive, l'idée d'une autonomie de la dynamique représentationnelle, qui est au centre des théories de la clôture, laisse encore la possibilité d'une voie alternative pour la stabilisation des processus cognitifs — à côté de la stabilisation interne —, qui est la stabilisation en couplage fonctionnel avec l'environnement.

3.2.1. Définir l'hypothèse de la clôture opérationnelle

Pour caractériser l'hypothèse de la clôture opérationnelle, nous partirons des travaux de Heinz von Foerster sur la clôture. Heinz von Foerster est un des personnages clefs du mouvement de la seconde cybernétique, mouvement dont l'objet était le développement d'une théorie générale des systèmes autour de concepts centraux comme l'auto-organisation, la non-linéarité ou l'auto-référentialité. Ses travaux ont inspiré de nombreux chercheurs en biologie et en neurosciences comme Humberto Maturana, Francisco Varela et Henri Atlan, et également de nombreux chercheurs en sciences sociales comme Gordon Pask, Gregory Bateson ou Niklas Luhmann pour n'en citer que quelques-uns⁷⁹.

La clôture opérationnelle ou systémique d'un système peut se définir de la façon suivante : une entité est un système clos quand l'interaction des

79 Sur le mouvement de la seconde cybernétique, voir par exemple HEIMS S.J., *The Cybernetics Group*, Cambridge (MA), 1991 ; VAN DE VIJVER G., *Van cybernetica naar connectionisme, op. cit.*, ou SÉGAL L., *Le rêve de la réalité. Heinz von Foerster et le constructivisme*, trad. par Anne-Lise Hacker, Paris, 1990.

éléments qui le constituent engendre précisément ces éléments à l'intérieur de ce système⁸⁰. On peut illustrer cette définition à partir d'un exemple simple de système clos, un système chimique autocatalytique, que l'on a représenté à la figure 2.6.

Figure 2.6. Un système chimique autocatalytique (source : SÉGAL L., 1986, p. 162).

Le système de la figure 2.9. représente la production d'un composé instable AB par l'intervention d'un catalyseur C et la décomposition, en fin de processus, de ce composé en A et B. L'intervention du catalyseur C est nécessaire parce que A et B ne réagissent pas l'un avec l'autre. Cependant, AC peut réagir avec B pour former ABC. En rejetant C on obtient alors AB. Ce système présente clairement un caractère cyclique. La composante C est le produit de la décomposition de ABC en AB et C, mais intervient elle-même dans la production de ABC dont elle provient par la réaction de AC avec B. C'est cette caractéristique qui en fait un système opérationnellement clos.

Les processus circulaires décrits par von Foerster se rencontrent typiquement dans l'organisation interne des êtres vivants. Dans une cellule organique par exemple, le matériel génétique, l'ADN, contient le code pour la production des enzymes. Mais ces mêmes enzymes sont responsables pour la régulation de la reproduction de l'ADN et de la transcription de l'information génétique vers les sites de production des enzymes. Un autre exemple est la relation entre la cellule et la membrane cellulaire. D'un côté

80 Cf. SÉGAL L., *op. cit.*, p. 162.

l'existence et le fonctionnement de la cellule n'est possible que grâce à la présence de la membrane cellulaire. De l'autre, la membrane est constituée à partir du fonctionnement de la cellule. L'observation de telles circularités dans les êtres vivants a amené deux biologistes chiliens, Humberto Maturana et Francisco Varela, à appliquer l'idée de la clôture opérationnelle à l'organisation du vivant. Ils caractérisent la clôture opérationnelle des systèmes vivants par le concept d'autopoièse, du grec *auto* (soi-même, par soi-même) et *poiesis* (création, production) : les systèmes vivants sont des systèmes qui produisent eux-mêmes leur propre organisation⁸¹.

3.2.2. La clôture dans les systèmes cognitifs

Que devient l'idée de la clôture dans le cas des systèmes cognitifs ? Remarquons d'abord que du point de vue de la théorie de la clôture, l'hypothèse de la clôture du système cognitif se résume à l'hypothèse de la clôture du système nerveux, étant donné que l'on ne peut plus parler de représentations de l'environnement, mais uniquement de corrélations neuronales internes. La clôture du système nerveux est illustrée par Maturana et Varela à partir du système sensori-moteur de la grenouille⁸². Pour montrer que le fonctionnement du système sensori-moteur de la grenouille est simplement une expression de corrélations internes et pas un système de représentation - action, ils considèrent l'expérience étonnante suivante, due à Sperry⁸³. Par une intervention minutieuse, on peut retourner l'oeil du têtard, la larve de la grenouille, de 180 degrés, sans abîmer le nerf optique. La larve, et ensuite la grenouille, se développe normalement après cette opération. Si maintenant on recouvre l'oeil opéré de la grenouille adulte et que l'on présente une mouche dans son champ visuel, alors la grenouille sort sa langue et attrape correctement la mouche. Cependant, si on couvre l'oeil normal et qu'on présente une seconde fois une mouche, alors la grenouille envoie sa langue précisément avec une déviation de 180 degrés. On peut répéter cette expérience plusieurs fois, chaque fois on observera la même déviation. La grenouille envoie sa langue comme si la partie de la rétine sur laquelle se forme l'image de la proie se trouvait en position normale et n'était pas retournée de 180 degrés.

Cette expérience montre de façon convaincante que pour le système visuel de la grenouille il n'y a pas de haut et de bas, de devant et d'arrière en

81 Cf. *ibid.*, p. 164.

82 Cf. Maturana H. et Varela F., *De boom der kennis. Hoe wij de wereld door onze eigen waarneming creëren*, trad. par T. Maas et C. Oks, Amsterdam, 1988, pp. 100-101.

83 Cf. *ibid.*, p. 100.

relation avec le monde externe, même si ces catégories spatiales sont utilisées par un observateur qui analyse les relations entre la grenouille et son environnement. Il y a uniquement une corrélation interne entre une perturbation sur la rétine et les contractions musculaires qui font bouger la langue. Le fonctionnement du système nerveux est l'expression de sa cohésion interne et le comportement est causé par ce système de relations internes.

Pour mieux comprendre le changement de perspective, du traitement de l'information à la clôture opérationnelle, invoqué par Maturana et Varela considérons un autre exemple, donné par von Foerster. "Quand j'écris avec une craie sur le tableau, je dois établir un lien entre deux sources de données sensorielles, mon oeil et ma main"⁸⁴. Pour établir ce lien, on utilise des connexions neuronales intermédiaires, entre la surface rétinienne et la sensation du mouvement dans les cellules proprioceptives. Tout changement au niveau de la rétine cause des changements au niveau de la correction du mouvement et tout mouvement cause des changements sur la rétine. C'est le système des corrélations intérieures qui produit les traces de craie au tableau.

On objectera peut-être qu'il n'y a pas de différence entre l'image du système cognitif dans les termes de corrélations sensori-motrices internes et le représentationnalisme classique. En effet, il suffit de supposer qu'au niveau des connexions neuronales intermédiaires le système construit des représentations du monde, pour retrouver le représentationnalisme. Cependant, pour von Foerster, Maturana et Varela, le langage représentationnel n'a de sens que pour un observateur extérieur qui interprète le comportement⁸⁵. Tout ce que fait le système cognitif, c'est maintenir un ensemble de corrélations internes. On peut comparer une telle situation à un aviateur qui pilote son avion aux instruments⁸⁶. Même si pour un observateur extérieur le pilote est par exemple en train d'atterrir dans un brouillard épais, tout ce que fait le pilote, qui ne voit rien à cause du brouillard, c'est manipuler un certain nombre de relations internes dans l'avion, de façon à ce que les variables mesurées par les instruments prennent les valeurs souhaitées. De même, dans

84 SÉGAL L., *op. cit.*, p. 170.

85 Une des conséquences de l'hypothèse de la clôture opérationnelle est donc que les représentations ne sont pas des traits objectifs du système, mais seulement des projections d'un observateur extérieur, qui interprète le fonctionnement interne du système dans les termes d'une fonction de traitement d'information qui serait réalisée par le système. Comme dans l'interprétation kantienne du concept de finalité, l'attribution de fonctions appartiendrait à un jugement réfléchissant et pas au jugement déterminant de la science (les fonctions ne joueraient pas de rôle *déterminant* dans la stabilisation du système).

86 Cf. *ibid.*, p. 167.

le cas du système nerveux central, il n'y aurait pas de traitement sémantique d'information de l'environnement, mais uniquement le maintien d'une organisation interne.

L'image de l'architecture du cerveau proposée par Maturana, Varela et von Foerster — et que l'on retrouve également dans les travaux d'Atlan — a clairement une certaine plausibilité. Il est très probable que des boucles sensori-motrices, avec un plus ou moins grand nombre de connexions intermédiaires, jouent un rôle important dans le cerveau et que pour ces processus on puisse considérer l'interaction avec l'environnement comme une simple perturbation des corrélations internes et pas comme une source d'information. Un exemple de telles boucles dont on connaît relativement bien la neurophysiologie est le réflexe vestibulo-oculaire⁸⁷. La fonction de ce réflexe consiste à stabiliser le champ visuel sur la rétine malgré les déplacements de la tête. Il produit des mouvements oculaires de compensation de même vitesse que le mouvement de la tête mais dirigés en sens opposé. De la sorte, le regard, somme du mouvement de l'oeil dans l'orbite et du mouvement de la tête, est stable dans l'espace.

Pour montrer le caractère purement interne de cette coordination visuo-motrice, des chercheurs ont étudié l'effet sur le système visuel de distorsions provoquées par le port de lunettes grossissant, réduisant ou encore inversant l'image rétinienne. Après quelques jours d'expérience active avec les lunettes, le monde paraît à nouveau stable et les coordinations visuo-motrices sont parfaitement recalibrées⁸⁸. Le système s'est adapté pour maintenir la valeur de la variable qui mesure le glissement de l'image rétinienne égale à zéro. Le port de lunettes provoque une modification de l'activité de certaines synapses, au niveau du cervelet et du tronc cérébral⁸⁹, dans le seul but de rétablir les invariances organisationnelles du système. Si on compare cet exemple à l'exemple de la grenouille, on peut voir également comment les systèmes cognitifs peuvent se complexifier en ajoutant des boucles sensori-motrices supplémentaires. La grenouille dispose seulement de la boucle réflexe qui se situe entre sa rétine et le centre de mouvement de la langue, mais n'a pas de système de stabilisation du regard. Si, en plus de cette première boucle, elle disposait d'un réflexe vestibulo-oculaire plastique, comme c'est le cas chez l'être humain, elle pourrait produire un comportement approprié, même après que l'on aura retourné son oeil de 180 degrés.

Les exemples de clôture du système nerveux donnés par Maturana, Varela et von Foerster concernent la présence de boucles sensori-motrices

87 Cf. BOISACQ-SCHEPENS N. et CROMMELINCK M., *op. cit.*, t. 1, p. 73 et t. 2, p. 220.

88 Cf. *ibid.*, p. 220.

89 Cf. *ibid.*, p. 222.

dans le système nerveux. Cependant, à partir de ces exemples, ils avancent également la thèse que le principe de la clôture est un principe général d'architecture du système nerveux⁹⁰. Si on considère les expériences et les différents modèles que l'on a rencontrés dans le premier chapitre, le rejet des représentations en général, et l'extension de l'interactivité à l'ensemble du système, ne paraissent pourtant pas justifiés. On se souvient des expériences d'Alain Berthoz sur les représentations spatiales. Dans ce cadre, on a distingué entre les référentiels égocentriques, prenant comme point de référence de la représentation spatiale un endroit du corps, et les représentations allocentrées, prenant comme point de référence un objet de l'environnement. L'existence des référentiels allocentrés montre clairement la possibilité de construire des représentations de l'environnement. L'existence de différents types de représentations est confirmée également par des expériences en éthologie animale. Les babouins et les chimpanzés par exemple sont capables de construire des représentations très élaborées de leur environnement à la fois physique et social. Des représentations spatiales leur permettent de calculer à l'avance le trajet le plus court pour atteindre une destination, et des représentations sociales les aident à se situer dans la hiérarchie des relations de domination⁹¹. La clôture opérationnelle ne semble donc pas pouvoir se généraliser à l'ensemble du système nerveux.

Les boucles sensori-motrices fermées ne sont qu'un cas particulier d'organisation rencontrée dans les systèmes cognitifs. Pour mieux définir ce type d'organisation, on peut comparer les boucles fermées à un autre exemple de boucles sensori-motrices, étudié au premier chapitre : les sous-systèmes sensori-moteurs des agents autonomes de R. Brooks. On se souvient que les agents autonomes, de petits robots mobiles, sont composés de sous-systèmes sensori-moteurs indépendants et capables d'exécuter des tâches simples sans utiliser un système de représentation de leur environnement. Comme illustration du principe d'architecture des agents autonomes, Brooks donne l'exemple d'un petit robot simple dont la tâche est d'avancer sans se heurter aux objets de l'environnement⁹². Dans une architecture classique on munirait le robot d'un sous-système perceptif, un sous-système moteur et un sous-système central de contrôle, qui planifie et contrôle le mouvement. Cependant la tâche de ce robot peut également être réalisée sans contrôle centralisé, à partir de deux boucles sensori-motrices extrêmement simples, une qui déclenche un mouvement dans une direction

90 Cf. MATURANA H. et VARELA F., *op. cit.*, pp. 131–132.

91 Cf. VAUCLAIR J., *L'intelligence de l'animal*, Paris, 1992, chapitre 3 : l'usage d'outils et la représentation de l'espace, chapitre 4 : la cognition sociale.

92 Cf. VARELA F., THOMPSON E. et ROSCH E., *op. cit.*, p. 285.

aléatoire en absence d'obstacles et une qui arrête le robot si on touche un obstacle. A partir de ces deux boucles simples le robot produit le comportement désiré. L'architecture de ces agents autonomes est caractérisée en toute généralité par la décentralisation et l'orientation vers l'action des sous-systèmes : chaque sous-système a un fonctionnement indépendant et produit une activité qui connecte à sa manière la sensation à l'action⁹³. On retrouve ces principes également dans le robot à six pattes d'Espenschied. Les agents autonomes sont uniquement constitués de boucles connectant la sensation à l'action. Mais est-ce qu'ils constituent pour autant des illustrations du principe de la clôture ? Nous pensons que non. En effet, comme on l'a montré au premier chapitre, la cohérence entre les différents sous-systèmes est établie par leur interaction avec un environnement opératoire commun. L'environnement ne joue donc pas simplement le rôle d'une perturbation extérieure, mais intervient activement dans l'établissement de cette cohérence. La simplicité des solutions aux problèmes moteurs proposées par les théoriciens des agents autonomes ne tient donc pas à la clôture des systèmes qu'ils construisent, mais — au contraire — à l'intégration ouverte de ces systèmes dans un environnement opératoire adéquat⁹⁴.

Les boucles sensori-motrices directement opératoires correspondent bien au cas de la clôture ; on retrouve le rejet de la notion classique de représentation et le caractère fermé du système sur lui-même. Cependant, une théorie de la cognition qui se limite à une telle analyse des boucles opératoires, comme la théorie de la clôture opérationnelle, reste purement subjective. Elle ne parvient pas à intégrer l'ouverture constitutive des systèmes par rapport à la dynamique environnementale. Mais, la définition par le système de son propre contexte de significations, telle qu'on l'a vu dans la section précédente, n'implique pas nécessairement une fermeture par rapport à la cohérence opératoire de l'environnement. Au contraire, comme notre discussion des modèles de la clôture le montre, cette anticipation contextuelle est tout à fait compatible avec l'inscription des opérations cognitives dans un environnement opératoire autonome. La prise en compte de cette inscription permet alors d'entrevoir une pratique de modélisation qui

93 Cf. CLARK A., *Being There*, *op. cit.*, p. 13.

94 L'ouverture n'implique pas encore une ouverture flexible à un environnement dynamique et changeant. Pour construire des systèmes représentationnels flexibles, il faut en plus une certaine modularisation, de sorte que dans un contexte systématiquement nouveau on peut amender le système de façon locale, un module à la fois, sans influencer l'ensemble des connexions dans le système. Ou, pour utiliser le concept introduit au premier chapitre, il faut que le système soit capable d'une généralisation contextuelle transformatrice (cf. CLARK A., *Associative Engines*, *op. cit.*, p. 73).

articule l'anticipation contextuelle de significations avec l'inscription de cette anticipation dans la dynamique opératoire autonome du contexte.

Chapitre 3

Limitations internes et limitations contextuelles de la modélisation de l'intentionnalité

1. Introduction

Dans les deux premiers chapitres de cet ouvrage, nous avons examiné la cohérence d'une opération de modélisation de l'intentionnalité qui tient compte de la problématique de l'origine des significations et des visées intentionnelles. Cette opération de modélisation se distingue d'une approche purement computationnelle ou opératoire de la cognition, en ce qu'elle vise à situer l'effectivité des opérations cognitives dans leur contexte sémantique et fonctionnel. Comme nous l'avons montré, c'est seulement en dépassant les clôtures trop facilement opérées entre l'étude de l'intentionnalité tantôt comme opération mécanique (algorithme), tantôt comme orientation vers un but (comme fonction) ou encore tantôt comme source de significations (comme représentation) que l'on peut obtenir une opération de modélisation spécifique à l'étude de l'intentionnalité. Nous avons défini cette opération de modélisation dans les termes d'une théorie écologique de la cognition, à partir des notions de computation, de représentation et de couplage fonctionnel.

L'opération de modélisation de l'intentionnalité que nous avons ainsi définie se distingue de deux autres approches naturalistes de l'intentionnalité, le fonctionnalisme et les théories de la clôture opérationnelle. D'un côté, le fonctionnalisme intègre dans la pratique de modélisation l'articulation des opérations et des fonctions, mais évacue la question de la création des significations. De l'autre côté, les théories de la clôture opérationnelle étudient l'émergence des significations dans les systèmes cognitifs à partir des opérations, mais elles supposent une autostabilisation des significations sans lien fonctionnel avec l'environnement¹. Toutefois, l'étude naturalisée de

1 Comme nous l'avons vu, dans la perspective de la clôture opérationnelle, l'origine des significations est purement interne. L'attribution d'une relation fonctionnelle entre les significations et l'environnement dépend d'un observateur extérieur qui *interprète* la relation entre les significations et l'environnement dans ses propres termes. Ainsi, en

l'intentionnalité telle qu'on l'a définie à partir de la théorie écologique suppose non seulement de dépasser le mécanisme algorithmique, comme c'est déjà le cas dans le computationnalisme classique de David Marr, mais également de croiser le computationnalisme avec les perspectives ouvertes par les théories de la clôture et avec l'étude de la stabilisation des systèmes en couplage avec leur environnement. A partir de ce croisement peut alors s'envisager une pratique de modélisation particulière, que nous avons désignée par le terme de la modélisation contextuelle, qui se distingue autant des paradigmes d'inspiration fonctionnaliste, que des paradigmes d'inspiration biologique.

Notre examen de la modélisation de l'intentionnalité en sciences cognitives n'en est pas pour autant terminé. Jusqu'ici nous avons uniquement obtenu une spécification formelle de l'opération de modélisation. En effet, même si la modélisation contextuelle à laquelle nous aboutissons ne s'intéresse pas uniquement aux opérations computationnelles, mais également aux représentations et aux interactions fonctionnelles, cela ne change rien au fait que les représentations et les fonctions sont toujours conceptualisées dans des modèles mathématiques formels. Par exemple, le concept de représentation dans les réseaux de neurones est défini dans les termes d'une corrélation statistique entre des catégories de signification et une distribution de données présentées à l'entrée du réseau. D'autre part, le concept de couplage fonctionnel est défini dans les termes des principes de la cybernétique et de la théorie des systèmes émergents. En examinant la cohérence de l'opération de modélisation de l'intentionnalité de façon interne aux modèles mathématiques formels du champ des sciences cognitives, nous avons occulté la question des conditions de possibilité d'une modélisation formelle de l'intentionnalité.

La question que nous nous poserons dans ce troisième chapitre sera alors celle des limitations de la modélisation de l'intentionnalité en sciences cognitives. Dans l'introduction générale, nous avons défini, suite aux réflexions d'Atlan, un premier niveau de limitation, la limitation naturaliste, qui concerne la réduction méthodologique générale à laquelle procèdent les sciences cognitives. Cette réduction porte sur le traitement de la conscience intentionnelle comme "croyance effective" ou comme "procédure effective" et renvoie à une abstraction méthodologique qui consiste à mettre entre parenthèses l'expérience subjective. Atlan situe cette réduction dans le même

reprenant l'exemple donné par Henri Atlan, l'attribution d'une fonction de reconnaissance de formes à un système auto-organisé est le résultat "d'une projection de nos propres expériences cognitives de reconnaissance de formes sur le comportement observé" (ATLAN H., "Projet et signification", *op. cit.*, p. 280).

cadre méthodologique que les sciences de la nature. Toutefois, par là, il n'interroge pas le pouvoir spécifique qu'ont les modèles formels élaborés au sein de l'attitude naturelle à investir des domaines spécifiques de modélisation, comme le domaine des opérations intentionnelles, dans leur différence spécifique avec le domaine de la biologie ou de la physique. Afin de pouvoir définir les limitations de la modélisation de l'intentionnalité, il faudra donc déplacer la position de Atlan sur les limitations, afin de prendre en compte également la capacité sémantique qu'ont les modèles formels d'investir des champs particuliers de modélisation. En effet, la connexion qui peut s'établir entre un formalisme donné et le domaine d'application auquel il se rapporte peut se déployer à un double niveau. Tout d'abord, à l'interne, le formalisme anticipe dans son auto-compréhension un domaine d'application. Cette auto-compréhension définit un équilibre entre la fécondité sémantique du formalisme d'une part, sa capacité à investir des domaines nouveaux, et la nécessité de se rapporter à des domaines concrets du réel de l'autre. Elle constitue la limitation interne de la fécondité sémantique du formalisme. Cette limitation est bien prise en compte par Atlan à partir de la limitation naturaliste entendue comme méthode d'abstraction spécifique à l'entreprise de modélisation de l'intentionnalité. Ensuite, il faut encore prendre en compte les conditions qui interviennent au niveau de l'application du formalisme au réel. Ces conditions déterminent l'extension du formalisme ou encore sa limitation contextuelle. Par l'examen de cette deuxième limitation, nous voulons mettre en évidence la limitation spécifique de la modélisation de l'intentionnalité, au lieu d'en rester, comme Atlan, à une stabilisation autoréférentielle du processus de généralisation formelle.

Afin de construire de façon précise le dépassement de la position de Atlan nous procéderons, dans le chapitre qui suit, en trois étapes. En un premier temps, nous donnerons une définition plus précise des deux niveaux de limitation à partir d'une analyse des propriétés sémantiques de la modélisation. Ensuite, nous appliquerons cette analyse à la modélisation de l'intentionnalité en sciences cognitives afin de mettre en évidence la limitation contextuelle. Finalement, nous tenterons de montrer qu'il faut dépasser ce premier point de vue structurel sur la limitation contextuelle afin d'éviter une reprise formaliste et non réflexive de l'opération de contextualisation sur laquelle elle s'appuie.

Pour le premier point, nous partirons d'un aperçu critique du modèle standard empirico-formel en philosophie des sciences². D'après le modèle standard, la modélisation en sciences est soumise à la double exigence de la

2 Cf. SUPPE F., *The Semantic Conception of Theories and Scientific Realism*, Urbana/Chicago, 1989.

cohérence logique et de l'adéquation empirique. A partir de ce modèle, les limitations liées à la formalisation de l'intentionnalité en sciences cognitives seraient les mêmes que toute autre approche naturalisée du réel : l'application des formalismes est soumise à une contrainte logique et à une contrainte empirique. Toutefois, comme nous le verrons, ce modèle ne rend compte que de façon très imparfaite des limitations internes et contextuelles de la modélisation, liées à la dimension sémantique des modèles. C'est ce qui nous amènera à élargir le modèle standard à partir de la conception sémantique des théories scientifiques. La conception sémantique est une approche contemporaine de la science qui met au centre de l'entreprise scientifique le concept de modèle. On retrouve cette conception de façon explicite chez des auteurs comme Frederic Suppe et Bas Van Fraassen³ ou chez des auteurs proches de cette conception, sans s'en réclamer de façon explicite, comme Mario Bunge ou Jean Ladrière⁴. A partir de ces travaux, nous étudierons la question des limitations liées à la possibilité d'utiliser les formalismes au-delà de leur domaine normal d'interprétation, ou encore la question des limitations de la productivité sémantique.

Pour le deuxième point, il faudra retourner à la question spécifique des limitations de la modélisation de l'intentionnalité en sciences cognitives. Tout d'abord, il s'agira de mettre en évidence en sciences cognitives le rôle de la fécondité sémantique des modèles mathématiques formels, comme les modèles cybernétiques ou les modèles des réseaux de neurones, et la limitation interne de cette fécondité, liée au niveau de généralité de ces modèles. Ensuite, il faudra définir leur relation à un domaine d'application, c'est-à-dire leur relation aux modèles particuliers de la théorie et la limitation contextuelle spécifique liée à cette relation. Comme nous le montrerons, la spécification d'un domaine d'application pour la modélisation de l'intentionnalité s'appuie — à la différence de la modélisation en physique — sur des hypothèses auxiliaires qui portent sur la relation entre le système et son environnement opératoire. A cause de la présence de ces hypothèses auxiliaires, il ne sera pas possible de trancher entre les différents modèles de l'intentionnalité par des méthodes de recherche scientifique seules et donc de spécifier le domaine d'application des principes formels de façon univoque.

Finalement, la prise en compte de la limitation contextuelle a également des conséquences sur l'abstraction méthodologique mise en évidence par

3 Cf. SUPPE F., *The Semantic Conception of Theories*, op. cit. ; VAN FRAASSEN B., *The Scientific Image*, London, 1980.

4 Cf. BUNGE M., *Method, Model and Matter*, Dordrecht/Boston, 1973 ; LADRIÈRE J., *Les enjeux de la rationalité. Le défi de la science et de la technologie aux cultures*, Paris, 1977.

Atlan et qui caractérise le projet des sciences cognitives dans leur ensemble. En effet, nous verrons que la mise en évidence de l'autonomie de la construction contextuelle des modèles particuliers amène une question en retour sur le projet de la formalisation. L'équilibre contextuel entre l'opération formelle et les modèles particuliers ne peut se stabiliser qu'en prenant en compte l'activité réflexive qui constitue l'horizon d'effectuation du projet de la formalisation et dont le pouvoir de modalisation permet d'ouvrir en même temps vers de nouveaux équilibres contextuellement opérants. Dans le dernier paragraphe de ce chapitre, nous tenterons de montrer comment le modèle de la phénoménologie génétique permet de prendre en compte un tel traitement réflexif et modal de la contextualisation.

2. La critique du traitement standard de la limitation

Dans ce premier paragraphe, nous abordons la question des limitations des formalismes mathématiques dans le cadre général d'une réflexion sur la modélisation en sciences. Tout d'abord, dans une première section, nous présentons le modèle standard empirico-formel des théories scientifiques, issu de l'empirisme logique du cercle de Vienne. Deux difficultés importantes auxquelles le modèle standard est confronté indiquent l'insuffisance de cette approche pour rendre compte des conditions d'application du formalisme, le problème de la charge théorique de l'observation et le problème de la présence dans les théories de termes sans correspondance empirique. Ensuite, dans une deuxième section, nous montrerons qu'une prise en compte de l'importance du rôle des modèles en sciences permet de donner une réponse plus satisfaisante au problème des limitations. Pour cette deuxième section nous nous baserons sur la conception sémantique des théories scientifiques telle qu'on la trouve chez des auteurs comme Frederic Suppe ou Bas Van Fraassen.

2.1. Du modèle standard à la conception sémantique

Le modèle standard des théories scientifiques présente les théories sous la forme d'énoncés d'un langage formel dont l'interprétation est fixée par des règles de correspondance, comme on l'a représenté schématiquement à la figure 3.1⁵. D'après ce modèle, on peut classer les termes des théories

5 Cf. SUPPE F., *The Semantic Conception of Theories*, *op. cit.*, p. 38. Suppe utilise le terme *conception reçue* au lieu de *modèle standard*. En utilisant le terme *modèle standard* nous adoptons l'usage de l'encyclopédie *Enzyklopädie Philosophie und Wissenschaftstheorie*, qui est également l'usage le plus courant (Cf. MITTELSTRAß J. (ed.), *Enzyklopädie*

scientifiques en deux sous-ensembles, les termes théoriques et les termes empiriques. Les règles de correspondance permettent d'associer aux termes du langage théorique, définis dans le langage formel, des termes du langage empirique ou observationnel. Par exemple, au concept théorique de température, définie comme le paramètre d'équilibre de la fonction de distribution d'un système de particules en interaction, on peut associer le concept empirique de température, la propriété que l'on mesure au moyen d'un thermomètre.

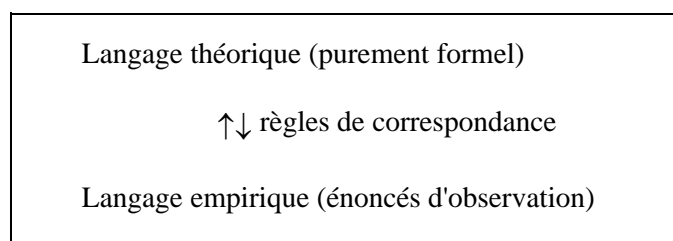


Figure 3.1. Le modèle standard de l'empirisme logique (figure de l'auteur).

Les auteurs de la version originale du modèle standard dans l'empirisme logique défendaient la thèse que le contenu des théories scientifiques peut être entièrement fixé par l'observation des faits empiriques, observations décrites dans les énoncés du langage empirique. D'après leur conception, le contenu non empirique de la théorie est purement syntaxique, défini à partir d'axiomes de base et de règles de déduction. Il n'apporte rien au contenu réel des théories scientifiques et sert simplement à systématiser les données de l'observation. On retrouve cette conception originale, qui sépare strictement les composantes empiriques des composantes purement formelles des théories scientifiques, par exemple dans le *Manifeste du cercle de Vienne* :

“Les concepts figurant dans les axiomes sont fixés non pas d'après leur contenu mais uniquement dans leurs relations mutuelles, au moyen de ces mêmes axiomes, ce qui est une façon de les définir. Un tel système d'axiomes n'acquiert une signification pour le réel que lorsqu'on y ajoute d'autres définitions, c'est-à-dire les 'définitions de coordination' (*Zuordnungsdefinitionen*). [...] Grâce à l'application de la méthode axiomatique à ces problèmes, les composants empiriques de la science se séparent en tous points des composants purement conventionnels : un contenu d'énoncé se sépare d'une définition. Un jugement synthétique a priori n'a plus sa place ici”⁶.

Philosophie und Wissenschaftstheorie, Mannheim/Wien/Zürich, 1984, t. 2, p. 267).

6 SOULEZ A. (ed.), *Manifeste du cercle de Vienne et autres écrits*, Paris, 1985, pp. 122–123.

Plusieurs objections ont été soulevées contre cette conception. Deux objections historiquement importantes sont l'argument contre l'existence d'énoncés observationnels neutres — le réel est toujours interrogé à la lumière de certaines théories préalables⁷ — et l'impossibilité de faire correspondre à tous les termes théoriques des termes observationnels⁸. Cependant, comme le remarque Frederic Suppe dans une discussion critique du modèle standard, ces objections n'entraînent pas nécessairement le rejet de la conception du modèle standard. D'abord, les théories auxiliaires qui interviennent dans l'interprétation et la mise en oeuvre d'une expérimentation peuvent être inscrites dans les règles de correspondance. De cette façon on peut sauvegarder la séparation entre le langage théorique et le langage observationnel, malgré la charge théorique de l'observation⁹. Ensuite, la deuxième objection, la présence d'éléments non-empiriques dans la théorie, a été neutralisée par des améliorations ultérieures du modèle standard. Dans ces versions ultérieures les règles de correspondance ne donnent pas une interprétation empirique exhaustive du langage théorique, mais seulement une interprétation partielle. Dire qu'un terme théorique est interprété par l'intermédiaire d'une règle de correspondance, ne veut pas dire que cette interprétation fixe de façon exhaustive le contenu de ce terme, mais simplement qu'une partie du contenu de ce terme est donné par le contenu empirique spécifié dans la règle de correspondance¹⁰. Cette thèse peut être illustrée à partir de l'exemple du contenu empirique du concept d'électron :

“Si l'on considère un terme théorique comme *électron*, nous constatons qu'une grande partie du contenu associé à ce terme concerne des associations non-empiriques qui ne sont reliées aucunement aux manifestations empiriques des électrons. Pour les électrons, ces associations peuvent inclure différents aspects du modèle des boules de billard, diverses intuitions classiques portant sur des masses ponctuelles macroscopiques, etc. [...] Ceci ne signifie pas que, quand les scientifiques parlent d'électrons, ils s'attachent à des amalgames métaphysiques, mais que, quand les scientifiques parlent d'électrons, seulement une partie du contenu du terme 'électron' est empirique”¹¹.

Deux interprétations alternatives du modèle standard peuvent être données si l'on tient compte de la thèse de l'interprétation partielle. Soit on dit que le contenu des termes théoriques est entièrement empirique, l'électron existe réellement dans le monde extérieur, mais que ce contenu empirique

7 Cf. par exemple FEYERABEND P., “Problems of Empiricism”, in COLODNY R.G. (ed.), *Beyond the Edge of Certainty*, Englewood Cliffs (NJ), 1965, pp. 145–260.

8 Pour un aperçu de différents exemples de termes théoriques qui n'ont pas de correspondance empirique cf. VAN FRAASSEN B., *op. cit.*, pp. 59–64.

9 Cf. SUPPE F., *The Semantic Conception of Theories*, *op. cit.*, p. 64.

10 Cf. *ibid.*, p. 46.

11 *Ibid.*, p. 47 ; notre traduction.

n'est pas encore entièrement connu. De cette façon, on obtient une interprétation réaliste des termes théoriques. Soit on dit que les éléments théoriques résiduels, que l'on ne peut pas soumettre à l'épreuve expérimentale, sont des croyances rationnelles d'une certaine communauté d'interprétation. Ces croyances rationnelles sont des conjectures qui permettent de donner un certain pouvoir de prospection à la théorie, mais qui ne peuvent être vérifiées de façon directe. On obtient ainsi une conception des théories scientifiques telles qu'on peut la trouver dans l'empirisme de Quine par exemple. Dans la conception de Quine des théories scientifiques, l'opposition entre le langage théorique et le langage empirique est remplacée par l'opposition entre un champ théorique central et la périphérie expérimentale. En témoigne par exemple l'introduction au paragraphe 'l'empirisme sans les dogmes' de l'article célèbre "Les deux dogmes de l'empirisme" :

"La totalité de ce qu'il est convenu d'appeler notre savoir ou nos croyances, des faits les plus anecdotiques de l'histoire et de la géographie aux lois les plus profondes de la physique atomique ou même des mathématiques pures et de la logique, est une étoffe tissée par l'homme, et dont le contact avec l'expérience ne se fait qu'aux contours. [...] Si un conflit avec l'expérience intervient à la périphérie, des réajustements s'opèrent à l'intérieur du champ."¹²

Ni la thèse de la charge théorique de l'observation, ni l'existence de termes sans correspondance empirique ne remettent entièrement en question le modèle standard de l'empirisme logique. Pour faire face au problème de la charge théorique de l'observation d'un côté, on peut inclure dans les règles de correspondance les théories auxiliaires qui interviennent dans l'interprétation de l'observation et dans la définition des procédures expérimentales. De l'autre, la présence d'éléments non-empiriques dans la théorie peut être intégrée dans le modèle standard à partir de la thèse de l'interprétation partielle.

Cependant on peut se demander si l'image de la science qui ressort de cette version améliorée du modèle standard reflète encore fidèlement la science telle qu'elle se fait effectivement. Si les théories auxiliaires font partie des règles de correspondance, chaque fois que l'on change une théorie auxiliaire, par exemple la théorie de l'instrument de mesure, on change également le contenu de la théorie. En thermodynamique, il y aurait alors au moins autant de théories qu'il y a de théories de la mesure des grandeurs thermodynamiques. D'autre part, dans le modèle standard, chaque formulation langagière différente de la théorie donne également une théorie

12 Cf. QUINE W.V.O., "Les deux dogmes de l'empirisme", in JACOB P., *De Vienne à Cambridge. L'héritage du positivisme logique*, Paris, 1980, pp. 87-112, p. 108.

différente. Cependant des théories formulées différemment au niveau du langage peuvent être équivalentes au niveau mathématique. Ainsi, en mécanique quantique, on peut utiliser deux modèles pour expliquer les spectres d'émission des atomes, l'un faisant appel au concept d'onde, le modèle de Schrödinger, et l'autre faisant appel au concept de probabilité de transition d'un niveau d'énergie à l'autre, le modèle matriciel de Heisenberg. On peut montrer que ces deux modèles sont *mathématiquement* équivalents¹³. De même, il existe des formulations alternatives de la mécanique newtonienne qui ne font pas appel au concept de force. Dans la formulation hamiltonienne par exemple, les équations de mouvement s'obtiennent en calculant l'extremum de l'hamiltonien, qui exprime la différence entre l'énergie cinétique et l'énergie potentielle du système. De nouveau, on peut montrer que ces deux formulations sont mathématiquement équivalentes¹⁴.

Le modèle standard, dans sa version améliorée, entraîne donc une multiplication des théories scientifiques. Si on modifie la formulation langagière d'une théorie, on obtient une nouvelle théorie et si on modifie une règle de correspondance sans modifier la formulation langagière, on obtient également une nouvelle théorie. On peut alors se demander si le dénominateur commun choisi pour la reconstruction rationnelle des théories scientifiques, des expressions langagières d'une théorie et des règles de correspondance, est bien le meilleur dénominateur pour clarifier les principes méthodologiques de la science. Il y a trop de formulations alternatives possibles d'une même théorie et trop de possibilités d'interprétation différentes du matériel expérimental au sein d'une même théorie pour que l'on puisse synthétiser une théorie dans les termes d'un ensemble de formules syntaxiques et de règles de correspondance. Cette constatation a amené certains auteurs, comme Toulmin, Hanson et Feyerabend, à rejeter entièrement le projet de la clarification des principes méthodologiques. Selon ces auteurs, il n'y a pas de reconstruction formelle générale possible des théories scientifiques et de leur évolution¹⁵.

Mais est-ce qu'une telle conclusion suit nécessairement des difficultés du projet de l'empirisme logique ? Pour rejeter le projet de la clarification méthodologique il faut d'abord se mettre d'accord sur ce que l'on veut clarifier. Si on veut étudier la cohérence et la validité des déductions d'une théorie scientifique, il semble qu'une présentation syntaxique soit

13 Cf. MARGENAU H. et LINDSAY R.B., *Foundations of Physics*, Woodbridge, 1981, p. 458.

14 Cf. *ibid.*, p. 130.

15 Cf. MORMANN T., "Husserl's Philosophy of Science and the Semantical Approach", in *Philosophy of Science*, 58 (1991), pp. 61–83, p. 64.

extrêmement utile. L'axiomatisation de la géométrie par Hilbert, par exemple, a permis de mettre en évidence des présupposés de raisonnement dans les démonstrations de la géométrie d'Euclide dont on pouvait difficilement s'apercevoir sans passer par la formalisation¹⁶. Dans ce cas, la présentation syntaxique permet de vérifier la validité des déductions et d'obtenir une formulation plus complète d'une théorie. En thermodynamique, l'axiomatisation a permis de montrer que la théorie classique, la thermodynamique de l'équilibre, était bien un cas particulier de la thermodynamique généralisée, qui couvre également les processus loin de l'équilibre. Ici, l'axiomatisation permet de vérifier la cohérence d'une théorie plus générale, en vérifiant si elle reproduit bien les résultats de la théorie particulière¹⁷. L'approche syntaxique d'une théorie semble donc bien utile dans la clarification des propriétés logiques¹⁸.

Cependant, si on s'intéresse aux propriétés sémantiques des théories scientifiques, comme au contenu et à la relation de référence des termes théoriques, l'approche syntaxique semble faire défaut. Dans la version améliorée du modèle standard on rencontre deux types de problèmes. Tout d'abord les règles de correspondance spécifient à la fois le contenu empirique des termes théoriques et les relations de vérification expérimentale, en formulant par exemple une théorie des procédés de mesure. Des

16 L'utilité de l'axiomatisation de Hilbert est clairement affirmée par Saunders Mac Lane : "In elementary Euclidean geometry, proofs of various facts about figures are developed from the axioms, with occasional implicit use of intuitively evident properties of the figures concerned. [...] If *all* facts are to be deduced from axioms, then these plausible observations must also be proved from the axioms, rather than from inspection of the figure. Once this austere requirement was recognized, it turned out to be possible to add suitable new axioms to those of Euclid so that all arguments could be free of any intuitive or pictorial content. [...] This austere program was systematically carried out by David Hilbert in a book, *Grundlagen der Geometrie*" (Cf. MAC LANE S., *Mathematics : Form and Function*, New York, 1986, p. 63.).

17 En témoigne par exemple l'introduction de David Jou *et al.* au paragraphe sur la formulation axiomatique de la thermodynamique des phénomènes irréversibles : "Here rational thermodynamics [la formulation axiomatique] is utilized more as a method than as a theory. It is shown that the present version of extended irreversible thermodynamics confirms the results derived previously within the classical scheme, which attests to the coherence of the formalism" (JOU D., CASAS-VAZQUEZ J. et LEBON G., *Extended Irreversible Thermodynamics*, Berlin/Heidelberg, 1993, p. 54 ; les commentaires entre crochets sont de l'auteur).

18 C'est également la position défendue par M. Bunge. Dans un article sur la méthode axiomatique en physique il énumère encore d'autres raisons qui rendent la clarification logique utile pour les sciences, comme l'élimination de démonstrations superflues, la clarification des fondements ou le contrôle précis des dérivations (cf. BUNGE M., "The Axiomatic Method in Physics", in BUNGE M., *Method, Model and Matter*, *op. cit.*, pp. 67-87, pp. 85-87).

changements dans les relations de vérification qui sont indépendants de changements dans les relations de contenu empirique ne peuvent être clarifiés. Ensuite on a vu que le dénominateur commun aux différentes versions d'une théorie n'est pas une *formulation linguistique formelle* de cette théorie, mais des *structures mathématiques* communes. Ici aussi, l'approche syntaxique, qui formalise le contenu à partir de structures linguistiques, ne semble pas très adéquate pour clarifier les propriétés sémantiques.

Au lieu de rejeter entièrement le projet de clarification des principes méthodologiques des sciences, on pourrait donc tenter une autre formalisation des théories scientifiques qui serait plus adéquate pour étudier les propriétés sémantiques. C'est précisément ce dernier projet que l'on peut retrouver dans la conception sémantique des théories scientifiques. Nous verrons ci-dessous que ce qui est au centre de la formalisation des théories dans cette conception, ce sont précisément des structures mathématiques — au lieu des expressions linguistiques — et qu'au lieu d'avoir un seul ensemble de règles de correspondance, on aura deux types de règles, les unes spécifiant les relations de vérification et les autres la définition des contenus. De cette façon, la conception sémantique semble bien pouvoir donner une réponse adéquate à la clarification des propriétés sémantiques, clarification qui reste problématique tant qu'on se laisse inspirer par les différentes tentatives d'amélioration de la conception du modèle standard de l'empirisme logique.

2.2. La prise en compte de la dimension sémantique des modèles

La conception sémantique veut proposer un formalisme plus adéquat pour clarifier les propriétés sémantiques des théories scientifiques. Comme tout programme de recherche, il ne s'agit pas d'une conception homogène, que l'on retrouve de façon identique d'un auteur à l'autre, mais d'une famille d'approches plus ou moins semblables. Dans cette famille on peut rassembler des travaux pionniers comme ceux de Patrick Suppes et d'Evert Beth, des travaux de systématisation comme ceux de Frederic Suppe, Bas Van Fraassen, Joseph Sneed et d'autres, ainsi qu'un ensemble d'applications ponctuelles à des problèmes plus spécifiques¹⁹. Malgré les accents différents de chacun de ces auteurs, on peut néanmoins caractériser la conception sémantique à partir de quelques traits généraux. D'après la conception

19 Cf. SUPPES P., *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*, Dordrecht, 1969 ; BETH E., *Inleiding tot de Wijsbegeerte der Wiskunde*, Antwerpen, 1940 ; SUPPE F., *The Semantic Conception of Theories*, *op. cit.* ; VAN FRAASSEN B., *op. cit.* ; SNEED J., *The Logical Structure of Mathematical Physics*, Dordrecht, 1971. Pour une bibliographie plus étendue et un aperçu historique détaillé, voir SUPPE F., *The Semantic Conception of Theories*, *op. cit.*, pp. 1–37.

sémantique, les théories ne sont pas caractérisées par des structures syntaxiques, que l'on peut interpréter ensuite par des règles de correspondance, mais d'emblée par leur domaine d'application, en donnant une famille de modèles particuliers des théories. Une telle famille est spécifiée par les structures mathématiques qui sont communes à l'ensemble des modèles. De cette façon, comme on l'a représenté à figure 3.2., on obtient une représentation des théories scientifiques en trois étages : les structures mathématiques, le phénomène que l'on veut étudier à partir de ces structures et des modèles particuliers comme médiation entre les structures mathématiques générales et le phénomène concret.

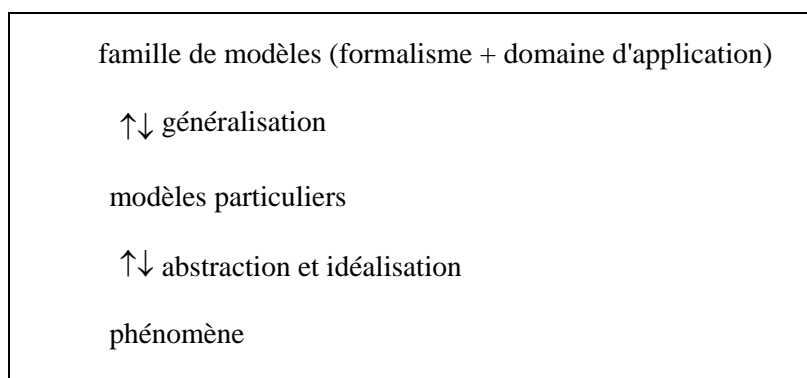


Figure 3.2. La conception sémantique des théories scientifiques (figure de l'auteur).

En mettant au centre de l'activité scientifique le concept de modèle, la conception sémantique opère deux déplacements par rapport au modèle standard²⁰. D'après la conception sémantique :

- a. l'objet direct des théories scientifiques n'est pas les phénomènes ou les observations neutres, mais les modèles du réel ; ces modèles ne sont pas des structures linguistiques (un ensemble d'énoncés) mais des structures conceptuelles non-linguistiques qui décrivent les opérations expérimentales sur le réel et les interactions qui résultent de ces opérations ;
- b. l'outil le plus commode pour caractériser de façon formelle les modèles n'est pas la logique formelle (un système formel d'énoncés)

20 Cf. MORMANN T., *op. cit.*, p. 65.

mais les mathématiques formelles (un système formel d'éléments primitifs, d'opérations et de relations).

Nous pouvons préciser ces deux déplacements opérés par la conception sémantique dans l'élaboration des modèles et dans la caractérisation formelle des théories par la distinction entre une filière théorique évidentielle et une filière théorique référentielle. Cette distinction a été introduite par M. Bunge pour indiquer la différence entre la dimension sémantique des théories, spécifiée dans la filière référentielle, et la dimension de vérification expérimentale, spécifiée dans la filière évidentielle²¹. C'est bien la prise en compte d'une filière référentielle, à côté de la filière évidentielle, qui caractérise la conception sémantique. D'un côté, le déplacement des énoncés d'observation aux modèles vise à tenir compte de l'intervention de concepts théoriques au sein de l'observation. De l'autre, le déplacement des énoncés logiques formels à la famille de structures mathématiques qui caractérisent la théorie vise à tenir compte de la référence au réel des principes théoriques. Dans cette section, nous étudierons de façon plus précise ces deux déplacements, d'abord pour la problématique de l'élaboration des modèles et ensuite pour la problématique de la fécondité des formalismes.

2.2.1. Les conditions d'élaboration des modèles particuliers

Le premier déplacement, du phénomène au modèle, nous permet de mieux étudier l'intervention des concepts théoriques au sein de l'observation des faits empiriques. En effet le scientifique, quand il interroge le réel, n'est pas confronté au réel à l'état brut, dans toute sa complexité. Plutôt, pour pouvoir obtenir des informations utiles, il doit opérer une abstraction sur le réel. C'est cette abstraction qui définira le passage du phénomène au modèle.

En mécanique classique par exemple, on s'intéresse uniquement aux propriétés mécaniques des corps comme la position, la masse et la vitesse. D'autres propriétés, comme la couleur ou la température ne sont pas prises en considération. De plus, on s'intéresse à ces propriétés dans des conditions idéalisées. On fait par exemple abstraction du frottement et on fait comme si toute la masse de l'objet était concentrée dans un point. L'objet de la mécanique classique sera alors l'étude de systèmes isolés de masses ponctuelles, sans extension, qui interagissent dans le vide. Ces systèmes sont

21 Cf. BUNGE M., "The Axiomatic Method in Physics", *op. cit.* ; ID., *Scientific Research*, Berlin/Heidelberg/New York, 1967, p. 496. Pour un exposé plus détaillé des distinctions introduites par Bunge, cf. FELTZ B., *Croisées biologiques*, Louvain-la-Neuve, 1991, pp. 259–274.

caractérisés par la valeur des variables de position et de quantité de mouvement des masses ponctuelles à un moment donné. La théorie de la mécanique classique vise donc à décrire le comportement de systèmes abstraits. Le but de cette abstraction est, d'un côté, d'isoler les traits qui nous semblent déterminants dans la production du comportement (la vitesse mais pas la température par exemple). De l'autre, le but de l'idéalisation est l'intervention des mathématiques. Par l'idéalisation, on fait abstraction des contenus réels du système (les corps deviennent des points géométriques par exemple) et on étudie le système à partir du seul aspect des relations formelles entre les éléments, ou des opérations formelles que ces éléments effectuent.

L'abstraction et l'idéalisation ne sont cependant pas uniquement des opérations sémantiques, simples interprétations des phénomènes à partir des schèmes conceptuels et constructions mathématiques du scientifique. L'étude du comportement des systèmes abstraits suppose une intervention active dans le réel. D'un côté, il faudra isoler l'objet étudié pour que les traits dont on a fait abstraction ne puissent pas influencer le comportement (éviter que le système s'échauffe par exemple). De l'autre, l'intervention des mathématiques suppose une manipulation délibérée du système. La réduction des corps massifs à des corps ponctuels par exemple, vise à étudier la masse sous son seul aspect de quantité ou de poids, et à associer un nombre précis à cette quantité. Cependant les nombres sont des entités idéales, mathématiques, et ne sont pas donnés en tant que tels dans le réel. Comment alors utiliser cette entité idéale dans la description des phénomènes ? C'est ici qu'interviennent les procédures expérimentales. L'intervention expérimentale vise à reproduire, dans la pratique, les opérations abstraites de définition des entités idéales. Les nombres naturels par exemple peuvent se définir, en utilisant les axiomes de Peano, à partir d'une unité de base (le nombre zéro) et l'opération de succession (plus un). Alors en partant d'un nombre défini de façon conventionnelle (l'unité de mesure) on peut obtenir, dans la pratique, d'autres nombres en démultipliant cette unité de base. De cette façon on peut associer, par exemple, au poids un nombre défini de façon exacte. Evidemment tout ce que l'on peut faire dans la pratique expérimentale, c'est construire des procédures analogues aux opérations mathématiques exactes. Si les opérations expérimentales ne reproduisent pas certaines propriétés mathématiques, l'erreur ne sera jamais du côté des mathématiques, mais du côté de l'imperfection de nos manipulations expérimentales. L'intervention expérimentale devra donc s'accompagner d'une estimation des erreurs commises et des facteurs de correction à introduire.

L'abstraction et l'idéalisation supposent donc une intervention délibérée dans le système. D'un côté, pour rendre le système conforme aux conditions d'abstraction, il faut l'isoler convenablement de son environnement et de l'autre, pour pouvoir construire un modèle mathématique, il faut manipuler le système en suivant des procédures systématiques d'interaction expérimentale. Précisons que l'interaction expérimentale ne se limite aucunement aux seules procédures de mesure, c'est-à-dire à l'étude quantitative du système. Le caractère construit des mathématiques n'est pas limité à l'arithmétique, où l'on trouve les opérations de construction des grandeurs, mais c'est une propriété universelle des mathématiques²². Si l'on veut étudier les propriétés de symétrie spatiale de la disposition d'un ensemble de points matériels par exemple, les opérations élémentaires qui interviendront ne seront pas des opérations de construction de grandeurs, mais seront des opérations géométriques de projection, de rotation et de translation de figures. Ou pour étudier un système de relations, on peut définir de façon algébrique des relations de transitivité, de réflexivité ou de permutabilité entre les éléments. Ensuite, en réeffectuant de telles opérations dans la pratique, on peut construire différents types de modèles mathématiques. Ainsi les méthodes structuralistes en sciences humaines construisent des modèles non quantitatifs en utilisant les définitions de l'algèbre pour définir la structure relationnelle d'un système²³.

En étudiant de façon précise le rapport entre les modèles scientifiques et les phénomènes étudiés, la conception sémantique des théories scientifiques permet de définir de façon plus précise les relations de référence et les relations de vérification des concepts théoriques mis en oeuvre. La référence sera définie à partir de l'abstraction et de l'idéalisation mise en oeuvre, et la vérification à partir de l'interaction expérimentale et la maîtrise des conditions d'isolement du système. Le rapport du système au phénomène est donc à la fois un rapport référentiel et évidentiel, une sollicitation à partir d'un certain regard et un enregistrement aussi précis que possible des réactions renvoyées.

2.2.2. Les conditions de la productivité sémantique des formalismes

L'abstraction et l'idéalisation permettent de définir un premier niveau de généralité auquel on peut étudier les phénomènes. C'est le niveau de l'élaboration des modèles théoriques particuliers et de l'étude expérimentale

22 Cf. MAC LANE S., *op. cit.*, p. 410 ; LADRIÈRE J., *L'articulation du sens*, deux tomes, Paris, 1984, t. 1, p. 166.

23 Cf. LADRIÈRE J., *L'articulation du sens*, 2 tomes, *op. cit.*, p. 178.

des propriétés de fonctionnement. Ce niveau combine la projection d'un regard dans le réel (l'élaboration du modèle) et l'interaction avec le réel (l'intervention expérimentale). Cependant, il y a un deuxième niveau de généralité qui intervient dans la construction des théories scientifiques, qui se situe au-dessus des modèles particuliers : c'est le niveau des présuppositions théoriques plus générales des domaines étudiés. En effet, quand on élabore un modèle particulier, par exemple un modèle qui veut décrire les positions dans l'espace d'un électron — en reprenant l'exemple cité dans notre discussion du modèle standard —, on partira de certaines présuppositions théoriques générales, comme des intuitions classiques portant sur des masses ponctuelles macroscopiques, l'idée d'interactions par chocs entre particules, etc. Ensuite on se laissera guider par ces présuppositions dans notre construction du modèle particulier. Ces présuppositions sont exprimées dans les modèles théoriques généraux sous la forme de principes généraux de la discipline (ainsi on parle des principes de Newton, des principes de la cybernétique, les principes de la thermodynamique, etc.). Ce que les principes ont de remarquable, c'est qu'ils permettent non seulement de synthétiser, à un niveau d'abstraction plus élevé que les modèles particuliers, l'ensemble des faits connus, mais d'anticiper également le comportement de faits qui n'appartiennent pas à la classe de faits déjà étudiée. Le regroupement de différents modèles particuliers à partir de principes généraux donne une véritable fécondité à la théorie, fécondité qui permet de découvrir de nouveaux horizons de la réalité.

Pour clarifier le statut des modèles utilisés en sciences, il faut clarifier également la fécondité des principes théoriques utilisés. S'agit-il simplement de conjectures, basées sur des croyances d'une certaine communauté d'interprétation, comme dans la conception des théories scientifiques de Quine ? Ou les concepts théoriques et les principes ont-ils également un contenu réel et caractérisent-ils des structures d'intelligibilité intrinsèques du réel ? Pour répondre à cette question il nous faut regarder de plus près le second déplacement opéré par la conception sémantique des théories scientifiques. Selon la conception sémantique, l'outil le plus commode pour caractériser les modèles théoriques n'est pas la logique formelle, comme dans le modèle standard, mais les structures mathématiques. Nous verrons que l'enjeu de ce déplacement est une prise en compte du rapport au réel des principes généraux des théories scientifiques.

Tout d'abord, en toute généralité, pour rendre effectif son pouvoir d'anticipation, la théorie recourt également au formalisme des systèmes déductifs. En déduisant toutes les conséquences de quelques principes de base, on peut élargir le domaine d'interprétation originale qui nous a inspiré

dans la construction de la théorie et ainsi projeter nos présuppositions dans de nouvelles classes de systèmes²⁴. Comme le pouvoir de déduction de la théorie est un corollaire de son organisation formelle (organisation à partir d'éléments de base et de règles de déduction), la fécondité de la théorie semble résider dans son caractère formel. Cette première perspective sur la fécondité des théories scientifiques est la perspective que l'on a rencontrée dans le modèle standard de l'empirisme logique. Si la fécondité tient au pouvoir de déduction, on peut clarifier les édifices théoriques en les traduisant dans le langage de la logique formelle, c'est-à-dire en donnant une description purement syntaxique de la théorie. L'interprétation des symboles formels du langage théorique entièrement syntaxique sera fixée par des règles de correspondance qui permettent d'associer un contenu empirique aux termes du langage théorique. Les significations théoriques résiduelles, que l'on ne peut pas entièrement traduire sous la forme de règles de correspondance, sont alors soit dénuées de sens, comme dans la première version de l'empirisme logique, soit des conjectures à vérifier, comme dans la version améliorée à partir de la thèse de l'interprétation partielle.

D'après le modèle standard, la fécondité de la théorie tient donc à son caractère formel. La formulation syntaxique de la théorie permet d'obtenir la classe des systèmes logiquement possibles, respectant le principe de non-contradiction, et pour retrouver, à partir de là, les systèmes physiquement possibles il suffit d'appliquer les règles d'interprétation (règles de correspondance) aux propositions valides du système formel (c'est-à-dire aux propositions qui découlent bien des principes). Cependant, le rapport entre les systèmes logiquement possibles, les énoncés valides, et les systèmes physiquement possibles, les énoncés vrais, n'est pas aussi simple que l'on pourrait le croire. Les énoncés valides ou théorèmes d'un système formel sont les énoncés qui peuvent être obtenus par déduction à partir des axiomes, ce sont les énoncés prouvables. Toutefois, suite aux travaux de Kurt Gödel, on sait qu'on ne peut pas définir la notion de vérité, et donc ici la possibilité physique d'une proposition, à partir de la notion de prouvabilité. En effet, Gödel a montré qu'un énoncé (intuitivement) vrai n'est pas nécessairement prouvable. De même, dans les systèmes formels, on peut avoir des énoncés faux sans pouvoir prouver la vérité de la négation de ces énoncés, qui devraient être des énoncés vrais²⁵. La notion de vérité, et donc l'étude de la possibilité physique des systèmes, requiert plus que la simple prouvabilité ou

24 Cf. LADRIÈRE J., "Le formalisme et le sens", in MOULOU D. N. (coord.), *Les langages, le sens, l'histoire*, Villeneuve d'Ascq., 1977, pp. 241–277, p. 274 ; cité dans MAESSCHALCK M., *Normes et Contextes*, op. cit., p. 192.

25 Cf. BETH E., op. cit., p. 189.

la déduction à partir des principes. En effet, comme l'a montré Alfred Tarski, pour définir la vérité d'un énoncé il faut préciser en plus le domaine d'interprétation des prédicats et des variables de la théorie purement formelle²⁶. Dès lors, la vérité ne peut pas être définie à un niveau d'abstraction entièrement formel, sans référence à un domaine d'application. Pour traiter de la vérité des énoncés, et donc de la possibilité physique des modèles théoriques, il faut introduire du concret.

La théorie qui étudie de façon systématique les relations qui existent entre un système formel et les domaines qu'il veut représenter est la théorie des modèles sémantiques. Un modèle sémantique²⁷ d'une théorie formelle est une interprétation des énoncés valides du système formel sur un domaine d'objets D qui rend ces énoncés vrais²⁸. Comme on peut définir différents domaines d'application, une même théorie formelle pourra admettre différents modèles sémantiques. Si deux modèles sémantiques sont isomorphes (c'est-à-dire lorsqu'il y a entre eux une correspondance biunivoque qui conserve les relations), on peut les considérer comme deux présentations différentes d'une même réalité²⁹. Pour rendre compte de la fécondité des théories scientifiques, il faut également renvoyer à des modèles sémantiques qui représentent les domaines possibles de validité. Il ne suffit donc pas de se référer seulement à leur pouvoir déductif, c'est-à-dire à leur capacité d'exprimer des possibilités logiques. En se situant à un niveau d'abstraction trop élevé, l'empirisme logique a seulement conservé de la fécondité des théories scientifiques l'aspect logique, mais a complètement éliminé leur référence à un domaine d'application. C'est effectivement ce que l'on retrouve dans la thèse du statut purement conventionnel ou conjectural des principes généraux et des concepts théoriques.

Les modèles sémantiques, qui sont des interprétations d'un système formel sur un certain domaine d'objets D, correspondent à un niveau d'abstraction intermédiaire. D'un côté ils renvoient à un domaine d'application, et par là permettent de caractériser une classe de systèmes physiquement possibles, mais de l'autre ils conservent, par leur aspect formel, une certaine généralité et donc un certain pouvoir d'anticipation. On pourrait dire, en empruntant à l'analyse de la productivité sémantique des formalismes de Jean Ladrière, que les modèles sémantiques sont des paliers

26 Cf. *ibid.*, pp. 189–190.

27 Pour désigner l'interprétation d'un système formel on utilise parfois le terme *modèle* d'un système formel, mais, pour distinguer ce dernier concept du concept de modèle que nous avons utilisé jusqu'ici, nous parlerons de *modèle sémantique* d'un système formel.

28 Cf. LADRIÈRE J., "Les limites de la formalisation", in PIAGET J. (dir.), *Logique et connaissance scientifique (Encyclopédie La Pléiade)*, Paris, 1967, p. 319.

29 Cf. *ibid.*, p. 320.

de stabilisation “dans un processus de généralisation reliant par paliers l'effectivité à la généralité de telle sorte que chaque palier soit à la fois (dans l'ordre inverse du schème kantien) la forme d'un plus concret et le support d'un plus abstrait”³⁰. Il semble que l'on peut obtenir une caractérisation adéquate du contenu des principes théoriques à partir de telles structures d'un niveau d'abstraction intermédiaire. Un bon exemple en est donné par la représentation des théories quantitatives dans un espace des phases³¹. L'espace des phases est un espace n -dimensionnel où les n coordonnées de l'espace représentent les n paramètres de la théorie. Par exemple un système classique de 2 particules sera représenté dans l'espace des phases par les 6 coordonnées de position et les 6 coordonnées de quantité de mouvement des deux particules ($n = 12$). Un système classique de N particules sera représenté par les $3N$ coordonnées de position et les $3N$ coordonnées de quantité de mouvement ($n = 6N$). Chaque point dans l'espace des phases représente alors un vecteur dont les composantes sont les positions et les quantités de mouvement des particules et correspond donc à un état *logiquement possible* du système. Maintenant, pour pouvoir appliquer ce formalisme au réel, nous avons besoin d'une caractérisation des états *physiquement possibles*. Pour cela, on définit les configurations et les changements de configurations physiquement admissibles parmi les configurations possibles. Cette définition est donnée par les principes généraux de la théorie. Ces principes peuvent être déterministes, statistiques ou autres. En mécanique classique les principes sont donnés par exemple sous la forme des équations de Hamilton. Pour chaque forme spécifique de l'Hamiltonien on obtient alors un modèle particulier de la théorie.

La représentation des théories quantifiables dans un espace des phases³² illustre bien la représentation des théories par la structure mathématique commune à une famille de modèles particuliers. D'un côté, l'espace des phases est une structure purement formelle, un espace géométrique tout à fait abstrait. Et de l'autre, les équations d'évolution du système introduisent des contraintes sur cet espace pour préciser les états physiquement possibles. Ensuite, on peut obtenir les modèles particuliers en spécifiant la valeur des

30 MAESSCHALCK M., *Normes et Contextes*, *op. cit.*, p. 198.

31 Cf. SUPPE F., *The Structure of Scientific Theories*, Chicago, 1977, pp. 226–229.

32 La représentation de la famille de modèles en utilisant le formalisme de l'espace des phases correspond à la version de la conception sémantique privilégiée par Van Fraassen. Suppe tente de généraliser cette représentation au cas des théories non quantifiées. Pour cela, en se basant sur un article de Tarski, il représente une théorie en toute généralité comme un système relationnel, défini à partir du formalisme de l'algèbre des relations. De cette façon, on peut couvrir aussi bien les théories quantitatives que les théories qualitatives (cf. SUPPE F., *The Semantic Conception of Theories*, *op. cit.*, p. 84).

variables sur un sous-ensemble D' du domaine de validité D . Ainsi, si les principes généraux décrivent la dynamique d'une masse ponctuelle en oscillation autour de l'origine des axes de référence, des valeurs spécifiques des paramètres permettent d'appliquer la théorie au mouvement d'un pendule en absence de frottement, d'un ressort parfaitement élastique ou d'une corde vibrante par exemple. Les principes généraux de la théorie décrivent donc les contraintes physiques communes à toute une famille de modèles particuliers. Ces modèles particuliers sont obtenus en appliquant les principes à un sous-ensemble d'objets du domaine de validité de la théorie.

Finalement, à partir de la représentation des théories scientifiques par la structure mathématique commune aux modèles particuliers de la théorie, on peut reprendre la question de la limitation de la fécondité des principes théoriques. Dans le modèle standard, les principes théoriques sont simplement des systématisations commodes du matériel empirique, qui peuvent être clarifiées en les représentant sous une forme purement syntaxique. Cependant, à partir d'une représentation syntaxique, on ne peut pas rendre compte adéquatement de la fécondité des principes théoriques. Cette fécondité tient autant à leur caractère formel qu'au fait que ces principes gardent une certaine trace du réel. En effet, les principes généraux sont valables uniquement sur un domaine d'application particulier. Ce domaine caractérise l'ensemble des modèles spécifiques dont les principes décrivent les propriétés communes. C'est donc le croisement de la représentation formelle avec la relation à un domaine d'application possible qui garantit à la fois une fécondité et un rapport au réel des conceptions théoriques. La première dimension, celle de la formalisation, confère à la théorie un certain pouvoir d'anticipation, à partir de son auto-compréhension formelle. Cette anticipation détermine l'intention sémantique de la théorie, que l'on peut appeler sa limitation interne. La deuxième, la relation à un domaine d'application, garantit un certain niveau de vérification empirique. Elle détermine l'extension de la théorie ou sa limitation contextuelle. Ces deux limitations ne sont pas amenées de façon explicite à l'avant-plan de la modélisation. Comme nous le verrons dans la section suivante, selon le degré d'abstraction des principes théoriques, tantôt l'accent sera mis sur la vérification empirique stricte des principes, et tantôt sur leur pouvoir d'exploration du domaine des possibilités physiques en général.

3. Du formalisme aux modèles particuliers

La conception sémantique des théories scientifiques nous donne un cadre théorique général pour réfléchir sur le rôle des modèles en sciences.

Tout d'abord, le recours à des formalismes en sciences n'élimine pas l'importance des modèles. En effet, contrairement à ce qu'on pourrait croire à partir d'une présentation syntaxique des théories, la validité des énoncés formels est relative à un contexte de modélisation particulier ou encore à une famille de modèles définie sur le domaine d'interprétation de la théorie. Ensuite, cette importance du rôle des modèles en sciences introduit des limitations supplémentaires par rapport à l'exigence de la cohérence logique et de l'adéquation empirique. D'un côté, la fécondité des formalismes est limitée par la nécessité de se rapporter à un domaine d'application particulier ou famille de modèles. Tout formalisme anticipe dans son auto-compréhension un tel domaine d'application, ce qui constitue sa limitation interne. De l'autre, les modèles particuliers de la théorie ne sont pas uniquement soumis à la condition de la vérification expérimentale, mais doivent également satisfaire les conditions d'abstraction et d'idéalisation mises en oeuvre dans leur élaboration. Le formalisme doit donc prendre en compte également la limitation contextuelle spécifique liée à cette élaboration des modèles particuliers.

Le point de départ de la réflexion sur le rôle des modèles en sciences dans le cadre de la conception sémantique est très clairement la physique mathématique. Ce choix est guidé par le fait que la fécondité de la modélisation en physique mathématique est bien établie et peut servir d'étalon à partir duquel on peut réfléchir sur la modélisation dans d'autres disciplines. Des auteurs comme Frederic Suppe et Mario Bunge accordent une très grande importance au type de modélisation qui est à l'oeuvre en physique, mais tentent également d'étendre leur réflexion à d'autres disciplines comme la biologie ou les sciences humaines. Dans ce deuxième paragraphe, nous voudrions montrer que la conception sémantique permet aussi de clarifier la fécondité et les limitations des modèles en sciences cognitives, tout en indiquant une différence importante par rapport à la modélisation en physique.

Tout d'abord, on semble bien retrouver en sciences cognitives l'importance de la fécondité sémantique des formalismes. Par exemple, comme nous l'avons vu dans le premier chapitre de ce travail, l'étude des opérations computationnelles en sciences cognitives était au départ liée à une interprétation particulière de la cognition dans le paradigme symbolique, mais peut être généralisée à l'étude des formes d'intelligence habituelle et exploratoire. Dans une première section, nous poserons la question de la limitation interne de cette fécondité. Cette question renvoie en un deuxième temps du formalisme au domaine d'application du formalisme, et par là à la question de la spécification des modèles particuliers de la théorie. Dans la

deuxième section de ce paragraphe, nous montrerons alors que la question de la spécification du domaine d'application des formalismes en sciences cognitives introduit une limitation contextuelle spécifique liée à la modélisation de l'intentionnalité.

3.1. Les limitations internes de la productivité sémantique du formalisme en sciences cognitives

La présentation de la conception sémantique dans le premier paragraphe pourrait donner l'impression que la relation du formalisme au domaine d'application n'est pas une relation problématique. En effet, nous avons essentiellement discuté des exemples de théories scientifiques bien établies, dont la vérification empirique ne pose plus de problèmes. Toutefois, en sciences cognitives, les modèles théoriques ne répondent pas nécessairement à des critères stricts de vérification. Il s'agit d'un champ en pleine évolution dans lequel interviennent également des théories avec un faible degré de vérification. Pour rendre compte des limitations internes des formalismes en sciences cognitives, il faudra donc compléter l'analyse des propriétés sémantiques par une analyse des degrés de généralité des modèles.

Mario Bunge est un auteur qui a analysé l'utilisation en sciences de modèles théoriques avec des degrés de généralité variable. Dans un article sur les critères de vérification utilisés en sciences, Bunge distingue trois niveaux de généralité des modèles théoriques, le niveau des modèles théoriques particuliers, le niveau des modèles théoriques généraux et le niveau des modèles théoriques abstraits, et trois niveaux de vérifiabilité correspondants³³. Tout d'abord, Bunge précise que seuls les modèles théoriques particuliers, qui décrivent le fonctionnement de systèmes particuliers, répondent à des critères de vérification empirique stricts, c'est-à-dire présentent la possibilité de tester les prédictions de la théorie (la confirmation) et de réfuter ces prédictions par l'expérience (la falsification). Des théories plus générales, comme la mécanique classique ou la théorie de

33 Cf. BUNGE M., "Testability Today", in *Method, Model and Matter*, *op. cit.*, pp. 27–43. Dans cet article, M. Bunge désigne les trois niveaux respectivement par les termes de théorie spécifique, théorie générique interprétée et théorie générique sémi-interprétée (cf. *ibid.*, p. 39). Remarquons que M. Bunge distingue encore un quatrième niveau de généralité, qui est celui des affirmations métaphysiques. Ces dernières affirmations ne peuvent plus être insérées dans la dynamique de la vérification. Par exemple la théorie de la mécanique classique est compatible aussi bien avec une théorie qui prône l'existence d'un espace absolu, comme le faisait Newton, ou une théorie purement relationnelle de l'espace, comme celle de Leibniz. Ces affirmations invérifiables peuvent cependant fonctionner comme des outils de clarification et de systématisation commodes des théories scientifiques (cf. *ibid.*, p. 41).

l'évolution ne peuvent plus être falsifiées, ni être confirmées directement. Ces théories spécifient des contraintes générales de toute une classe de modèles particuliers. Pour les tester, il faut les rendre plus spécifiques, en donnant des valeurs particulières aux variables et en ajoutant des hypothèses auxiliaires quant aux paramètres qui définissent les propriétés dynamiques du système étudié. Ces théories sont scientifiques dans la mesure où elles peuvent être soumises à des procédures de vérification de façon indirecte. Les modèles théoriques abstraits, finalement, spécifient uniquement des processus de transformation ou d'évolution généraux, sans préciser le système auquel ces processus s'appliquent. Des exemples de telles théories sont donnés par la théorie de l'information, la cybernétique ou la théorie des automates. Ces théories peuvent seulement être vérifiées quand elles sont enrichies d'hypothèses complémentaires qui permettent de les utiliser dans des théories générales.

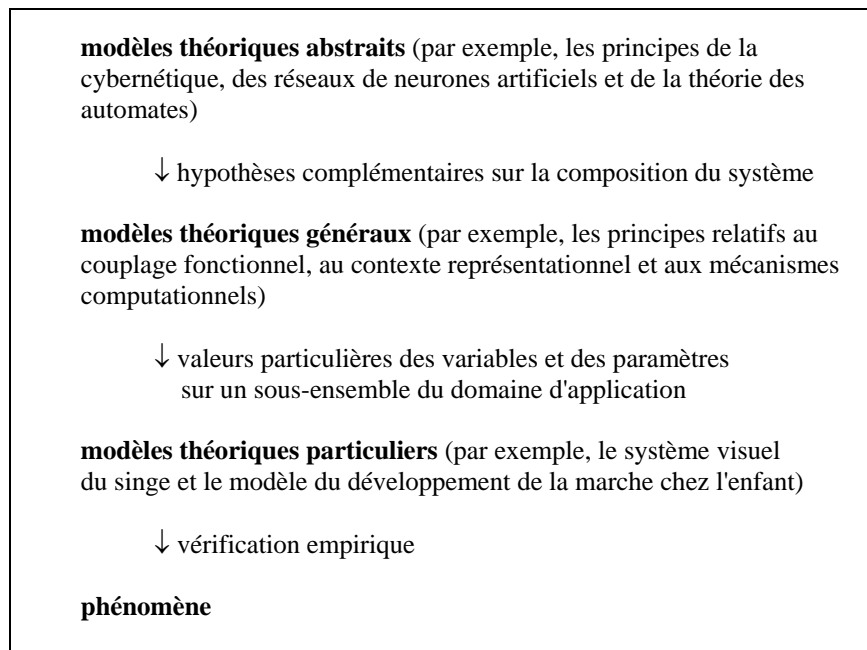


Figure 3.3. Filière évidentielle en sciences cognitives : degrés de généralité des modèles et procédures de vérification (figure de l'auteur).

En sciences cognitives, on rencontre également l'utilisation de tels modèles théoriques avec des degrés de vérification variable. Des exemples de modèles théoriques particuliers que nous avons étudiés sont le modèle du

système visuel du singe, le modèle du développement de la marche chez l'enfant ou encore le modèle de Hutchins des interactions entre les membres de l'équipage d'un voilier. Des exemples de modèles théoriques généraux sont la théorie cognitiviste classique, le connexionnisme ou la théorie écologique de la cognition. Ces modèles généraux donnent un contenu spécifique aux concepts de computation, de couplage fonctionnel et de représentation, qui caractérisent la modélisation en sciences cognitives. Finalement, les modèles théoriques abstraits sont également omniprésents en sciences cognitives, que l'on pense aux modèles de la simulation de réseaux de neurones artificiels ou à l'importance des principes de la cybernétique. Nous avons représenté de façon schématique ces différents niveaux à la figure 3.3.

Pour montrer la pertinence de l'analyse de Bunge pour le problème des limitations des formalisme en sciences cognitives, nous développerons plus dans le détail deux exemples caractéristiques de modélisation en sciences cognitives, l'étude des réseaux de neurones et l'étude de comportement de robots. Tout d'abord, nous monterons que ces modèles participent bien à la dynamique de la recherche en sciences cognitives. Ensuite, nous illustrerons la question de la vérification de ces modèles par l'exemple des réseaux de neurones.

Un premier exemple de modélisation que nous avons rencontré à de nombreux endroits dans ce travail est donné par les simulations du fonctionnement de réseaux de neurones artificiels. Malgré ce que laisse entendre le vocabulaire utilisé, il ne faut pas surestimer la similarité des réseaux de neurones artificiels avec les neurones réels dans le cerveau³⁴. Les réseaux de neurones artificiels ne tiennent pas compte de la multiplicité des types de neurones, des mécanismes de transmission synaptiques ou de structures de connexions entre neurones qui existent dans le cerveau. De plus, la règle de Hebb de modification des poids des connexions entre les neurones artificiels n'est observée dans le cerveau que pour certains processus simples. Les réseaux de neurones ne donnent donc aucune précision fiable sur les mécanismes d'implémentation qui réalisent les processus simulés. Alors, typiquement, en jouant sur la valeur des paramètres qui définissent le réseau de neurones, on pourra toujours obtenir une bonne simulation du comportement souhaité. C'est-à-dire que l'on peut forcer la confirmation en ajustant la valeur des paramètres. De cette façon,

34 Cf. CHURCHLAND P.M., *A Neurocomputational Perspective*, *op. cit.*, p. 184 ; CLARK A., *Being There*, *op. cit.*, p. 54 ; EDELMAN G.M., *Neural Darwinism*, New York, 1987, pp. 181-183.

les simulations de processus cognitifs à partir des réseaux de neurones artificiels peuvent échapper à toute possibilité de réfutation.

Néanmoins, l'étude des réseaux de neurones a permis d'élargir notre compréhension des mécanismes cognitifs et de fournir de nouveaux concepts pour interpréter les données expérimentales des neurosciences cognitives et de la psychologie cognitive. En particulier, l'utilisation massive de représentations distribuées et de modules interactifs dans les simulations a permis de montrer que l'implémentation de processus cognitifs parfois complexes ne nécessite pas nécessairement une organisation modulaire et hiérarchique. Les méthodes de simulation se prêtent donc particulièrement bien à l'exploration de modèles qui pourraient réaliser telle ou telle propriété. Parfois les résultats de la simulation suggèrent des solutions inhabituelles à partir desquelles on pourra jeter sur le réel un regard nouveau .

Un autre type de modèles avec un faible degré de vérification, rencontré déjà auparavant, sont les modèles d'exploration de comportements de robots. Ces modèles sont apparentés aux modèles de simulation, en ce sens qu'ils étudient également des objets simplifiés du réel pour explorer la modélisation de certaines propriétés conceptuelles. Cependant, au lieu d'étudier la possibilité de certains schèmes conceptuels *a priori*, à partir de calculs sur ordinateur, on construit un double réel du système en question. Cette fois, ce sera la façon concrète dont on doit s'y prendre pour mettre ensemble un certain artefact qui pourra suggérer des idées parfois tout à fait inattendues pour les structures ou les principes de fonctionnement des systèmes cognitifs. Ainsi, dans le cas des robots de Brooks, on a vu que deux boucles sensori-motrices indépendantes dans un robot permettent d'obtenir un comportement pour lequel de manière classique on supposait un système central de représentation et de contrôle de mouvement. De même, dans le cas de la construction d'un bras de robot automatique, on a vu qu'il est parfois plus utile d'introduire une pièce élastique, comme un joint en caoutchouc, que de construire un système plus précis dont on ajuste le mouvement par des calculs exacts de trajectoire. Ici la robotique montre qu'une solution "mentaliste", de type calculatoire, est parfois moins indiquée qu'une solution qui s'appuie sur les ressources dynamiques de l'environnement.

Ces exemples tirés de la robotique montrent comment la construction d'artefacts peut jouer un rôle important dans la suggestion de nouvelles idées qui interviennent dans le progrès théorique. Historiquement, c'est d'ailleurs dans la conception de tels systèmes artificiels que se trouve l'origine de la formulation des principes de la cybernétique et des principes d'auto-organisation des systèmes dynamiques³⁵. D'un côté, la cybernétique de

35 Cf. HEIMS J., *op. cit.*

Wiener s'est développée dans le cadre de problèmes concrets, comme les problèmes de transport de messages dans des réseaux de communication ou la conception de systèmes de vol automatique. De l'autre, les principes de l'auto-organisation, comme le principe d'ordre par le bruit de von Foerster, trouvent leur origine dans l'étude de systèmes artificiels non-linéaires, comme les systèmes laser, capables d'une certaine forme d'auto-organisation³⁶. À partir de ces exemples, on constate donc une collaboration étroite, dans la recherche en sciences cognitives, entre la démarche technologique et la démarche scientifique proprement dite. D'un côté les principes abstraits des théories, comme ceux de la cybernétique ou ceux de la théorie des systèmes auto-organisés, permettent de suggérer de nouveaux artefacts physiquement possibles. Et de l'autre, le bricolage concret de nouveaux systèmes artificiels permet de suggérer des principes de fonctionnement dont on peut se servir ensuite dans la construction de nouvelles théories.

Les exemples de la simulation et de l'étude du comportement de systèmes artificiels montrent bien l'importance de l'utilisation de modèles avec un faible degré d'adéquation empirique dans la construction des théories en sciences cognitives. Ces modèles étudient des principes généraux, sans donner des informations plus spécifiques sur les systèmes auxquels ils peuvent s'appliquer. Ils se présentent donc d'abord comme des outils d'exploration conceptuelle. Toutefois, ces modèles ne peuvent être réellement féconds pour la recherche en sciences cognitives que si on peut spécifier leur relation à un domaine de modèles particuliers. Il faut donc pouvoir croiser, comme dans l'analyse de Bunge, la dynamique de l'exploration conceptuelle et la dynamique de la vérification.

Finalement, nous pouvons illustrer cette limitation de la fécondité des formalismes en appliquant l'analyse des procédures de vérification de Bunge à l'exemple de l'étude des réseaux de neurones artificiels. Tout d'abord, nous avons vu que les réseaux de neurones artificiels permettent d'étudier des modèles d'encodage de représentations. En particulier, les simulations permettent de montrer la possibilité de l'encodage distribué et interactif dans les systèmes cognitifs. Ensuite, pour pouvoir utiliser ces modèles dans l'étude des systèmes cognitifs naturels, il faut prendre en considération les principes généraux qui interviennent dans l'étude de la cognition. Un principe que nous avons étudié dans ce travail, c'est la nécessité d'articuler la

36 Dans les systèmes laser on génère un faisceau lumineux cohérent à partir d'oscillations locales. Il s'agit donc d'une génération d'ordre à partir de mouvements désordonnés. C'est dans le cadre de travaux sur de tels systèmes que von Foerster a formulé la première fois son célèbre principe d'ordre par le bruit.

question des mécanismes computationnels à la question de l'organisation fonctionnelle des systèmes cognitifs. Le problème de l'encodage de représentations, que l'on peut étudier dans les réseaux artificiels sur le plan conceptuel, prendra alors la forme de l'étude de différentes hypothèses d'organisation fonctionnelle du système cognitif, comme l'hypothèse de l'organisation modulaire ou l'hypothèse de l'organisation en modules interactifs. Finalement, l'étude de l'organisation fonctionnelle doit s'appuyer sur la vérification empirique de différents modèles théoriques particuliers. Ainsi, nous avons étudié l'hypothèse de l'organisation en modules interactifs, à partir de modèles particuliers du système visuel et de la coordination motrice. C'est uniquement en spécifiant les principes abstraits de cette façon sur une famille de modèles théoriques particuliers, que l'on peut garantir leur fécondité pour la recherche en sciences cognitives.

3.2. Modèles et interprétation dans la modélisation de l'intentionnalité

La problématique des limitations de la fécondité sémantique des formalismes en sciences cognitives, que nous avons étudiée à partir de l'analyse des procédures de vérification de Bunge, renvoie à la problématique de la spécification du domaine des modèles particuliers auquel le formalisme peut s'appliquer. Toutefois, le renvoi du formalisme au domaine d'application n'est pas une limitation spécifique liée aux formalismes utilisés en sciences cognitives. Comme nous l'avons vu dans le premier paragraphe, il s'agit d'une limitation également présente dans la modélisation en physique. Dans cette section, nous voudrions montrer que la problématique de la modélisation de l'intentionnalité en sciences cognitives introduit une limitation spécifique par rapport à la modélisation en physique.

En toute généralité, les modèles particuliers sont construits à partir des principes généraux en spécifiant les hypothèses auxiliaires qui permettent de donner une valeur particulière aux variables et aux paramètres du modèle³⁷. En physique, ces hypothèses auxiliaires portent sur la définition des conditions frontières et des conditions initiales du système³⁸. Par exemple, à partir des équations de mouvement de la mécanique classique, on peut construire différents modèles particuliers en résolvant les équations par rapport à un ensemble de conditions frontières et un état initial. A la différence de la physique, la définition des modèles particuliers de l'intentionnalité en sciences cognitives ne dépend pas uniquement de conditions qui portent sur la frontière entre le système et l'environnement,

37 Cf. BUNGE M., "Testability Today", *op. cit.*, p. 30.

38 Cf. SUPPE F., *The Semantic Conception of Theories*, *op. cit.*, p. 71.

mais dépend également d'un modèle du contexte opératoire du système cognitif, à la fois dans sa dimension fonctionnelle et représentationnelle. Tout d'abord, nous verrons que cette prise en compte du contexte modifie le rôle des hypothèses auxiliaires dans la modélisation. En particulier, les lois d'évolution du système seront différentes selon les hypothèses auxiliaires qui spécifient le contexte opératoire. Ensuite, nous montrerons comment l'importance de la prise en compte du contexte de modélisation modifie la problématique de la spécification du domaine d'application des formalismes.

3.2.1. Le caractère contextuel de la modélisation de l'intentionnalité

Nous pouvons illustrer l'importance du caractère contextuel de la modélisation des opérations intentionnelles à partir des exemples de modèles du contexte représentationnel et fonctionnel étudiés dans le premier chapitre. Une première série d'exemples que nous avons étudiée dans le premier chapitre porte sur la spécification du contexte représentationnel du système. Ainsi, nous avons étudié l'exemple de la tâche de la reconnaissance de pentes dangereuses chez le jeune enfant. Les opérations qui permettent de réaliser cette fonction à partir de l'espace de locomotion d'un enfant rampeur ne sont pas les mêmes que celles utilisées dans une représentation spatiale différente, par exemple à partir de l'espace moteur d'un enfant qui marche. Dans le premier chapitre nous avons également présenté des exemples où les opérations cognitives sont différentes selon le modèle du couplage. Les modèles de coordination du mouvement chez l'enfant par exemple seront très différents selon les dynamiques corporelles intrinsèques sur lesquelles s'appuie le système de coordination. Le but poursuivi par le système de régulation sera différent selon que le problème auquel l'enfant est confronté est d'atténuer la dynamique de son corps ou, au contraire, d'engendrer suffisamment d'élan pour réaliser un certain mouvement.

A la différence de la physique, les hypothèses auxiliaires qui permettent de définir les modèles particuliers modifient donc également les lois d'évolution du système. Les opérations computationnelles sont différentes selon les modèles du contexte représentationnel et fonctionnel. Remarquons que ceci ne serait pas le cas dans le cadre d'une perspective computationnelle classique, comme par exemple dans la théorie de la vision de David Marr. Chez Marr, les hypothèses auxiliaires qui portent sur le contexte représentationnel et fonctionnel se limitent à la spécification des variables d'entrée du système et à la spécification de la signification des opérations algorithmiques par rapport à une fonction computationnelle. Nous retrouvons donc chez Marr l'idée physicaliste de la définition des conditions frontières, sous la forme de la spécification des variables d'entrée (les

données) et la spécification des variables de sortie (la fonction computationnelle) du système. Comme en physique, la modification des conditions frontières pour un système donné ne modifie pas les mécanismes computationnels. Par exemple dans le premier chapitre nous avons vu qu'une fonction de reconnaissance de visages ou de phonèmes dans un réseau de neurones est réalisée par le même algorithme de catégorisation de données, basé sur les principes de l'apprentissage associatif et de la généralisation par similarité. La théorie de la vision de Marr reste donc une théorie physicaliste en ce qu'elle ne prend en compte ni la définition du contexte représentationnel par les opérations du système, ni l'importance du rôle des fonctions dans la définition des buts poursuivis par le système.

Pour comprendre l'origine du caractère contextuel de la modélisation de l'intentionnalité en sciences cognitives, il faut donc comprendre les effets sur la modélisation de la prise en compte de la dimension représentationnelle et de la dimension fonctionnelle. Nous aborderons d'abord la question de la spécificité des modèles fonctionnels, en rapport avec le concept de système téléonomique. Ensuite, nous montrerons quelle modification supplémentaire il faut introduire quand on tient compte également de la dimension représentationnelle dans la modélisation.

Dans le deuxième chapitre de ce travail, nous avons étudié la spécificité de la modélisation des fonctions, à partir du concept de système téléonomique. Les systèmes téléonomiques sont des systèmes qui tendent vers un certain but selon la dynamique de l'interaction avec l'environnement, interaction qui est définie par des boucles de rétroaction. Cependant, comme le montre Frederic Suppe, à partir d'une analyse des lois téléonomiques, l'étude de la seule dynamique de l'interaction ne suffit pas pour connaître le but vers lequel tendra le système³⁹. Ce but dépend également des conditions de stabilisation de la dynamique d'interaction, conditions qui sont réalisées par la présence de certains éléments stables dans l'environnement avec lequel le système interagit.

Un exemple intéressant de l'importance des conditions de stabilisation des systèmes téléonomiques est donné par la théorie de la sélection naturelle en biologie. D'après la théorie de la sélection naturelle, la compétition des espèces dans un écosystème entraîne la disparition des espèces les moins adaptées et la survie des espèces les plus adaptées. Toutefois, cette adaptation dépend de l'environnement dans lequel on opère. Si par exemple on modifie artificiellement les sources de nutrition dans l'écosystème, on pourra obtenir soit une augmentation, soit une diminution de la biodiversité dans le système. L'état d'équilibre vers lequel tend le système dépend des

39 *Ibid.*, pp. 162–165.

contraintes présentes dans l'environnement. Cet exemple illustre bien que, dans l'étude des systèmes téléonomiques, les éléments de l'environnement avec lesquels le système interagit font partie du modèle. Et, comme le précise bien Suppe "pour un certain but donné, spécifié par une fonction, [...] la définition des conditions frontières qui portent sur les paramètres de stabilisation dans l'environnement implique une nouvelle loi d'évolution du système"⁴⁰. Dans le cas des systèmes téléonomiques, les lois d'évolution du système varient selon les contraintes présentes dans l'environnement.

C'est précisément l'importance des conditions de stabilisation dans la modélisation en sciences cognitives que nous avons pu étudier dans le paragraphe du premier chapitre consacré à la détermination écologique. Cependant, si l'on veut tenir compte également de la dimension représentationnelle, en plus de la dimension téléonomique, l'analyse de Suppe doit être modifiée. En effet, dans le cas des systèmes cognitifs, le contexte de fonctionnement n'est pas l'environnement naturel mais un contexte représentationnel ou un milieu de vie. Comme nous l'avons vu, l'environnement des systèmes cognitifs est un environnement représentationnel construit à partir de schèmes d'encodage généraux implémentés dans le cerveau. Un même environnement stabilisateur sera donc représenté d'une façon différente selon la niche sémantique du système cognitif en question. Par conséquent, les effets du contexte stabilisateur sur le système cognitif seront également différents selon le contexte représentationnel en question.

3.2.2. Le rôle de l'interprétation

Nous pouvons tirer deux conclusions de notre discussion de l'importance de la prise en compte du contexte dans l'étude de l'intentionnalité. Premièrement, comme nous l'avons vu à partir des exemples de la représentation de l'espace et de la coordination motrice chez l'enfant, les lois d'évolution du système seront différentes selon le modèle du contexte. Deuxièmement, en approfondissant notre analyse de la dimension fonctionnelle à partir de l'importance des conditions de stabilisation, nous avons pu préciser l'interdépendance entre les modèles du contexte fonctionnel et du contexte représentationnel : le rôle du système de représentation varie

40 *Ibid.*, p. 165 ; notre traduction. D'après cette analyse, les lois d'évolution des systèmes téléonomiques contiendront donc deux types de variables différentes, les variables qui restent constantes (les buts ou les éléments stabilisateurs de la dynamique d'interaction) et les variables dont la valeur change au cours de la dynamique de l'interaction. D'un point de vue formel, on aura donc l'équation suivante : $(X_1 \dots X_i X_{i+1} \dots X_p)_{t=t_0} \rightarrow (X'_1 \dots X'_i X_{i+1} \dots X_p)_{t=t_1}$ avec $X'_1 = X_1 \dots X_i = X'_i$ (cf. *ibid.*, p. 164).

en fonction du contexte fonctionnel (modification de la dynamique de stabilisation) et les effets du contexte fonctionnel dépendent du système de représentation (modification de la signification du contexte).

Cette interdépendance des facteurs contextuels permet de construire des interprétations alternatives d'une même expérience. Par exemple, dans l'expérience de la coordination motrice chez l'enfant, nous avons fait varier les modèles du couplage fonctionnel, mais nous avons supposé le modèle du système de représentation du corps constant. En l'occurrence, dans ces modèles du couplage fonctionnel, le système cognitif représente uniquement les aspects du corps qui interviennent dans la modulation du mouvement. Autrement dit la représentation du corps est une représentation pratico-sociale. Toutefois, on pourrait construire une autre interprétation de la même expérience en modifiant à la fois le modèle du contexte représentationnel et le modèle du couplage. Ainsi, l'expérience de la coordination motrice chez l'enfant peut être modélisée également à partir d'un modèle mentaliste de contrôle centralisé où le rôle du corps se réduit à fournir des informations à un système de représentation exhaustive du mouvement.

Le choix entre différentes interprétations d'une même expérience n'est évidemment pas arbitraire. Il devra s'appuyer sur des hypothèses auxiliaires qui permettent de justifier tel ou tel modèle du contexte représentationnel et fonctionnel. Toutefois, en l'absence d'hypothèses auxiliaires bien établies, le choix entre les modèles ne sera plus déterminé par des critères scientifiques seuls. Dans ce cas, le choix entre les modèles sera fonction d'une interprétation de ce qu'est l'intentionnalité. Pour montrer que ce phénomène de l'interprétation n'est pas un phénomène marginal dans la modélisation de l'intentionnalité, nous reprenons deux exemples discutés dans ce travail où le choix entre les modèles a été abordé de façon explicite. Le premier exemple concerne l'étude du contexte représentationnel dans le cadre des expériences de V. S. Ramachandran et le deuxième est lié au choix entre des modèles de couplage à partir des concepts développés par Rodney Brooks. Remarquons au préalable que cette problématique de l'interprétation est souvent occultée dans la recherche en sciences cognitives. En effet, comme dans les exemples de la coordination motrice et de la représentation de l'espace chez l'enfant, on s'intéresse soit à la dimension fonctionnelle, soit à la dimension représentationnelle du contexte. Le modèle de l'autre dimension est présupposé à titre d'arrière-fond de la modélisation. C'est précisément en explicitant cet arrière-fond que l'on peut se rendre compte du facteur interprétatif qui guide la modélisation.

Le premier exemple, la théorie de Ramachandran montre l'importance de l'utilisation de représentations partielles et circonstanciées dans les

opérations cognitives. Ainsi, pour construire une représentation d'une surface de points mobiles — par exemple une représentation du dos d'un léopard en mouvement — on peut inférer le mouvement des points individuels à partir du mouvement d'ensemble, sans devoir représenter le mouvement de chaque point individuel de façon explicite. Toutefois, même si une telle solution est adéquate dans des contextes particuliers, elle pourrait tomber en défaut si on était confronté à une surface où tous les points ne se meuvent pas de façon solidaire. Pour éviter ce problème, on peut partir de modèles plus classiques, où l'on suppose la construction de représentations exhaustives de l'environnement. Ramachandran justifie son choix pour les représentations partielles en invoquant un critère d'utilité auquel doivent satisfaire les représentations : “On pourrait argumenter, plutôt, que la perception est essentiellement un “stock d'astuces” [...] qui n'ont pas été adoptées pour des raisons d'attrait esthétique ou d'élégance mathématique, mais simplement parce qu'elles marchaient (d'où le terme de théorie “utilitariste” de la perception)”⁴¹. Dans l'interprétation de Ramachandran, la perception est essentiellement un guide d'action efficace, par rapport à des tâches spécifiques.

Ramachandran argumente pour la plausibilité de sa position en invoquant, comme nous l'avons vu, des arguments tirés de la théorie de l'évolution et des arguments computationnels. Les premiers arguments visent à montrer l'importance du couplage des systèmes cognitifs à un environnement de vie. Les deuxièmes veulent établir la performance d'un système algorithmique basé sur des dispositifs de calculs circonstanciés. Toutefois, ces arguments sont simplement des hypothèses auxiliaires de sa modélisation et ne justifient pas entièrement le choix pour l'interprétation utilitariste. Une interprétation alternative plus classique, qui attribue au système perceptif la tâche de construire des représentations exhaustives de l'environnement reste également plausible. Les hypothèses auxiliaires qui pourraient appuyer une telle hypothèse mettraient davantage l'accent sur l'indépendance du système par rapport à l'environnement, plutôt que sur le couplage, et sur l'importance des capacités d'abstraction des systèmes cognitifs, plutôt que sur leur flexibilité contextuelle.

Un deuxième exemple de conflit d'interprétation qui joue un rôle central dans notre travail intervient dans le choix entre les modèles de couplage fonctionnel avec l'environnement. Dans le deuxième chapitre, nous avons discuté deux interprétations alternatives du couplage avec l'environnement de systèmes cognitifs composés de modules interactifs. Ces systèmes

41 RAMACHANDRAN V.S., “Interactions between Motion, Depth, Color and Form : the Utilitarian Theory of Perception”, *op. cit.*, pp. 346–360, p. 347.

peuvent être modélisés à partir d'une hypothèse de stabilisation des processus cognitifs internes sous contrôle de l'environnement, comme dans les recherches de Brooks sur les agents autonomes. Toutefois, l'organisation fonctionnelle des systèmes cognitifs en modules interactifs n'exclut pas une interprétation alternative du couplage, dans les termes d'une autostabilisation du système, sans lien fonctionnel avec l'environnement. Brooks argumente pour le modèle du contrôle environnemental, à partir d'un critère de simplicité des opérations et de rapidité de la réponse du système⁴². Nous retrouvons donc une interprétation de l'intentionnalité en termes d'action efficace, comme dans la théorie utilitariste de Ramachandran.

Une interprétation alternative, qui se base sur l'hypothèse de l'autostabilisation, est développée par exemple par Francisco Varela, Evan Thompson et Eleanor Rosch dans *L'inscription corporelle de l'esprit*⁴³. Dans leur discussion des recherches de Brooks, ils proposent une interprétation dans les termes du paradigme de l'enaction, qui vise à “embrasser la temporalité de la cognition entendue comme histoire vécue”⁴⁴. Toutefois, ils précisent également qu'une telle interprétation “dépend fortement du degré d'intérêt que nous avons à rester proches de la réalité biologique”⁴⁵. Cette réalité biologique, c'est la réalité de l'évolution conçue comme “dérive naturelle”, c'est-à-dire selon une conception de l'évolution où la relation à l'environnement est simplement une relation de “couplage viable”⁴⁶. Dans une telle conception, “l'environnement n'intervient plus [...] dans les explications que dans les occasions où les systèmes subissent des ruptures ou des événements auxquels leurs structures ne peuvent faire face”⁴⁷. De nouveau, on rencontre une certaine interprétation de l'intentionnalité dans les termes d'une histoire vécue, articulée à un ensemble d'hypothèses auxiliaires, ici en lien avec la théorie de l'évolution.

Ces exemples de conflits d'interprétation en sciences cognitives confirment l'importance des critères non scientifiques qui interviennent dans la modélisation de l'intentionnalité. Le champ des modèles particuliers ne peut donc être spécifié de façon univoque sans l'intervention d'hypothèses auxiliaires, venant d'autres disciplines, qui appuient une certaine interprétation de ce qu'est l'intentionnalité. L'analyse de la dimension sémantique des formalismes dans ce chapitre montre donc non seulement la

42 Cf. CLARK A., *Being There*, op. cit., p. 21.

43 VARELA F., THOMPSON E. et ROSCH E., *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, op.cit., pp. 282–289.

44 *Ibid.*, p. 289.

45 *Ibid.*, p. 288.

46 *Ibid.*

47 *Ibid.*, p. 281.

nécessité, pour un formalisme donné, de spécifier le domaine d'application de ce formalisme dans les termes d'une famille de modèles, mais également le fait que ce domaine d'application peut avoir une signification contextuelle propre, indépendante du formalisme, déterminée par une interprétation de ce qu'est l'intentionnalité.

4. Pour un traitement réflexif de la contextualisation : la solution phénoménologique et sa radicalisation

Dans ce dernier paragraphe, nous voulons tenter de montrer l'apport d'une *radicalisation* de notre questionnement sur les limitations de la modélisation de l'intentionnalité en sciences cognitives. Nous procéderons à partir d'une mise en question de l'abstraction méthodologique qui caractérise le traitement formel de l'intentionnalité en sciences cognitives : la réduction naturaliste. La radicalité de cette critique consistera à s'appuyer sur un concept réflexif de contextualisation, tel qu'il se dégage du dépassement phénoménologique de la position cognitiviste à l'égard des limitations de sa propre modélisation de l'intentionnalité. Pour introduire à cette radicalisation, reprenons d'abord brièvement le parcours effectué dans ce chapitre. Tout d'abord, nous avons étudié la fécondité sémantique des formalismes en sciences cognitives et la limitation contextuelle de cette fécondité, le renvoi à une famille de modèles particuliers. Plus précisément, l'étude de la dimension sémantique nous a amené à mettre en évidence la dualité des formalismes et des modèles, ou encore l'inadéquation entre les principes formels, en nombre fini, et le champ concret des modèles, en nombre infini : pour un formalisme donné, il y aura toujours différentes familles de modèles que l'on peut proposer comme domaine d'application de ce formalisme. D'où l'importance de la dynamique de la vérification des modèles particuliers pour délimiter un domaine d'application spécifique des formalismes. Nous avons vu que ce domaine ne peut être spécifié de façon univoque pour la modélisation de l'intentionnalité. En effet, les modèles des opérations intentionnelles dépendent d'hypothèses auxiliaires, qui portent sur le contexte représentationnel et fonctionnel du système cognitif. En absence d'hypothèses auxiliaires bien établies, le choix entre les modèles sera fonction d'une interprétation de ce qu'est l'intentionnalité.

La problématique de l'interprétation montre bien que le domaine des modèles particuliers de l'intentionnalité ne peut être spécifié par des critères scientifiques seuls. Toutefois, si l'on peut saisir le champ des modèles particuliers dans sa signification contextuelle propre, on peut se poser la question du sens de l'entreprise de formalisation de l'intentionnalité, en tant

que ce sens est différent du champ des modèles particuliers. L'usage contextuel déterminé des modèles particuliers d'un formalisme n'est qu'un élément dans la dynamique d'élargissement, d'extension et de modification des modèles de l'intentionnalité que l'on a rencontrée tout au long de ce travail. Afin de saisir cette dynamique pour elle-même et donc de comprendre le sens de l'entreprise de la formalisation, on ne peut pas avoir recours de nouveau à une démarche de modélisation — puisqu'il s'agit de comprendre précisément le sens d'une telle démarche — et il faudra donc sortir de l'attitude naturaliste. Nous retrouvons de la sorte la deuxième limitation de la formalisation de l'intentionnalité mise en évidence par Atlan dans *Projet et signification*, la limitation phénoménologique. Cette limitation porte sur l'impossibilité de rendre compte de la problématique du sens à l'intérieur de l'attitude naturaliste et la nécessité d'adopter une autre attitude, l'attitude phénoménologique, qui nous donne accès à la constitution réflexive de l'horizon de sens qui motive l'entreprise de formalisation. C'est ce dernier niveau de limitation qui nous permettra d'aborder la question du sens ou de la pertinence du formalisme dans la modélisation de l'intentionnalité.

Dans ce paragraphe, nous allons tenter de montrer la nécessité de radicaliser le concept de limitation phénoménologique tel qu'il est défini par Atlan dans l'article *Projet et signification*. En confrontant la perspective de Atlan avec celle de la phénoménologie husserlienne sur laquelle elle prend appui, nous verrons que celle-ci reste bien en deçà de la phénoménologie transcendante de Husserl. En effet, l'attitude phénoménologique adoptée par Atlan, en mettant l'accent sur le rôle productif de la réflexivité comme limitation de l'attitude objectivante, reste au niveau d'une eidétique de la manifestation du sens, dans son apparaître multiple et indéterminé. Pour donner une réponse satisfaisante à la question du sens du formalisme, il faudra dépasser la limitation phénoménologique envisagée par Atlan. En particulier, nous montrerons que dans la perspective phénoménologique husserlienne le sens n'est jamais simplement un domaine de manifestations effectives, mais toujours en même temps un horizon eidétique, qui est la source par laquelle se manifeste la phénoménalité originare pour la conscience⁴⁸. En évacuant la question de l'origine de la manifestation, Atlan évacue en même temps toute la problématique de la réduction phénoménologique transcendante et, à partir de là, il manque le dépassement que permet d'opérer la phénoménologie husserlienne de la question des limitations vers l'autoconscience possible des donations de sens

48 Cf. MAESSCHALCK M. et KOKOSZKA V., "Phénoménologie et auto-organisation", in FELTZ B., CROMMELINCK M. et GOUJON PH. (eds.), *Auto-organisation et émergence dans les sciences de la vie*, Bruxelles, 1999, pp. 405–420.

dans la vie de la conscience. Nous voudrions montrer que c'est précisément un tel dépassement que permettent d'opérer les travaux de Husserl sur la logique transcendantale dans les années vingt, en concentrant plus particulièrement notre attention sur l'ouvrage *Logique formelle et logique transcendantale*⁴⁹.

4.1. Le dépassement de la question des limitations dans la phénoménologie génétique

On peut déjà s'apercevoir du dépassement de la question du sens de l'entreprise de formalisation comme limitation dans la phénoménologie husserlienne au niveau des *Recherches logiques*, par le thème du retour aux choses elles-mêmes, et au niveau des *Ideen I*, où ce thème est retravaillé dans le cadre de la réduction phénoménologique transcendantale⁵⁰. Toutefois, dans le cadre des travaux des années vingt sur la phénoménologie de la genèse, ce dépassement prend une tournure plus radicale. En effet, dans ces travaux, la question de la source originaire du sens de l'évidence logique formelle est remise sur le chantier à partir d'une interrogation sur la production originaire de l'évidence dans le flux concret de la conscience. Par là, la question de l'évidence peut être abordée pour elle-même, au niveau des synthèses passives dans le flux, et non plus simplement en rapport à la recherche d'un soubassement fondamental pour les opérations logiques supérieures⁵¹.

Pour articuler la perspective ouverte par les travaux des années vingt et la question du dépassement du concept de limitation phénoménologique, nous étudierons plus particulièrement un ouvrage clef publié dans cette période, l'ouvrage *Logique formelle et logique transcendantale*. L'intérêt de cet ouvrage par rapport à la problématique des limitations est double.

49 HUSSERL E., *Logique formelle et logique transcendantale. Essai d'une critique de la raison logique*, trad. par Suzanne Bachelard, Paris, 1965 (l'ouvrage sera cité *Logique formelle et logique transcendantale*) ; pour le texte original nous utilisons l'édition critique, HUSSERL E., *Husserliana XVII : Formale und transzendente Logik. Versuch einer Kritik der logischen Vernunft. Mit ergänzenden Texten*, édité par P. Janssen, Den Haag, 1974.

50 Cf. BRISART R., "La réduction et l'irréductible phénoménologiques. Husserl critique de Heidegger", in BRISART R. et CÉLIS R., *L'évidence du monde. Méthode et empirie de la phénoménologie*, Bruxelles, 1994, pp. 139–185.

51 Cf. HUSSERL E., *De la synthèse passive. Logique transcendantale et constitutions originaires*, trad. par Bruce Bégout et Jean Kessler avec la collaboration de Natalie Depraz et Marc Richir, Grenoble, 1998, p. 10 de l'introduction par Bruce Bégout et Natalie Depraz. Comme d'usage, et de manière générale, nous renverrons à l'édition critique des œuvres de Husserl par le sigle Hua.

Premièrement, en tant qu'il tente de tracer le chemin qui va de la logique formelle à la logique transcendantale, il permet de montrer la genèse de l'évidence de la logique formelle, ou *mathesis universalis*, à partir d'évidences plus originaires, au niveau de l'expérience perceptive du monde. Ce lien génétique entre l'évidence de la logique formelle et l'évidence du monde est simplement évacué dans la réflexion sur les limitations de la formalisation en sciences. En effet, cette réflexion problématise uniquement l'attitude naturaliste, qui est celle des sciences de la nature, sans voir le lien originaire de cette attitude avec l'attitude naturelle, qui caractérise notre confiance naturelle dans les évidences du monde de l'expérience. Par là, la limitation phénoménologique de la modélisation ne problématise pas l'évidence naturelle du monde et évacue tout le domaine de réflexion auquel nous donne accès la réduction phénoménologique transcendantale. Deuxièmement, le questionnement dans Logique formelle et logique *transcendantale* sur la genèse de l'évidence logique montre que cette évidence s'appuie sur une production originaire de l'évidence au niveau de la passivité. La conscience intentionnelle n'est donc pas pur pouvoir de projection de sens, comme chez Atlan, mais également épreuve passive d'un pouvoir-être plus originaire.

Le point de départ de l'analyse phénoménologique du sens du formalisme dans *Logique formelle et logique transcendantale* est la mise en évidence des présuppositions idéalisantes qui accompagnent l'activité logique formelle. C'est à partir de ces présuppositions idéalisantes que l'on peut caractériser l'évidence logique et son lien génétique avec les évidences du monde de l'expérience. Avant d'étudier le dépassement de la problématique des limitations à partir de la mise entre parenthèses de l'évidence du monde, il nous faudra donc caractériser cet horizon d'idéalités tel qu'il est mis en évidence par l'analyse phénoménologique. Notre analyse de la question du sens dans *Logique formelle et logique transcendantale* se fera alors en trois étapes. Premièrement, nous étudierons l'analyse que fait Husserl de la logique formelle moderne et de ses présuppositions. Deuxièmement, nous examinerons le lien génétique avec les évidences du monde de l'expérience perceptive et le chemin qui conduit vers la réduction phénoménologique transcendantale. Finalement, nous verrons comment la réflexion phénoménologique ouvre un questionnement sur l'épreuve originaire du sens au niveau des synthèses passives dans le flux de la conscience.

4.1.1. L'horizon eidétique de la logique formelle moderne

Pour caractériser les présuppositions idéalisantes qui accompagnent l'activité logique formelle d'un point de vue phénoménologique, nous partirons de la réflexion sur le sens de la logique formelle moderne dans *Logique formelle et logique transcendantale*. Le dessein de l'ouvrage, tel que Husserl l'annonce dans l'introduction, est de réfléchir sur la possibilité de principe des sciences en général et en particulier sur la logique en tant qu'elle donne des normes universelles au savoir scientifique. D'abord Husserl indique une similarité d'un tel projet avec le projet des *Méditations métaphysiques* de Descartes, projet radical d'une théorie de la science universelle, précédant les sciences particulières et garantissant leur possibilité. Toutefois le chemin poursuivi dans cet ouvrage ne sera pas aussi radical que celui des *Méditations*, mais sera un chemin plus graduel qui part d'une science particulière, ici la logique formelle, pour en expliciter étape par étape la constitution originaire.

4.1.1.1. La manifestation effective du sens : les orientations fondamentales de la logique formelle moderne

Dans la première partie de l'ouvrage, Husserl décrit le champ et les structures de la logique formelle objective. Le but d'une telle description est de dégager le sens du projet d'élaboration d'une logique formelle, *tel qu'il se manifeste "dans l'intention dominante des mathématiciens"*⁵². Ce sens, c'est celui "d'une *analytique pure de la non-contradiction* dans laquelle le concept de vérité n'entre pas comme thème"⁵³. La logique, dans la pratique scientifique courante, remplit le rôle d'une critique de la connaissance en tant qu'elle garantit — et vérifie — la cohérence logique du discours, son absence de contradiction interne. La consistance du discours scientifique ne lui procure évidemment pas encore la vérité, mais permet de remplir au moins une condition nécessaire *a priori* de tout discours vrai possible.

Le sens de la logique formelle moderne réside donc dans l'analyticité de son domaine de propositions et de déductions. Dans un appendice à *Logique formelle et logique transcendantale*, Husserl précise que par le concept d'analytique il faut entendre "la compossibilité de jugements en tant que tels [...], compossibilité dans l'unité de l'effectuation de jugement qui est

52 *Logique formelle et logique transcendantale*, p. 17 ; Hua XVII, pp. 15–16. C'est nous qui soulignons.

53 *Ibid.*

effectuation explicite et proprement dite⁵⁴. Cet usage large du concept d'analytique vise en fait à couvrir toute activité de jugement dont la validité est indépendante du contenu et qui dépend uniquement de l'activité du jugement en tant que tel. Par là, la logique formelle telle que Husserl l'entend ne contient pas uniquement la logique formelle traditionnelle, mais également les mathématiques formelles modernes. En effet, comme le montre Husserl, avec la formalisation des mathématiques modernes, la séparation entre la logique formelle — l'étude formelle du jugement — et les mathématiques formelles — l'étude formelle de *quelque chose en général* — a disparu. En combinant la logique formelle traditionnelle et les mathématiques formelles, on obtient alors l'idée d'une logique formelle généralisée ou *mathesis universalis*⁵⁵.

A côté de cette première orientation qui caractérise l'activité du mathématicien en tant que telle, l'orientation vers l'analyticité, Husserl distingue également deux autres orientations. Tout d'abord, les raisonnements et les constructions de propositions peuvent également se faire de façon automatique, simplement en respectant la grammaire de base du système formel et sans penser de façon effective à leur analyticité. Ce type d'activité n'est plus une activité de connaissance à proprement parler, mais nous pouvons néanmoins encore parler de production (impropre) de propositions formelles et d'une certaine unité de la pensée. Cette première orientation, c'est l'orientation vers la simple grammaticalité. Ensuite, Husserl distingue l'orientation qui caractérise le scientifique qui veut utiliser la logique formelle pour la connaissance en vérité d'un certain domaine d'objets. Pour ce scientifique se pose la question de l'interprétation du système formel, au-delà des considérations sur son analyticité, en vue de caractériser la vérité et la fausseté des propositions.

De cette façon, on obtient donc trois orientations fondamentales ou propriétés essentielles que l'on peut distinguer dans le domaine de la logique formelle : l'orientation vers la simple grammaticalité, l'orientation vers l'analyticité et l'orientation vers la vérité. Ces propriétés fondamentales déterminent les idées essentielles qui gouvernent la logique formelle moderne, ou encore l'*eidōs* des idéalités de la logique formelle⁵⁶. Evidemment, si l'on veut utiliser la logique comme une pièce fondamentale d'une théorie de la science, ces trois orientations devraient être combinées : la grammaticalité est une condition nécessaire pour l'analyticité possible des

54 *Ibid.*, p. 426 ; Hua XVII, p. 329.

55 Cf. *ibid.*, pp. 102–105 ; Hua XVII, pp. 78–80.

56 Sur le concept d'*eidōs* dans *Logique formelle et logique transcendantale*, cf. *ibid.*, pp. 329–335 ; Hua XVII, pp. 252–257.

propositions et l'analyticité est une condition nécessaire pour la vérité possible des propositions.

4.1.1.2. *Les présuppositions idéalisantes de la logique formelle moderne*

Après avoir dégagé le sens de la logique telle qu'il se manifeste dans la logique comme pratique culturelle objective, Husserl poursuit, dans la deuxième partie de l'ouvrage, la méditation sur le sens (*Besinnung*) de la logique formelle, pour amener ce sens à l'état de sens authentique, pleinement assuré de ses conditions de possibilité. Comme le précise Husserl dans l'introduction, il s'agit de "l'élucidation du sens "authentique" des concepts logiques et de la logique elle-même — non pas d'un sens qui précède la théorie de la connaissance et qui est déjà là, mais d'un sens qui ne doit être créé que par la théorie de la connaissance et qui ne doit être exploré que dans les horizons de sa propre portée"⁵⁷. En effet "les sciences, par la non-compréhension de leurs propres effectuations en tant qu'effectuations d'une intentionnalité effectuant qui reste pour elles non-thématique, sont incapables de clarifier le sens-d'être authentique de leur domaine et des concepts qui les expriment — elles sont donc incapables de dire, au sens propre et dernier, quel sens a l'existant dont elles parlent, quels horizons du sens cet existant présuppose, horizons dont elles ne parlent pas et qui pourtant contribuent à déterminer ce sens"⁵⁸.

Le passage d'une considération des sciences dans leur positivité naïve à une élucidation de leur sens authentique est désigné par Husserl dans la deuxième partie de l'ouvrage de façon plus précise comme le passage d'une critique *analytique* de la connaissance à une critique *transcendantale* : "À côté de la critique des données préalables, des processus et des résultats qui apparaissent de façon manifeste dans le champ de la conscience, nous avons affaire ici à une tout autre sorte de critique de la connaissance, à celle des sources originelles constitutives du sens positionnel et du droit de ces données, processus et résultats, donc à la critique des *effectuations cachées* dans l'activité de recherche et de théorétisation qui est tournée immédiatement vers le domaine. C'est la critique de la *raison*, ou, [...] la *critique transcendantale de la connaissance*"⁵⁹. Dans l'activité quotidienne du scientifique, la possibilité de sa démarche est constamment présupposée de façon implicite. La critique transcendantale demande un changement du regard, un détournement de la présence naïve auprès du domaine pour

⁵⁷ *Ibid.*, p. 20 ; Hua XVII, p. 18.

⁵⁸ *Ibid.*, pp. 19–20 ; Hua XVII, p. 17.

⁵⁹ *Ibid.*, pp. 231–232 ; Hua XVII, p. 179.

s'interroger sur sa constitution originaire. Par l'introduction du concept de *critique transcendantale*, nous dépassons donc une simple conception du sens comme domaine de manifestations effectives, pour nous interroger sur la constitution originaire de ces manifestations. La première étape de la critique transcendantale de la raison logique consistera alors dans une enquête sur les *effectuations cachées* présupposées dans l'activité de connaissance logique.

Une première série de présuppositions mise en évidence par Husserl concerne les propositions en tant que simples unités grammaticales. Il s'agit des présuppositions les plus fondamentales de la logique, en ce que la grammaticalité est une condition nécessaire aussi bien pour l'analyticité que pour la vérité possible. Les propositions, comme de simples unités grammaticales, sont manipulées dans les activités logiques comme des entités qui sont identifiables par delà leur donnée intuitive. En effet, les objets et les résultats de la logique sont donnés à titre de "formations durables, [...] des formations qu'on peut *toujours à nouveau réactiver et identifier*"⁶⁰. Par exemple, au cours d'une démonstration, je forme une certaine proposition, je manipule une autre, je reviens à la première proposition, et ainsi de suite. Même si je ne vise pas la proposition initiale de façon active tout au long de la démonstration, au moment où elle revient dans mon champ d'attention je la vise en tant qu'étant la même proposition que celle qui a été produite de façon originelle au début de la démonstration. Ce qui est présupposé, c'est l'identité des propositions à travers de multiples apparitions dans le temps et même pour différentes personnes ou communautés.

Dans l'attitude naïvement immédiate du théoricien qui manipule ces propositions, l'identité durable des objets et des résultats de la logique ne pose aucun problème. Il considère les objets formels qu'il manipule comme des objets qui ont une identité en soi. Toutefois, si l'on se pose la question de la source constitutive de cet en-soi dans la vie du sujet, on se rend compte qu'il s'agit d'une idéalisation. Ainsi, quand nous revenons sur une proposition qui a été produite au préalable, "on en a conscience à nouveau par l'intermédiaire d'un ressouvenir"⁶¹. Mais pour restituer exactement la même proposition on devrait opérer "la restitution de tous les moments particuliers ou étapes particulières du processus originel, et même quand cela aurait lieu, [...], est-il sûr que ce serait la restitution de l'évidence antérieure ?"⁶². Par conséquent on est confronté au dilemme suivant : "ou bien la logique opère

60 *Ibid.*, p. 250 ; Hua XVII, p. 192.

61 *Ibid.*

62 *Ibid.*

avec une *fiction* universelle et alors elle n'est elle-même rien moins que donatrice des normes ; ou bien elle est science qui fournit des normes, alors cet idéal est précisément une *norme fondamentale effective*⁶³. Si on écarte la première solution, celle du sceptique, alors on est confronté à la situation suivante : la logique présuppose partout un idéal — la possibilité de viser ses objets et ses résultats *comme identique* à travers différentes apparitions particulières —, dont la justification transcende la donnée actuelle de ces objets.

La discussion de la présupposition d'identification durable met bien en évidence le caractère idéalisant des effectuations cachées dans l'activité de la connaissance logique. C'est pourquoi Husserl parle des "présuppositions idéalisantes"⁶⁴ de la logique formelle moderne. Après la présupposition de l'orientation vers la grammaticalité, Husserl considère ensuite les présuppositions fondamentales qui accompagnent l'orientation vers l'analyticité et vers la vérité possible. Tout d'abord, l'orientation vers l'analyticité présuppose la possibilité de pouvoir amener tout jugement ou son contraire à l'état de jugement non-contradictoire⁶⁵. Ensuite, l'orientation vers la vérité possible présuppose la possibilité d'attribuer à tout jugement le prédicat vrai ou faux⁶⁶. Elle présuppose donc implicitement la possibilité de pouvoir décider de la vérité et de la fausseté de l'ensemble des propositions interprétées du système formel. Comme cet ensemble est infini, une telle présupposition a clairement un caractère idéalisant.

Cette brève discussion des présuppositions idéalisantes de la logique nous apporte une clef importante de la méthode qui mènera vers la réduction phénoménologique transcendantale. Dans sa discussion, Husserl caractérise les présuppositions comme étant des *normes fondamentales effectives*. Cette précision sur le statut des présuppositions indique clairement le but qui est visé par la *critique transcendantale* telle que Husserl veut l'entreprendre. Ce but, ce n'est pas tellement de *fonder* les normes que de les *comprendre* dans leur effectivité, ou encore de les expliciter dans leurs horizons constitutifs. Le rôle de la mise en évidence des présuppositions d'arrière-fond est de décrire les normes dans leur vie même, telles qu'elles donnent sens à l'activité logique. La critique transcendantale devra ensuite retracer la genèse de l'assomption motivée de telles normes dans la vie du sujet. C'est par cette critique transcendantale que nous retrouverons alors un horizon de sens et des synthèses d'identification plus originaires.

63 *Ibid.*, p. 253 ; Hua XVII, p. 195.

64 *Ibid.*, p. 248 ; Hua XVII, p. 191.

65 Cf. *ibid.*, § 88 ; Hua XVII, pp. 222–223.

66 Cf. *ibid.*, § 79 ; Hua XVII, pp. 203–205.

4.1.2. La réduction phénoménologique transcendantale

Nous pouvons maintenant aborder le premier dépassement que permet d'opérer la phénoménologie husserlienne par rapport à la problématique des limites de l'attitude naturaliste, le dépassement opéré par la réduction phénoménologique transcendantale. Les chemins qui mènent vers la réduction transcendantale sont multiples, et ce n'est pas le lieu ici d'en discuter. Rappelons toutefois que le chemin poursuivi dans *Logique formelle et logique transcendantale* prend comme *fil conducteur* une science particulière, la logique formelle moderne, pour expliciter la constitution originnaire du sens de cette science. Comme nous l'avons vu, c'est ce chemin qui commande la structure générale de l'ouvrage : dans la première partie Husserl clarifie le sens d'un domaine d'objets particuliers — les données et les résultats de la logique formelle considérée dans son objectivité immédiate — et dans la deuxième partie Husserl entreprend une réflexion critique sur les sources originelles constitutives du sens positionnel et du droit de ces données et résultats, appelée critique transcendantale ou critique constitutive. La poursuite de cette critique dans cette deuxième section nous mènera en un premier temps des activités logiques à l'horizon du monde de l'expérience et, en un deuxième temps, à la mise en suspens de toute croyance dans la validité du monde, considéré comme monde de vérités existant en soi.

4.1.2.1. De l'horizon de la logique formelle à l'horizon du monde de l'expérience

La mise en évidence du caractère idéalisant des présuppositions de la logique indique que l'on devra dépasser la sphère de la logique pour retrouver la constitution originnaire de ces présuppositions. En effet, les présuppositions sont des idéaux qui ne peuvent jamais être satisfaits par un acte logique particulier et elles ne peuvent donc trouver leur justification dans la logique elle-même.

Pour dépasser la sphère de la logique, Husserl montre d'abord que les présuppositions de la logique formelle sont déjà présentes au niveau des jugements logiques élémentaires, les jugements qui portent sur des objets individuels donnés dans l'expérience. En effet, déjà au niveau des jugements d'expérience, nous constatons la production de formes grammaticales identifiables par delà leur donnée intuitive. Dans l'activité du jugement, des substrats de jugement reçoivent une forme logique objective (une détermination prédicative, une détermination relationnelle, etc.), forme qui est identifiable, à laquelle on peut revenir, que l'on peut communiquer, etc.

On retrouve donc la présupposition idéalisante attachée à la visée de grammaticalité dans la logique formelle. De même, nous pourrions retrouver les présuppositions idéalisantes attachées à la visée de l'analyticité ou de la vérité possible des jugements. La tâche de la critique constitutive sera alors de montrer la genèse de ces horizons de présuppositions déjà présents au niveau du jugement d'expérience à partir des activités intentionnelles de l'expérience elle-même.

Or, la détermination de vécus d'expérience d'après les formes logiques du jugement s'appuie constamment sur la présupposition de l'existence d'un monde d'expérience pré-donné, potentiellement déterminable par les formes du jugement. Tout d'abord, la visée d'identification durable attachée à la visée de la grammaticalité renvoie à un monde d'expérience dont on peut extraire des formes logiques stables, identifiables de façon durable. Ensuite, les présuppositions attachées à l'analyticité et à la vérité possibles renvoient respectivement à un domaine d'expérience potentiellement unitaire — donnant lieu à des substrats et des jugements *compossibles* — et à l'existence d'un domaine d'objets et de vérités objectives existant en soi. De la question de l'horizon des présuppositions de la logique on est donc renvoyé à la question des horizons plus originaires du monde de l'expérience, à la constitution originaire de visées d'identification, d'unification et d'adéquation comme pôles intentionnels des vécus d'expérience.

La mise en évidence de ce lien génétique entre les activités intentionnelles de la logique et les activités intentionnelles de l'expérience n'est pas développée dans le détail dans *Logique formelle et logique transcendantale*. Ce n'est qu'au niveau des travaux publiés dans *Expérience et jugement*⁶⁷ que Husserl entreprend l'analyse systématique de l'explicitation des horizons du monde de l'expérience et de leur lien constitutif avec les formes logiques du jugement. Ici — et c'est probablement aussi un des enjeux fondamentaux de *Logique formelle et logique transcendantale* — il s'agit de bien comprendre le sens de cette entreprise. En effet, la question de la genèse subjective des activités de la logique pourrait donner l'impression d'une réduction psychologisante de ces activités. Selon une telle interprétation, les formes logiques trouveraient leur origine dans les structures du sujet connaissant, et plus précisément dans les activités qui structurent le monde de l'expérience.

67 HUSSERL E., *Expérience et jugement. Recherches en vue d'une généalogie de la logique*, trad. par Denise Souche-Dagues, Paris, 1970 (l'ouvrage sera cité *Expérience et jugement*) ; trad. de HUSSERL E., *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, rédigé et édité par L. Landgrebe, Hamburg, 1954.

A partir des développements qui précèdent sur le caractère idéalisant des présuppositions d'arrière-fond, il devrait déjà être clair qu'un tel traitement psychologisant du concept d'horizon est très problématique. Toutefois, il nous semble que ce n'est qu'en combinant les deux volets de *Logique formelle et logique transcendantale* que l'on peut véritablement résister à une telle reprise psychologisante du projet husserlien. Ce n'est qu'à partir des développements de la première partie sur la logique dans sa positivité objective, que l'on peut saisir ensuite dans la deuxième partie le caractère spécifique de l'intentionnalité d'horizon comme intentionnalité constituante d'objectivités avec leur consistance propre. Ou, pour le formuler autrement, les présuppositions idéalisantes de la logique sont des caractères d'objets autant que des visées intentionnelles. La grammaticalité, l'analyticité et la vérité sont des caractéristiques qui, malgré le fait qu'elles soient des idéalités, sont visées comme étant des caractéristiques des propositions elles-mêmes. De même les visées d'identification, d'unification et d'adéquation au niveau du monde de l'expérience sont des visées qui concernent l'identité durable, l'unité et la vérité des objets d'expérience. L'horizon des présuppositions est donc un horizon que l'on peut qualifier de *transcendant dans l'immanence*. Il a une origine constitutive dans la vie du sujet, mais en même temps il se constitue dans cette vie comme horizon objectif.

4.1.2.2. L'expérience perceptive comme donation des choses dans leur être même

La critique constitutive nous a ramené de l'horizon de présuppositions idéalisantes des activités logiques à un horizon de présuppositions plus originaires, au niveau du monde de l'expérience. Ou encore, on est ramené de l'eidétique des activités logiques à l'eidétique du monde de l'expérience. Toutefois, une dernière présupposition reste encore attachée à cet horizon, qui nous empêche de mener à bien la tâche de la critique constitutive, c'est la présupposition de l'existence du monde, comme monde de vérités existant en soi.

C'est dans les derniers chapitres de *Logique formelle et logique transcendantale* que Husserl argumente en faveur d'une radicalisation de la question de l'horizon du monde. En effet, la mise en évidence de la structure d'horizon du monde ne résout pas encore la question de départ sur l'origine et la justification ultime de l'objectivité de la logique. En quelque sorte on est ramené de la question de l'objectivité de la logique à la question de l'objectivité du monde, ou encore, dans le cadre de notre réflexion phénoménologique, on est ramené de la critique constitutive de l'évidence logique à la critique constitutive de l'évidence de l'expérience. C'est à ce

point crucial de son exposé que Husserl précisera le sens de la critique constitutive de l'expérience, ici en l'occurrence par rapport à la critique cartésienne de l'expérience. De prime abord, on pourrait rapprocher le projet husserlien de celui de Descartes. En effet, la critique de l'évidence de l'expérience prend la forme, comme nous l'avons vu, de l'explicitation de l'horizon du monde comme pôle intentionnel des vécus de conscience. Husserl opère donc bien une réduction de l'expérience à la vie de la conscience. Toutefois, si nous entendons cette vie de la conscience à la manière cartésienne comme la vie d'une âme réelle, avec ses lois de pensée réelles, on ne pourrait jamais obtenir une justification ultime de la logique puisqu'on présuppose précisément ce qui est en question, la possibilité d'une vie de la pensée. Comme le dit Husserl : "Déjà chez Descartes l'ego est établi par une évidence absolue comme une *parcelle du monde*, première, existant indubitablement (*mens sive animus, substantia cogitans*) et toute la question est alors d'y ajuster, par un *procédé de déduction* logiquement concluant, le reste du monde"⁶⁸. La critique cartésienne de l'évidence de l'expérience se base sur une nouvelle évidence, l'existence en soi de l'âme avec ses lois de la pensée.

La réduction de l'expérience à la vie de la conscience, si elle veut élucider le sens authentique de la logique formelle, ne peut aucunement présupposer ce sens comme étant pré-donné dans un domaine de vérités en soi. Elle doit donc opérer une réduction phénoménologique transcendantale qui met en suspens toute évidence et toute vérité quelle qu'elle soit. C'est à ce moment là que l'on pourra expliciter l'évidence de façon originnaire "dans la pure intériorité de l'ego, en tant que pôle intentionnel de l'expérience"⁶⁹. Par la réduction transcendantale on dépasse un questionnement sur le sens lié aux limites d'un savoir mondain, que ce soit le savoir de la logique, le savoir de l'expérience naturelle du monde ou le savoir de la vie subjective de l'âme. La phénoménologie transcendantale n'est donc pas une science mondaine, au sens où elle serait une science qui porte sur un domaine de vérités en soi, mais elle est une science "avec des vérités essentielles, avec des théories qui valent exclusivement pour moi, pour moi l'ego"⁷⁰. Plus précisément "la *phénoménologie entière* n'est rien de plus que *la prise de conscience par soi-même de la subjectivité transcendantale* [*Selbstbesinnung der transzendentalen Subjektivität*], [...] une prise de conscience qui va du fait aux nécessités d'essence, au *logos primitif* [*Urlogos*] d'où procède tout ce qui a le statut du

68 *Logique formelle et logique transcendantale*, p. 306 ; Hua XVII, p. 235.

69 *Ibid.*, p. 310 ; Hua XVII, p. 238.

70 *Ibid.*, p. 359 ; Hua XVII, p. 276.

'logique''⁷¹. De la méditation sur le sens (*Besinnung*) authentique de la logique formelle on est donc reconduit à l'autoconscience (*Selbstbesinnung*) du sens dans les effectuations de la conscience absolue, dont la manière d'être est un "être-pour-soi-même"⁷², en dehors de tout préjugé portant sur l'existence de vérités *en soi*.

Par cette réflexion transcendantale sur le sens, Husserl dépasse clairement une réflexion sur le sens liée comme questionnement des limites d'un savoir mondain donné. Toutefois, avant de conclure, Husserl considère une objection que l'on pourrait soulever à une telle radicalisation de la problématique des limitations. En effet, si on laisse tomber toute présupposition quant à l'existence d'une vérité en soi absolue, sur quoi peut-on encore baser la vérité de la logique et, puisque la logique formule les conditions nécessaires de tout savoir rationnel possible, sur quoi peut-on baser la vérité des sciences en général ? De nouveau, l'aveuglement par la présupposition de la vérité absolue occulte le fait que la vérité des sciences est une vérité qui ne pourra jamais donner accès au sens d'être authentique de l'objet. Comme le dit Husserl : "On doit finalement cesser de se laisser aveugler par les idées idéales et régulatrices et par les méthodes des sciences 'exactes', et en particulier dans la philosophie et dans la logique, comme si l'en-soi de celles-ci était réellement une norme absolue aussi bien en ce qui concerne l'être de l'objet qu'en ce qui concerne la vérité"⁷³. Le sens d'être authentique de l'objet ne se dévoile que dans la méditation sur ce sens, à même l'*expérience* de l'objet dans la vie de la conscience. L'explicitation originelle du sens est une explicitation sur base de la donation de "*la chose elle-même*, dans son être-présent qui lui est propre"⁷⁴.

L'horizon de l'expérience auquel on est reconduit à partir du questionnement sur l'origine du sens et du droit de l'objectivité logique est donc à interroger dans la vie même de l'expérience, en dehors de tout préjugé, et c'est au sein de cette vie que l'on pourra étudier l'assomption critique, à travers les effectuations intentionnelles, des visées originaires d'identification, d'unité et d'adéquation. Cet horizon du *logos primitif*, c'est, comme le précise Husserl dans *Expérience et jugement*, l'horizon de "l'expérience la plus simple et la plus immédiate"⁷⁵, l'horizon de l'expérience perceptive des corps simplement sensibles. En effet, si l'on veut s'interroger sur l'origine constitutive du sens de la logique formelle, en tant que celle-ci

71 *Ibid.*, p. 363 ; Hua XVII, p. 280.

72 *Ibid.*, p. 364 ; Hua XVII, p. 281.

73 *Ibid.*, p. 369 ; Hua XVII, p. 284.

74 *Ibid.*, p. 373 ; Hua XVII, p. 287.

75 *Expérience et jugement*, p. 75 (*Erfahrung und Urteil*, p. 66).

se rapporte aux formes nécessaires de toute connaissance en général, on doit s'intéresser aux formations de sens dans la couche de l'expérience qui précède tous les autres, l'expérience "des substrats sensibles, de la couche naturelle du monde concret total"⁷⁶. Dans le cadre du projet d'une généalogie de la logique un privilège sera donc accordé à l'expérience perceptive des corps sensibles, c'est-à-dire à l'activité purement contemplative de la forme et de la qualité sensible des corps. C'est finalement dans l'autoconscience de l'expérience perceptive comme activité contemplative orientée vers la donation des choses dans leur être même (*Selbstgebung*) que l'on trouve la source du sens et du droit positionnel de l'entreprise de la logique formelle. Et il s'agira ensuite de montrer, comme le conclut Husserl, "comment, en s'élevant progressivement à partir de l'expérience sensible, on peut rendre compréhensible l'évidence en tant qu'effectuation et ce que signifie la vérité existant en soi en tant qu'effectuation de l'evidence"⁷⁷.

4.2. La structure d'éveil de la conscience comme forme de vie originnaire

A ce niveau de notre interrogation sur la constitution réflexive de l'horizon de sens qui motive l'entreprise de formalisation, nous pouvons déjà constater un premier déplacement par rapport à la problématique des limitations : d'une visée intentionnelle comme production active de sens dans la conscience vers une visée intentionnelle comme horizon de sens, constitué à partir de l'unité réflexive des effectuations intentionnelles de l'expérience. Ce déplacement permet d'interroger la structure intentionnelle de façon indépendante de la problématique des limitations, en formulant le projet de l'élucidation du sens tel qu'il se donne à même l'expérience dans son être même. Le modèle phénoménologique permet de cette façon de frayer un accès réflexif à l'interrogation de la dynamique de l'entreprise de la formalisation, dans ces opérations d'élargissement, de modification, d'extension et, finalement, de stabilisation en rapport à un usage contextuel déterminé, au lieu de présupposer un horizon transcendantal non interrogé qui fournirait les critères de cette dynamique. En effet, à partir du lien génétique qui existe entre l'évidence logique et l'évidence de l'expérience naturelle du monde, la démarche phénoménologique a montré que ces opérations de modalisation de l'incomplétude de la visée intentionnelle propre à la logique formelle sont déjà présentes au niveau de l'expérience perceptive du monde à titre de disposition cognitive à la stabilisation, une ouverture au remplissement possible de la visée par le monde. C'est dans ce

⁷⁶ *Ibid.*

⁷⁷ *Logique formelle et logique transcendantale*, p. 375 ; Hua XVII, p. 288.

sens que nous pouvons parler d'un traitement réflexif de la contextualisation de l'opération formelle, en définissant l'équilibre contextuel à partir de la disposition cognitive à la stabilisation possible de l'opération formelle et des usages contextuels déterminés.

Toutefois, comme nous voulons le montrer dans cette dernière section, les travaux des années vingt sur les synthèses passives permettent encore de radicaliser ce déplacement, par l'interrogation de la prédonation passive de sens à laquelle sont toujours liées les effectuations intentionnelles. Le modèle phénoménologique, tel que nous l'avons étudié jusqu'ici, propose une issue à la question des limitations, mais ne construit pas pour elle-même l'origine de son propre mode d'opérativité. C'est le rapport entre l'unité réflexive explicite et l'unité latente, pré-constituée passivement, tel que la phénoménologie tente de l'éclairer à partir des travaux sur les synthèses passives qui entre en jeu ici⁷⁸. Dans la section qui suit, nous tenterons de reconstruire les éléments clefs de ce rapport à partir des textes husserliens, tout en réservant une interprétation globale du rapport entre passivité et activité pour la dernière section.

4.2.1. Les synthèses passives dans le flux de la conscience

Le problème des synthèses passives apparaît dans *Logique formelle et logique transcendantale* en lien avec le problème de la constitution d'acquis persistants de connaissance dans le flux de la conscience. Sans de tels acquis persistants, dit Husserl, aucune science ne serait possible⁷⁹. En effet, la réunion de plusieurs objets de connaissance dans le jugement, ainsi que la réalisation de processus progressifs de jugement suppose un champ de conscience dans lequel ces objets et ces jugements peuvent apparaître de façon unitaire. De même, nous utilisons de façon permanente en sciences des propositions qui ont été produites au préalable dans des activités originaires et que nous considérons comme un fonds d'acquis durables de connaissance.

Pour interroger la constitution d'acquis durables dans le flux, il faut passer d'une analyse statique du sens de l'objet à une analyse génétique de ce sens dans le flux de la conscience. Comme l'écrit Husserl, pour caractériser cette modification de la méthode : "Tandis que l'analyse 'statique' est conduite par l'unité de l'objet intentionné [...], l'analyse intentionnelle génétique est dirigée vers l'ensemble concret tout entier dans lequel se

78 Cf. KÜHN R., *Husserls Begriff der Passivität. Zur Kritik der passiven Synthesis in der genetischen Phänomenologie*, Freiburg/München, 1998.

79 Cf. *Logique formelle et logique transcendantale*, p. 422 ; Hua XVII, p. 326.

situent toute conscience et son objet intentionnel en tant que tel⁸⁰. En interrogeant l'apparition de l'horizon du sens dans l'ensemble concret de la conscience en situation Husserl ouvre un nouvel ensemble de questions :

“Viennent alors aussitôt en question les autres renvois intentionnels qui appartiennent à la *situation* dans laquelle par exemple se trouve celui qui exerce l'activité de jugement ; vient donc aussitôt en question en même temps l'*unité immanente de la temporalité* de la vie qui a dans cette temporalité son *histoire* de sorte que dans cette unité temporelle tout vécu de conscience particulier a, en tant qu'il intervient temporellement, sa propre 'histoire', c'est-à-dire sa *genèse temporelle*”⁸¹.

D'un questionnement abstrait sur l'horizon comme explicitation intentionnelle de sens, nous passons, par l'analyse génétique, à un questionnement sur l'émergence de ce sens dans la conscience, dans l'*histoire* et la *situation* concrète de cette conscience. Dans le cadre de l'analyse génétique, les renvois à des donations de sens plus originaires deviennent alors des renvois temporels et des renvois aperceptifs :

“Tout mode original de donnée a un *effet génétique ultérieur double*. Premièrement sous la forme de reproductions possibles qui procurent le ressouvenir en passant par des rétentions qui s'enchaînent d'une manière génétiquement originelle et tout à fait immédiate ; et *secondement* on a affaire à l'effet 'aperceptif' conformément auquel, dans une situation nouvelle analogue, ce qui existe déjà, de quelque façon qu'il ait été constitué, devient, d'une manière analogue, objet de l'aperception”⁸².

Dans l'analyse génétique, l'horizon devient le dépôt de renvois temporels et aperceptifs. Plus précisément, Husserl distingue deux types différents de renvois temporels. D'une part, les rétentions qui se situent encore dans l'horizon du présent vivant du temps et, d'autre part, les reproductions possibles que procure le souvenir. Les premiers renvois, les renvois immédiats dans le mouvement de la rétention, donnent lieu à des acquis persistants dans le flux, qui rendent possibles des processus progressifs de jugements et la constitution de jugements composés de plusieurs membres. Les deuxièmes renvois, les reproductions dans le souvenir, conduisent à la possession durable des résultats du jugement, au-delà de la première acquisition vivante. Ces possessions durables sont désignées par Husserl par le terme de possessions habituelles ou d'*habitus*⁸³.

Le passage d'une conception statique de l'horizon à une conception génétique introduit une modification de la structure intentionnelle des activités du jugement. L'analyse statique de l'horizon de sens construisait l'horizon comme l'unité synthétique d'effectuations originaires constitutives.

80 *Ibid.*, p. 410 ; Hua XVII, p. 316.

81 *Ibid.*

82 *Ibid.*, p. 411 ; Hua XVII, p. 317.

83 Cf. *ibid.*, p. 412 ; Hua XVII, p. 318.

Maintenant, en interrogeant l'émergence de l'horizon dans le flux même de la conscience, on doit rendre compte du maintien durable de ces effectuations originaires, de façon rétionnelle ou habituelle, au cours de nouvelles activités du sujet. La constitution du jugement sera donc une constitution graduelle, partiellement effectuation originaire et partiellement présence passive d'effectuations originaires sédimentées.

4.2.2. La vie originaire comme vie en éveil dans le présent vivant du temps

La constitution de l'évidence du jugement s'appuie sur la possibilité de maintenir l'objet du jugement, dans son "soi-même", comme identique et identifiable dans le flux de la conscience. Comme nous l'avons vu, cette synthèse d'identification passive qui accompagne toute explicitation active est rendue possible de manière génétiquement originelle par les renvois associatifs dans le mouvement de la rétion. La phénoménologie génétique apparaît donc comme un approfondissement de la phénoménologie de la conscience intime du temps⁸⁴. Par le questionnement sur le sens nous sommes reconduits à l'autoconscience (*Selbstbesinnung*) de la vie de la conscience dans sa temporalité immanente.

C'est dans les cours sur *La logique transcendantale* des années vingt, édités dans les *Husserliana XI*⁸⁵ et traduits récemment dans l'ouvrage *De la synthèse passive*⁸⁶, que Husserl développe de manière plus détaillée cet approfondissement de la doctrine de la conscience intime du temps. Plus précisément vient en question la constitution même du mouvement de la rétion et de la protention, en lien avec le champ de la prédonation passive. Comme l'écrit Husserl, l'horizon rétionnel et l'horizon protentionnel du présent vivant restent des horizons vides sans la présence de contenus différenciés dans le flux et l'éveil affectif du sujet par de tels contenus :

"Il va de soi qu'est présupposée la synthèse qui s'effectue continuellement dans la conscience originaire du temps. [...] Mais cette manière dont la subjectivité devient conscience de sa vie passée et future avec les teneurs intentionnelles qui s'y trouvent est incomplète. S'il n'y avait pas d'éveil, elle serait pour le moi une modalité sans signification, car les rétions sont vides et sombrent même dans l'arrière-plan

84 Cf. également HOLENSTEIN E., *Phänomenologie der Assoziation. Zu Struktur und Funktion eines Grundprinzipes der passiven Genesis bei E. Husserl*, Den Haag, 1972, pp. 68–73.

85 HUSSERL E., *Husserliana XI : Analysen zur passiven Synthesis. Aus Vorlesungs- und Forschungs-manuskripten 1918–1924*, édité par Margot Fleischer, Den Haag, 1966. Les cours qui sont édités dans ce volume ont été prononcés dans le semestre d'hiver de 1920–1921, le semestre d'été de 1923 et le semestre d'hiver de 1925–1926. Le premier cours avait comme intitulé *Logique transcendantale* (Cf. Hua XI, Introduction p. XIV).

86 HUSSERL E., *De la synthèse passive*, op. cit.

rétentionnel indifférencié. A plus forte raison le futur protentionnel est une conscience vide⁸⁷.

C'est seulement la présence d'un champ impressionnel différencié qui permet l'éveil du sujet et la conscience rétentionnelle et protentionnelle de sa vie passée et future. Tout d'abord, la structuration interne du champ impressionnel se fait par des fusions originaires dans le champ impressionnel d'après l'homogénéité ou l'hétérogénéité des *data*. Ces fusions originaires donnent lieu à des configurations sensibles de *data*, par exemple le détachement d'une tache rouge sur un fond blanc. Ensuite, la constitution du champ originaire comme système d'objets situés dans le temps s'appuie sur des recouvrements entre les horizons temporels de différents *data* qui apparaissent dans ce champ. Par ces recouvrements "il se constitue *un* maintenant, qui confère au maintenant de l'un et de l'autre datum l'unité d'identité"⁸⁸. Par ces synthèses de recouvrement s'opère une unification du présent et du non-présent du champ entier et se produit "un seul temps dans lequel s'écoulent tous les écoulements temporels des objets"⁸⁹. De la même manière que l'individuation des *data* du champ se produit par rapport au temps, une individuation se produit par rapport à l'espace, de sorte que le champ "possède, de manière analogue au temps, la forme d'un système de places continuellement ordonné"⁹⁰.

Les fusions originaires et les synthèses temporelles et spatiales sont des conditions de possibilité encore formelles de la constitution d'un champ individué et unifié. Ce qui conditionnera la formation effective d'un champ structuré, c'est la force affective du champ, la stimulation affective du je par les impressions originaires. Cette stimulation affective a évidemment ses degrés et c'est pourquoi Husserl distingue deux phénomènes sous le titre de l'affection. D'abord l'affection comme "l'attrait spécifique qu'un objet exerce sur le moi"⁹¹, "la vitalité changeante d'un vécu"⁹² et ensuite l'affection comme "l'éveil d'une intention dirigée vers lui [*l'objet*]"⁹³. Certains *data* du champ sont particulièrement saillants pour le sujet et c'est précisément ces *data* qui éveillent son intérêt et vers lesquels il peut tourner son attention dans une saisie explicite d'objet.

Le cas limite de l'horizon affectif — et ce sera également la limite ultime de la théorie phénoménologique de la conscience —, c'est l'arrière-

87 *Ibid.*, p. 197 ; Hua XI, p. 125.

88 *Ibid.*, p. 199 ; Hua XI, p. 127.

89 *Ibid.*

90 *Ibid.*, p. 213 ; Hua XI, p. 143.

91 *Ibid.*, p. 217 ; Hua XI, p. 148.

92 *Ibid.*, p. 231 ; Hua XI, p. 166.

93 *Ibid.*, p. 219 ; Hua XI, p. 151.

fond [*Hintergrund*] de données sans vitalité et sans force affective pour le je. Husserl désigne cet arrière-fond par le terme d'horizon nul : "Aussi devons-nous ranger dans le présent vivant un horizon affectif nul qui constamment se modifie"⁹⁴. Cet horizon affectif nul est constitué de l'arrière-plan de données rétentionnelles qui disparaissent continuellement du champ impressionnel, et dont le sens — qui ne disparaît pas pour autant — devient de moins en moins affectif jusqu'à perdre toute vitalité. Ce sens reste présent de façon implicite à la conscience, par les "lignes particulières d'unité de différentes sortes"⁹⁵ qui le lient au présent de conscience. Toutefois par la sédimentation dans l'arrière-fond de non-vitalité ces données deviennent des données indifférenciées. Elles se sédimentent dans "l'horizon du passé entièrement inarticulé, entièrement indistinct, lequel est le terme du passé rétentionnel vivant, mouvant"⁹⁶.

Le questionnement sur le sens nous reconduit à l'autoconscience de la temporalité immanente de la vie en lien avec la différenciation originaire du champ impressionnel du présent vivant, avec sa vitalité affective mouvante. Ce que nous montre donc la réflexion phénoménologique, c'est que toute production de sens est explicitation originaire de sens à même l'expérience et que cette explicitation originaire s'appuie sur l'épreuve passive de la donation originaire du "soi" de l'objet dans le présent vivant de la conscience.

4.3. L'habitus pré-réflexif et la radicalisation de la méthode phénoménologique

La phénoménologie génétique des années vingt permet d'éviter l'aporie d'une activité d'explicitation de sens qui n'élucide pas la question de l'origine de sa propre opérativité. En effet, comme nous l'avons vu, ces travaux montrent que l'activité réflexive d'explicitation dépend à son tour d'une structure d'éveil de la conscience comme forme de vie originaire. Pour saisir la portée de ce déplacement pour la problématique de la théorie de la connaissance, il faut encore tenter de préciser la nature de la relation entre la réflexivité et la structure de l'éveil. Pour cela, nous procéderons en deux temps. Tout d'abord, nous tenterons d'approfondir l'ouverture vers les conditions d'opérationnalisation de la réflexivité dans la phénoménologie génétique à partir du développement de cette même problématique chez M. Henry. Ensuite, nous reprendrons de façon synthétique l'ensemble du

94 *Ibid.*, p. 232 ; Hua XI, p. 167.

95 *Ibid.*, p. 235 ; Hua XI, p. 171.

96 *Ibid.*

parcours effectué dans ce paragraphe à partir du fil conducteur de la problématique de la théorie de la connaissance.

4.3.1. L'ouverture à la sphère de déterminabilité de la conscience

Afin de saisir la portée de l'ouverture vers les conditions d'opérationnalisation pour la théorie de la connaissance, il faut retourner à la problématique de l'évidence. Comme nous l'avons vu, la réduction phénoménologique transcendantale permet une compréhension renouvelée de l'évidence comme *effectuation* et par là du sens comme *horizon eidétique* perceptible et explicitable à même l'expérience. Toutefois, le primat donné par la méthode phénoménologique à la perception et à l'évidence comme mode de donation originaire la conduit à identifier sa méthode (la pensée, l'évidence, l'intentionnalité) et son objet (également la pensée, la *perceptio*). La phénoménologie, en se basant sur l'évidence originaire dans la vie perceptive, ne permet pas d'élucider la question de son propre mode d'opérativité. En fait, l'accès frayé par la phénoménologie à l'eidétique de la conscience contribué en même temps à occulter la vie réelle et concrète de la pensée. Celle-ci, tout en se dérochant à la réflexivité phénoménologique, constitue néanmoins son point de départ. C'est précisément cette relation occultée entre la réflexion et la vie concrète qui est à l'origine des travaux sur les synthèses passives. Cette réflexion sur l'origine et le mode d'opérativité propre de la réflexivité dans la phénoménologie rejoint celle de M. Henry dans *Incarnation. Une philosophie de la chair*⁹⁷. D'après M. Henry, le déplacement vers les conditions opératoires de la réflexivité conduit à un *renversement* de la phénoménologie : du primat accordé à la perception par rapport à la vie concrète vers le primat de l'auto-affection de la vie qui rend possible le "voir" impliqué dans la perception. Le thème du *renversement* permet de préciser la portée de la réflexion sur les synthèses passives.

D'après M. Henry, le primat accordé par Husserl à la perception le conduit à une mésinterprétation du *cogito* cartésien⁹⁸. D'une part, comme nous l'avons vu ci-dessus, pour Husserl le mérite du geste cartésien est de remettre au centre de la réflexion philosophique la problématique de l'évidence. En même temps, Husserl reproche à Descartes de poser une nouvelle certitude indubitable, un nouvel *en-soi*, l'être *en soi* de la pensée ou du *cogito* au lieu de s'en tenir au projet d'une explicitation intentionnelle de toute évidence, quelle qu'elle soit, comme ce sera le cas dans la phénoménologie transcendantale. Toutefois, comme le montre M. Henry,

97 HENRY M., *Incarnation. Une philosophie de la chair*, Paris, 2000.

98 Cf. *ibid.*, p. 103.

cette interprétation de Husserl passe outre un autre trait important de la philosophie cartésienne, la méthodologie du doute, qui amène à la disqualification de l'évidence comprise comme auto-fondation du fondement⁹⁹. La critique, certes justifiée, qu'adresse Husserl à la position d'un *cogito comme en-soi* n'atteint donc pas la position cartésienne en tant que telle puisque celle-ci vise une position au-delà de l'évidence, un *cogito* qui n'est justement pas à comprendre comme un nouvel *en-soi*, mais comme la découverte de la vie qui marque de son trait décisif toute connaissance et donc, en particulier, la connaissance phénoménologique. En reprenant les termes de M. Henry, "la pensée est possible en tant que mode de la vie, parce que la certitude de son voir est en réalité celle de la vie"¹⁰⁰.

Cette réinterprétation de la problématique de l'évidence au profit d'un primat de l'auto-affection de la vie amène également à envisager autrement le mode d'opérativité propre des synthèses passives qui fondent l'habitualité de la conscience. Comme le montre bien M. Henry, ce n'est pas tellement la temporalité qui constitue ce qu'il y a de synthétique dans la passivité, mais l'épreuve de soi par un soi singulier¹⁰¹. L'accès à la vie par la rétention mène, comme nous l'avons vu dans la section précédente, à un horizon nul, sans affectivité vitale, c'est-à-dire au lieu vide de l'absence de toute vie. Toutefois, "nous savons, écrit M. Henry, que ce n'est jamais dans un voir que nous parviendrons à la vie, mais seulement là où elle parvient en soi — là où, depuis toujours, nous sommes déjà parvenus en nous-mêmes : dans la Vie absolue"¹⁰². L'épreuve de soi se fait en rapport à l'auto-affection de "ce corrélat 'noématique' qu'est l'essence' de la vie, une essence 'transcendante', offerte au regard de l'intuition eidétique"¹⁰³. Le moment synthétique de la passivité ne consiste donc pas dans une temporalité toujours évanescence en rapport à l'écoulement des contenus de conscience, mais dans la synthèse propre à la vie invisible qui habite notre voir. L'accès à celle-ci est frayée par l'image spéculative de la vie absolue¹⁰⁴. Cette image spéculative nous révèle la structure de la vie en éveil comme auto-affection qui conditionne toute réflexivité et qui maintient la possibilité de l'effectuation par-delà toute effectuation temporelle particulière¹⁰⁵.

99 Cf. *ibid.*, pp. 103–104.

100 *Ibid.*, p. 130. Cf. également KÜHN R., *Leben als Bedürfen. Eine lebensphänomenologische Analyse zu Kultur und Wirtschaft*, Heidelberg, 1996.

101 HENRY M., *op. cit.*, pp. 105–107.

102 *Ibid.*, p. 131.

103 *Ibid.*

104 Cf. *ibid.*

105 Pour une analyse de la potentialité de la conscience absolue comme vie en éveil, cf. également KOKOSZKA V., *Les conditions transcendantales du devenir commun.*

En reconstruisant l'accès à la structure habituelle de la vie en éveil par le chemin de la saisie spéculative de l'auto-affectation de la vie absolue, cette interprétation de la radicalisation de la méthode phénoménologique peut également être rapprochée des travaux sur la conscience absolue dans l'idéalisme allemand. Comme le montre M. Maesschalck dans un article sur les philosophies intermédiaires de Fichte et de Schelling¹⁰⁶, qu'il s'agisse de la problématique du *primum passivum* chez Schelling¹⁰⁷ ou de la problématique du *Moi vivant* chez Fichte¹⁰⁸, l'enjeu des recherches sur la conscience dans ces philosophies est bien un dépassement du transcendantalisme qui n'explicite pas l'origine du mode d'opérativité de la conscience de soi. Ce rapprochement entre la phénoménologie génétique et l'idéalisme allemand permet de mettre en évidence, comme dans l'interprétation de M. Henry, la dimension structurelle de la conscience absolue comme activité pure, dont Husserl tente de rendre compte à partir de la métaphore de l'éveil. En effet, comme l'écrit M. Maesschalck, "Cette métaphore a l'avantage d'introduire à l'activité pure de la conscience absolue sans supposer la conscience de soi qu'elle sous-tend précisément"¹⁰⁹. De cette façon, le pouvoir d'éveil apparaît "comme la métaphore de l'activité pure qui sous-tend la discontinuité des états de veille et de sommeil de la conscience de soi. La discontinuité des états d'une conscience tantôt éveillée, tantôt assoupie, suppose en effet la possibilité continue tant d'entrer en éveil que d'en sortir"¹¹⁰. De plus, cette persistance latente de l'unité de la vie de la conscience, sous le mode de la possibilité de devenir une conscience éveillée, n'est pas quelque chose de surajouté à l'activité de la conscience de soi, mais est considérée par Husserl comme étant l'*habitus pré-réflexif* de cette conscience. Plus précisément, en tant que structure originaire, cette disposition habituelle est la condition de possibilité de la conscience de soi dans un monde. Pour désigner le mode d'opérativité spécifique de cette structure, M. Maesschalck parle de l'ouverture de la conscience absolue à sa sphère de déterminabilité. "Pour Husserl, en effet, écrit M. Maesschalck, l'eidétique de la conscience absolue n'est pas condition de possibilité du monde, mais en tant que structure originaire du flux de la 'vie en éveil', elle est condition de possibilité de la conscience de

Ontologie et monadologie chez Husserl, Diss., Louvain-la-Neuve, 2001, pp. 58–59.

106 MAESSCHALCK M., "La conscience absolue comme éveil à la déterminabilité du monde", in GODDARD J.-C. (ed.), *Le transcendantal et le spéculatif dans l'idéalisme allemand*, Paris, 1999, pp. 77–96.

107 Cf. *ibid.*, p. 82.

108 Cf. *ibid.*, p. 94.

109 *Ibid.*, p. 81.

110 *Ibid.*

soi dans un monde. Dans son essence, la conscience apparaît donc déjà ouverte à une déterminabilité, même si elle est libre de toute détermination”¹¹¹.

Le rapprochement avec l'idéalisme allemand permet donc de déployer sous d'autres modalités le terme occulté par la problématique phénoménologique de l'évidence, le corrélat noématique de la conscience absolue, traité par Husserl à partir de la problématique des synthèses passives et par M. Henry à travers le problème de l'auto-affection. Par cette mise en perspective, nous pouvons mieux comprendre également l'articulation proposée entre la réflexivité et la structure d'éveil dans la phénoménologie génétique : la structure d'éveil comme forme de l'activité pure de la conscience absolue constitue, en tant que *structure ouverte*, une disposition habituelle à la constitution réflexive de la conscience de soi en rapport à un monde. La radicalisation de la phénoménologie n'implique donc pas que l'un des deux termes devienne indépendant de l'autre, mais montre l'interdépendance entre le pôle de l'activité réflexive d'une part et le pôle opératoire de la vie en éveil d'autre part. En effet, l'activité réflexive dépend de la structure habituelle de la conscience, préconstituée passivement à travers les expériences passées, mais, en même temps, la structure ouverte de la vie en éveil, en tant que structure latente, dépend de la vie réflexive pour pouvoir se constituer en conscience de soi effectuant en rapport à un monde¹¹².

111 *Ibid.*, p. 80. Cf. également pour un développement de cette problématique à partir de la problématique de la temporalité, KOKOSZKA V., “La conception husserlienne de la temporalité entre 1905 et 1910”, in *Tijdschrift voor Filosofie*, 58 (1996), pp. 314–341.

112 Il faudrait ici poursuivre la réflexion sur les conditions d'opérationnalisation de la réflexivité, dont on n'a étudié que l'aspect d'éveil et de sédimentation habituelle, en montrant comment aux différents niveaux de l'activité réflexive on peut faire correspondre différents niveaux dans l'habitus pré-réflexif. On peut trouver une tentative dans ce sens dans l'article de A. Schütz sur le rôle de la typique et de l'eidétique dans la philosophie tardive de Husserl (cf. SCHÜTZ A., “Typus und Eidos in Husserls Spätphilosophie”, in *Alfred Schütz. Gesammelte Aufsätze III. Studien zur phänomenologische Philosophie*, Den Haag, 1971, pp. 127–152). Dans cet article, A. Schütz montre comment les différents niveaux de généralité typique de notre expérience du monde se déploient dans un espace de jeu (*Spielraum*) habituel, dont les strates opératoires correspondent à la pré-constitution passive des différents niveaux de généralité du type. Les niveaux pris en considération par Schütz vont du type empirique, en passant par les idées présomptives concernant les objets en général jusqu'à la typique du monde de la vie rendant possible les idées présomptives.

4.3.2. Une théorie radicale de la connaissance

En guise de conclusion du parcours que nous avons effectué dans ce dernier paragraphe, nous voulons reprendre les différentes étapes de ce parcours à partir de deux axes de discussion. En premier lieu, nous voulons montrer la portée de l'analyse phénoménologique du formalisme par rapport à une critique du projet de la naturalisation cognitive de la conscience intentionnelle tel que nous l'avons présenté dans l'introduction. En un deuxième temps, nous situerons l'analyse phénoménologique par rapport à notre propre réflexion épistémologique sur les limites de la fécondité sémantique des formalismes.

La première étape de la réflexion phénoménologique sur le sens, la mise en évidence du caractère eidétique ou idéalisant du sens de la logique formelle, permet de montrer les limites de la naturalisation, sous la forme du projet de la formalisation mathématique de la conscience intentionnelle. A la lumière de l'analyse phénoménologique, le projet de la formalisation apparaît lui-même comme étant le résultat d'une visée intentionnelle. De plus, l'horizon de sens de ce projet ne peut pas se justifier uniquement à partir des effectuations intentionnelles actuelles, mais vise un sens qui appartient aux productions de l'activité scientifique au-delà de leur donnée actuelle. La vie intentionnelle de la conscience apparaît comme ce qui donne sens aux normes effectives de la pratique de la formalisation dans son ensemble et on ne pourra pas en rendre compte en ayant de nouveau recours à une démarche du même type. Le projet de la naturalisation cognitive de la conscience intentionnelle rencontre une limite phénoménologique, qui porte sur le sens intentionnel de l'entreprise de la naturalisation elle-même.

Toutefois, la vie intentionnelle n'est pas uniquement une *limitation* subjective de certaines attitudes objectivistes, comme celle des sciences, mais elle est affirmation originaire d'une appartenance subjective à un monde. Ce qui est en question dans la réflexion phénoménologique, ce n'est donc pas tant l'attitude naturaliste, qui est celle des sciences, mais l'attitude naturelle. Celle-ci caractérise notre confiance originaire dans l'évidence du monde factuel, envisagé comme monde de vérités existant en soi, et constitue le sol sur lequel s'appuie toute attitude objectiviste. Une interrogation sur le sens qui en reste à l'attitude naturelle ne peut donc pas dépasser les apories liées à la problématique de l'évidence et évacue par là tout le domaine de réflexion auquel nous donne accès la phénoménologie transcendantale.

La mise en œuvre de la réduction phénoménologique transcendantale ouvre la voie d'une recherche phénoménologique sur le sens et sur l'évidence en tant que tels, en dehors de tout rapport à une facticité

déterminée. Par la réduction phénoménologique transcendantale, nous dépassons donc la problématique de la limitation de la fécondité sémantique des formalismes, qui portait encore sur un problème de sens et d'évidence régional — lié à la modélisation mathématique —, pour nous interroger sur la constitution subjective du sens, en dehors de tout préjugé sur l'existence d'un domaine de vérités *en soi*.

Pour distinguer cette recherche phénoménologique transcendantale d'une recherche régionale, qui se rapporte à un domaine de connaissance particulier, nous pouvons parler d'une théorie radicale de la connaissance. La radicalité du projet phénoménologique se réfère d'abord au radicalisme d'une recherche normative ultime. Dans *Logique formelle et logique transcendantale*, elle prend la forme d'une interrogation sur la légitimité de principe du projet de la connaissance scientifique, telle qu'on peut la rencontrer, comme l'écrit Husserl dans l'introduction, "dans le chemin audacieux des *Méditations métaphysiques*"¹¹³. Ensuite, comme nous l'avons vu, cette recherche ne peut présupposer la validité du *cogito*, en tant que *cogito* "mondain", mais doit pouvoir expliciter sa constitution réflexive à même l'expérience du monde.

Nous avons analysé les conséquences de cette radicalisation de la théorie de la connaissance à deux niveaux de réflexion: premièrement au niveau d'une recherche sur le logos primitif (*Urlogos*), entendue comme une explicitation originelle des horizons de sens constitutifs de l'expérience sensible dans sa totalité, et deuxièmement au niveau d'un enracinement de ce logos primitif dans les synthèses passives de la vie de la conscience. A la différence de la réflexion épistémologique sur les sciences, la théorie radicale de la connaissance vise à retrouver un enracinement de la visée de connaissance dans la finalité immanente de la vie de la conscience, envisagée dans sa possibilité pure. Par là, le projet d'une théorie radicale de la connaissance, tel qu'il est formulé dans *Logique formelle et logique transcendantale*, permet de rétablir le lien entre la visée de la connaissance scientifique dans sa facticité contingente d'un côté et le projet philosophique d'une réflexion critique sur la justification ultime de cette visée de l'autre.

Un tel lien peut être conçu pour lui-même en tant qu'expérience de la "dynamique de la raison" au sens donné à ce terme par Jean Ladrière¹¹⁴. Mais il peut être analysé encore sur le plan même de sa structure spécifique comme manifestation de l'événementialité de la raison. Sur ce plan, plus spécifiquement spéculatif, c'est l'opérativité même de la raison qui est resaisie dans son *pouvoir réflexif* consistant à produire sa visée de

113 *Logique formelle et logique transcendantale*, p. 10 ; Hua XVII, p. 11.

114 Cf. LADRIÈRE J., "L'éthique et la dynamique de la raison", in *Rue Descartes*, 7 (1993), pp. 47–69.

connaissance *de manière* à aller à la rencontre de sa propre fin à travers elle. Ce qu'on pourrait nommer aussi "téléologie interne de la raison" ou "entéléchie" comme Aristote, voir "eschatologie" comme Ladière¹¹⁵ est un processus inférentiel de réflexivité par lequel l'auto-affection de la vie prend les modalités opératoires d'une conscience intentionnelle mobilisée par une destination, convoquée au-delà d'elle-même par la "possibilisation" des formes de vie.

115 Cf. *ibid.*, p. 61.

Conclusion

En guise de conclusion, nous voulons évaluer la contribution de notre parcours, à la fois épistémologique et empirique, à la problématique de l'intentionnalité. A cet effet, nous essayerons de synthétiser la configuration nouvelle qui émerge en combinant les deux axes de notre parcours : un premier axe qui nous a conduit à la définition d'une théorie écologique de la cognition dans la première partie de cet ouvrage et un deuxième qui veut écarter une réduction formaliste de cette théorie écologique, en montrant que la stabilisation de cette entreprise de modélisation dépend d'un équilibre particulier avec son contexte d'application dans la deuxième partie. Le but de cet exercice de synthèse est non seulement d'évaluer si oui ou non les hypothèses de départ ont été vérifiées, mais également d'entrevoir des prolongements possibles de ces hypothèses à partir de la configuration d'ensemble.

1. Le premier axe : l'opération de modélisation de l'intentionnalité dans les sciences cognitives

Dans la première partie de ce travail, nous sommes partis du défi naturaliste formulé par H. Atlan dans "Projet et signification", afin d'examiner la fécondité d'une prise en compte du rôle de l'action et du contexte à l'intérieur de l'opération de modélisation de l'intentionnalité en sciences cognitives. Loin d'être absente des débats en sciences cognitives, nous avons vu que cette problématique est au contraire au centre de controverses qui portent sur le dépassement d'une approche purement computationnelle ou opératoire de la cognition. Ce dépassement concerne d'abord l'élargissement des modèles de l'intentionnalité, mais a aussi des conséquences sur la méthodologie de la recherche en sciences cognitives.

Premièrement, sur le plan de l'élargissement des modèles, se pose le problème de la définition d'une opération de modélisation qui tienne compte à la fois de la dimension de la création des significations (le contenu intentionnel) et de la dimension de la satisfaction d'une fonction (la visée intentionnelle). Le problème crucial qui est soulevé dans ce contexte est, comme l'indique également Atlan dans "Projet et signification", celui de l'autonomie du champ des significations par rapport aux fonctions. Le système intentionnel définit lui-même les contenus sur lesquels il opère et

l'évolution du système est autant déterminée par la dynamique autonome des contenus que par les fonctions. Nous avons rendu compte de cette tension réversible et asymétrique entre les significations et les fonctions dans le premier chapitre à partir des concepts d'écologie subjective et de détermination écologique.

Ensuite, nous avons étudié, dans le deuxième chapitre, les conséquences de la prise en compte de la dimension écologique dans la modélisation sur la méthodologie de recherche en sciences cognitives. Cette analyse nous a permis de distinguer différentes formes de naturalisation. En effet, la naturalisation de l'intentionnalité n'implique pas nécessairement un physicalisme ou un évolutionnisme, mais donne lieu à une pratique de modélisation spécifique, que l'on a désignée par le terme de modélisation contextuelle. Tout d'abord, l'approche computationnaliste classique, que nous avons étudiée à partir de la méthodologie de recherche de David Marr, s'inscrit encore dans un paradigme profondément physicaliste. En effet, dans l'approche classique de Marr, ni la définition de la fonction globale à réaliser par le système, ni la définition des données sur lesquelles il opère ne sont prises en compte en tant que telles dans la modélisation. Elles sont simplement présentes à titre de conditions frontières de la modélisation. Par conséquent, la modélisation se limitera à l'étude des fonctions algébriques de transformation "entrée - sortie" d'un système computationnel. Pour modéliser l'intentionnalité, il faut cependant également tenir compte des contextes représentationnel et fonctionnel qui définissent l'environnement opératoire du système. Afin de prendre en compte cette exigence, nous avons établi, sur le plan méthodologique, une double progression par rapport à l'approche computationnelle classique. La première progression concerne le passage d'un paradigme physicaliste, qui étudie des mécanismes de transformation de données indépendamment du contexte, à un paradigme d'inspiration biologique, qui articule l'étude des mécanismes avec l'étude du caractère téléonomique des fonctions comme fonctions *biologiques*, en lien avec les ressources de stabilisation dans le contexte. La deuxième progression construit le passage d'un paradigme biologique à un paradigme "intentionnel", qui prend également en compte l'anticipation du contexte d'effectuation des opérations cognitives à l'intérieur des modèles. La prise en compte de cette double progression par rapport au fonctionnalisme permet alors d'entrevoir une pratique de modélisation qui croise l'anticipation contextuelle de significations avec l'inscription de cette anticipation dans la dynamique opératoire autonome du contexte.

2. Le deuxième axe : les limitations contextuelles de l'opération de modélisation

Après la définition d'une opération de modélisation spécifique de l'intentionnalité dans les sciences cognitives, nous avons abordé, dans le troisième chapitre de ce travail, la question des limitations de cette opération de modélisation. Ce deuxième temps de notre réflexion nous a amené à réaliser un déplacement par rapport à la position de Atlan face au problème des limitations, telle que nous l'avons exposée à partir de "Projet et signification".

Tout d'abord, dans "Projet et signification", Atlan situe la réduction naturaliste dans le même cadre méthodologique que les sciences de la nature. Toutefois, par là, il n'interroge pas le pouvoir spécifique qu'ont les modèles formels élaborés au sein de l'attitude naturelle à investir un domaine spécifique de modélisation, comme le domaine des opérations intentionnelles, dans sa différence spécifique avec le domaine de la biologie ou de la physique. Un même formalisme peut être interprété de différentes façons et, pour obtenir un modèle spécifique de l'intentionnalité, il faut encore définir le domaine des modèles particuliers auquel le formalisme s'applique. Il faut donc opérer un déplacement de la question des limitations par rapport à la limitation naturaliste mise en évidence par Atlan, un déplacement des modèles formels au champ concret des modèles particuliers. C'est uniquement à partir de ce déplacement que l'on peut mettre en évidence la limitation spécifique de la modélisation de l'intentionnalité, qui porte, comme nous l'avons vu, sur l'importance des conflits d'interprétation dans l'élaboration des modèles.

A partir de cette limitation contextuelle, nous avons montré que différentes interprétations de l'intentionnalité s'opposent dans les sciences cognitives. Ainsi, nous avons étudié les interprétations de l'intentionnalité comme guide d'action efficace, comme construction de représentations adéquates du monde et comme histoire vécue. En particulier, l'interprétation de Varela, Thompson et Rosch de l'intentionnalité comme histoire vécue permet de mesurer toute la distance de notre position par rapport à celle de Atlan. En effet, comme ce dernier, Varela, Thompson et Rosch défendent un formalisme d'autostabilisation des fonctions intentionnelles. Toutefois, à la différence de l'approche d'Atlan, ils articulent de façon explicite ce choix de formalisme à une interprétation particulière de l'intentionnalité et à un ensemble d'hypothèses auxiliaires qui permettent de justifier cette interprétation.

Finalement, c'est cette relation entre l'interprétation et les hypothèses auxiliaires, telle qu'on peut la mettre en évidence dans les contextes

particuliers de modélisation, qui permet de formuler au mieux la position à laquelle aboutit notre discussion des limitations dans le troisième chapitre de ce travail. Dans ce chapitre, pour montrer l'importance de l'interprétation dans la modélisation de l'intentionnalité, nous avons présenté des interprétations alternatives d'expériences que nous avons étudiées dans le premier chapitre. Par là, nous ne voulons aucunement remettre en question les acquis de notre analyse de l'opération de modélisation de l'intentionnalité dans les termes de la théorie écologique. Au contraire, c'est uniquement en se basant sur cet acquis que l'on peut prendre en compte le caractère contextuel de la modélisation et mettre en évidence les conflits d'interprétation. Par conséquent, si nous affirmons l'importance de critères non scientifiques dans la modélisation de l'intentionnalité, nous voulons en même temps argumenter en faveur d'une articulation de ces critères aux hypothèses auxiliaires qui sont mises en œuvre dans l'élaboration des modèles particuliers.

3. Contribution de la théorie écologique à l'hypothèse de la réversibilité asymétrique

Au départ de notre ouvrage, nous avons formulé le projet de vérifier si la prise en compte du rôle de l'action et du contexte dans les modèles contemporains de l'intentionnalité permettait de construire un prolongement empirique de l'hypothèse de la réversibilité asymétrique. Cette hypothèse stipule qu'une construction de la rationalité du domaine de l'agir intentionnel pour elle-même demande de prendre en compte la relation d'interdépendance à la fois réversible (la relation fonctionne dans les deux sens) et asymétrique (les deux éléments mis en relation gardent leur autonomie) entre l'action et son contexte. Or, la plupart des théories écologiques que nous avons étudiées ne vérifient pas cette hypothèse. En effet, la critique du fonctionnalisme inhérente à la perspective écologique s'en tient le plus souvent à mettre en évidence l'internalisme du contexte, en montrant la co-détermination entre l'acteur et le contexte, que ce soit à partir d'une perspective ontogénétique comme chez A. Berthoz ou E. Thelen ou d'une perspective phylogénétique comme chez F. Varela ou J. Gibson. De telles perspectives formalisent bien la relation d'interdépendance réversible entre l'action et le contexte : le contexte n'existe que par les opérations perceptives qui le constituent comme contexte des activités vitales d'acteurs intentionnels d'une part et l'action vitale est toujours interne à une certaine perception contextualisante de l'environnement de l'autre. Toutefois, elles ne prennent pas en compte l'asymétrie qui caractérise la relation. En effet, la théorie écologique

classique présuppose un contexte dont les dynamiques intrinsèques n'interviennent pas dans la stabilisation interne du système et reste pour cela une théorie écologique subjective. Elle ne rencontre pas l'exigence que nous avons formulée dans l'avant-propos à partir des travaux de M. Maesschalck, à savoir l'exigence de "concevoir le contexte comme un milieu qui interagit avec l'action et dont le 'comportement' doit être pris en compte par celle-ci"¹. Toutefois, comme nous l'avons vu, si l'on intègre dans la théorie écologique également la détermination écologique, alors il y a moyen de prendre en compte l'asymétrie de la relation entre l'action et le contexte. De ce point de vue, la perception contextualisante est alors modélisée comme une opération interne au devenir autonome du contexte dont les dynamiques intrinsèques jouent également un rôle dans la stabilisation.

La configuration particulière qui émerge en combinant la théorie écologique subjective avec la détermination écologique vérifie donc bien l'hypothèse de la réversibilité asymétrique. Toutefois, on peut se demander si cette reconstruction de la théorie écologique n'a pas également des effets en retour sur la limitation contextuelle de la modélisation. En effet, à partir de notre examen de la limitation contextuelle, nous avons vu que le domaine de l'action intentionnelle n'est pas un objet de référence "en général" d'une opération de formalisation, mais que sa spécification dépend d'une pratique interprétative qui définit un équilibre particulier entre la perception contextualisante du contexte d'une part et les conditions contextuelles de stabilisation de cette perception de l'autre. Or, le projet de la naturalisation de l'intentionnalité tel qu'il se construit dans la théorie écologique ne permet pas de stabiliser cet équilibre dans la mesure où il admet, comme nous l'avons vu, une pluralité de pratiques interprétatives en cohérence avec des hypothèses auxiliaires de modélisation. La position à laquelle nous aboutissons semble souffrir des mêmes insuffisances épistémologiques que le réalisme interne de Putnam dans *Reason, Truth and History*. En effet, comme dans le réalisme interne, la limitation contextuelle fournit un critère d'adéquation des modèles sous des conditions épistémiques suffisamment bonnes², formulées ici en particulier par les hypothèses auxiliaires. Toutefois, comme l'écrit Putnam dans un ouvrage récent sur la question du réalisme, on peut se poser la question de savoir "comment l'on peut avoir un accès référentiel ou autre aux 'conditions épistémiques suffisamment bonnes'"³. Une telle image reste prisonnière des présuppositions épistémologiques de la modernité en ce qu'elle "retient la prémisse d'une

1 MAESSCHALCK M., *Normes et contextes*, op. cit., p. 184.

2 Cf. PUTNAM H., *Reason, Truth and History*, Cambridge, 1981, p. 54.

3 PUTNAM H., *The Threefold Cord : Mind, Body and World*, New York, 1999, p. 18.

interface entre le sujet connaissant et tout le reste ‘à l’extérieur’⁴. L’interprétation contextuelle, tant qu’elle est comprise comme un schème mental qui s’entrepasse entre nos modèles et une réalité externe à modéliser ne pourra fournir un critère de stabilisation, étant donné que l’on n’a pas d’accès référentiel aux conditions auxiliaires de la modélisation. Toutefois, la stabilisation s’opère bel et bien et c’est ce qui amène Putnam à envisager, à la suite de philosophes comme W. James, E. Husserl, L. Wittgenstein et J. Austin⁵, la position alternative du “réalisme naturel” que l’on peut également appeler un réalisme non fondationnel : “notre croyance dans le réalisme ne tient pas au fait qu’une réalité externe à laquelle nous n’avons pas accès ‘cause’ des expériences mentales internes, selon l’hypothèse des *qualia*, mais tient au fait que les objets ‘externes’, choux et rois, peuvent être objet d’expérience”⁶. Cette hypothèse alternative d’un lien génétique entre la croyance dans une réalité externe qui opère dans le réalisme scientifique et celle qui opère dans l’expérience naturelle du monde permet d’interpréter autrement la stabilisation du projet de naturalisation. En effet, en nous basant sur la proposition de Putnam, nous pouvons concevoir le troisième terme, qui permet de stabiliser l’ajustement entre la perception contextualisante et les conditions contextuelles de stabilisation de la perception, de façon plus adéquate comme une disposition cognitive à la stabilisation, une ouverture au remplissement possible de la règle par le monde, telle qu’elle opère déjà dans nos “transactions”⁷ naturelles avec les objets de ce monde. Précisons tout de suite que ce que Putnam appelle le réalisme naturel ne peut aucunement être confondu avec le réalisme du sens commun. Il s’agit bien d’une posture réflexive interne au jeu de langage philosophique, une “seconde naïveté”⁸ comme le dit Putnam, qui ne vise pas à proposer une nouvelle théorie de la perception ou une “nouvelle image métaphysique”⁹, mais une reconstruction réflexive de “l’accès conceptuel que nous avons aux choses au sujet duquel nous parlons et réfléchissons”¹⁰.

Le traitement réflexif de la contextualisation, tel qu’on peut l’entrevoir à partir de la proposition de H. Putnam, peut être approfondi en le confrontant à l’analyse que fait M. Maesschalck des limitations contextuelles dans un autre domaine particulier, celui de la théorie de la norme. Tout d’abord, comme le note M. Maesschalck, la limitation contextuelle maintient une

4 *Ibid.*

5 Cf. *ibid.*, p. 24.

6 *Ibid.*, p. 20.

7 *Ibid.*, p. 169.

8 *Ibid.*, p. 44.

9 *Ibid.*, p. 41.

10 *Ibid.*, p. 45 ; cf., également, *ibid.*, p. 157.

conception holiste du contexte, même si c'est un holisme pragmatique qui maintient la possibilité d'une pluralité de pratiques interprétatives au sein d'un même contexte¹¹. Toutefois, "le holisme d'une perception contextualisante, écrit M. Maesschalck, n'accéderait jamais à l'intellection, s'il n'était lui-même réinterprété à partir d'une pratique volontaire de contextualisation dont la visée reproduit et entretient la contingence des liens factuels"¹². Dans le holisme, l'opération de contextualisation n'est donc pas construite pour elle-même, et "l'hypothèse selon laquelle l'accès à un exercice de la raison ou la volonté même de trouver les moyens rationnels de traiter certains types de problèmes constitue aussi une forme de vie et que cette forme de vie est susceptible d'autotransformation, cette hypothèse est laissée de côté, parce que la raison reste conçue en définitive comme un mécanisme de compensation par rapport à des contraintes factuelles"¹³. Or, une autre solution pour la construction de l'opération de contextualisation semble être plus prometteuse que la position d'un hypothétique mécanisme de compensation par une réalité à jamais inaccessible. Celle-ci consiste à prendre en compte la cohérence opératoire des contextes "non comme un anticipable-projetable", mais comme susceptible d'informer l'ajustement entre l'action et le contexte en entraînant un "processus réflexif de décentrement et de déplacement de l'activité normative vers les transformations de son contexte"¹⁴. Dans ce cadre, l'opération de contextualisation peut être comprise de façon plus précise comme une ouverture réflexive à des ajustements possibles entre l'action et le contexte dans des situations déterminées ou, selon le modèle phénoménologique, comme un "type" de la raison pratique, qui tire sa dimension concrète "de son expérimentation dans une situation effective par l'action"¹⁵. En prenant en compte la mouvance du contexte, le type ouvre à la logique modale¹⁶, à une logique de possibilisation d'ajustements nouveaux, là où la conception schématique inhérente au holisme permet seulement d'envisager une logique de subsomption des ajustements sous des interprétations contextuelles déjà données.

11 Cf. MAESSCHALCK M., "Typique transcendantale et typique phénoménologique, I. Les enjeux d'une recherche sur la typique de la raison pratique", in *Les Carnets du Centre de Philosophie du Droit*, n°60 (1998), pp. 11–12.

12 *Ibid.*, p. 12.

13 MAESSCHALCK M., "Habitue et lien social, IV. L'apport de Sartre à une typique du contexte", in *Les Carnets du Centre de Philosophie du Droit*, n°54 (1998), p. 21.

14 *Ibid.*, pp. 20–21.

15 MAESSCHALCK M., "Typique transcendantale et typique phénoménologique, I. Les enjeux d'une recherche sur la typique de la raison pratique", *op. cit.*, p. 14.

16 Cf. *ibid.*, p. 21.

Le modèle phénoménologique de la disposition cognitive d'ouverture aux possibles permet de préciser la conception réflexive de la contextualisation dans la perspective d'un "mécanisme inférentiel"¹⁷. A partir de là, nous pouvons tenter de formuler une réponse aux impasses d'une prise en compte non réflexive des limitations contextuelles de la théorie écologique. Comme nous l'avons vu, la configuration particulière qui émerge à partir de la critique épistémologique des limitations de la modélisation montre l'instabilité du projet de la naturalisation de l'intentionnalité quand celui-ci reste dans le cadre épistémologique d'un réalisme interne. Ce qui n'est pas pris en compte dans un tel cadre, c'est la construction de l'opérateur de contextualisation mis en œuvre dans la critique épistémologique. Une construction de cette opération pour elle-même, comme ouverture réflexive à des possibilités effectives d'ajustement entre l'action et son contexte, permet alors de comprendre les différentes pratiques interprétatives particulières comme des traductions circonstanciées de la contextualisation de l'action comme "opérateur de modalisation"¹⁸ de celle-ci. Ce dernier modèle permet de prendre en compte la pluralité des pratiques interprétatives, tout en reconnaissant la constitution réflexive de l'opération de contextualisation comme condition de possibilisation d'ajustements nouveaux entre l'action et le contexte.

17 C'est un tel mécanisme que tente de développer pour lui-même la recherche menée par Jacques Lenoble et Marc Maesschalck dans *The Action of Norms*, cf. LENOBLE J. et MAESSCHALCK M., *The Action of Norms*, London/New York, à paraître.

18 Cf. MAESSCHALCK M., *Normes et contextes*, *op. cit.*, p. 10.

Bibliographie

- ANDLER D., “Calcul et représentation : les sources”, in ANDLER D. (ed.), *Introduction aux sciences cognitives*, Paris, Gallimard, 1992, pp. 9–46.
- ATLAN H., “Self-organizing Networks : Weak, Strong and Intentional, the Role of their Undertermination”, in *La Nuova Critica*, 19–20 (1992), pp. 51–70.
- ATLAN H., “Projet et signification dans les réseaux d'automates : Le rôle de la sophistication”, in JANICAUD D. (ed.), *L'intentionnalité en question. Entre phénoménologie et sciences cognitives*, Paris, Vrin, 1995, pp. 261–288.
- BALLARD D., “Animate Vision”, in *Artificial Intelligence*, 48 (1991), pp. 57–86.
- BECKERMANN A., “Visuelle Informationsverarbeitung und phänomenales Bewußtsein”, in METZINGER TH. (ed.), *Bewußtsein. Beiträge aus der Gegenwartphilosophie*, Paderborn, Schöningh, 1996, pp. 663–684.
- BERK L. et GARVIN R., “Development of Private Speech among Low Income Appalachian Children”, in *Developmental Psychology*, 20 (1984), pp. 271–286.
- BERTEN A. et LENOBLE J., *Dire la norme*, Paris/Bruxelles, L.G.D.J/Bruylant, 1990.
- BERTHOZ A., *Le sens du mouvement*, Paris, Odile Jacob, 1997.
- BETH E., *Inleiding tot de Wijsbegeerte der Wiskunde*, Antwerpen, Standaard Boekhandel, 1940.
- BOISACQ-SCHEPENS N. et CROMMELINCK M., *Neuro-psycho-physiologie*, Paris, Masson, 1994.
- BORING E., *A History of Experimental Philosophy*, New York, Appleton-Century, 1929.
- BOURDIEU P., *Le sens pratique*, Paris, Ed. de Minuit, 1980.
- BRISART R., “La réduction et l'irréductible phénoménologiques. Husserl critique de Heidegger”, in BRISART R. et CÉLIS R., *L'évidence du monde. Méthode et empirie de la phénoménologie*, Bruxelles, Publications des Facultés universitaires Saint-Louis, 1994, pp. 139–185.
- BROOKS R., “Intelligence without Representation”, in *Artificial Intelligence*, 47 (1991), pp. 139–159.
- BUNGE M., *Scientific Research*, Berlin/Heidelberg/New York, Springer, 1967.
- BUNGE M., *Method, Model and Matter*, Boston/Dordrecht, D. Reidel, 1973.

- CHURCHLAND P.M., *A Neurocomputational Perspective*, Cambridge (MA), MIT Press, 1989.
- CHURCHLAND P.M., *The Engine of Reason, the Seat of the Soul*, Cambridge (MA), MIT Press, 1995.
- CHURCHLAND P.M. et CHURCHLAND P.S., "Stalking the Wild Epistemic Engine", in *Noûs*, 17 (1983), pp. 5–18.
- CHURCHLAND P.S., *Neurophilosophie. Vers une science de l'esprit cerveau*, Paris, P.U.F., 1999.
- CHURCHLAND P.S., RAMACHANDRAN V.S. et SEJNOWSKI T.J., "A Critique of Pure Vision", in KOCH C. et DAVIS J. (eds.), *Large-scale Neuronal Theories of the Brain*, Cambridge (MA), MIT Press, 1994, pp. 23–60.
- CLARK A., *Associative Engines : Connectionism, Concepts and Representational Change*, Cambridge (MA), MIT Press, 1993.
- CLARK A., *Being There : Putting Brain, Body and World Together Again*, Cambridge (MA)/London, MIT Press, 1997.
- COHEN J.D., ROMERO R.D., SERVAN-SCHREIBER D. et FARAH M.J., "Mechanisms of Spatial Attention : The Relation of Macrostructure to Microstructure in Parietal Neglect", in *Journal of Cognitive Neuroscience*, 6 (1994), pp. 377–387.
- COLTHEART M. et DAVIES M., "Le concept de modularité à l'épreuve de la neuropsychologie", in ANDLER D. (ed.), *Introduction aux sciences cognitives*, Paris, Gallimard, pp. 109–130.
- COTTRELL G., "Extracting Features from Faces Using Compression Networks", in TOURETSKY D. et al. (eds.), *Connectionist Models : Proceedings of the 1990 Summer School*, San Mateo (CA), Morgan Kaufman, 1991.
- CUMMINS R. et SCHWARZ G., "Connexionnisme, computation et cognition", in ANDLER D. (ed.), *Introduction aux sciences cognitives*, Paris, Gallimard, 1992, pp. 374–394.
- DAMASIO A., *Descartes' Error : Emotion, Reason and the Human Brain*, New York, Grosset-Putnam, 1994.
- DAMASIO A. et DAMASIO H., "Cortical Systems for Retrieval of Concrete Knowledge : The Convergence Zone Framework", in KOCH C. et DAVIS J. (eds.), *Large-scale Neuronal Theories of the Brain*, Cambridge (MA), MIT Press, 1994.
- DEDEURWAERDERE T., "Neural Networks and the Brain : Associative Learning and/or Self-organisation ?", in DUBOIS D. (ed.), *Proceedings of the First International Conference on Computing Anticipatory Systems*, Liège, Chaos, 1998, pp. 170–179.

- DEDEURWAERDERE T., “Le renversement cognitiviste et les théories de la conscience”, in *Revue Philosophique de Louvain*, 98 (2000), pp. 732–760.
- DEDEURWAERDERE T., CASAS-VAZQUEZ J., JOU D. et LEBON G., “Foundations and Applications of a Mesoscopic Thermodynamic Theory of Fast Phenomena”, in *Physical Review E.*, 53 (1996), pp. 498–506.
- DENNETT D., *The Intentional Stance*, Cambridge (MA), Cambridge University Press, 1987.
- DREYER W. et STRUCHTRUP H., “Heat Pulse Experiments Revisited”, in *Continuum Mechanics and Thermodynamics*, 5 (1993), pp. 3–50.
- DREYFUS H.L., “La portée philosophique du connexionnisme”, in ANDLER D. (ed.), *Introduction aux sciences cognitives*, Paris, Gallimard, 1992 , pp. 352–373.
- DREYFUS H.L. et DREYFUS S.E., “How to Stop Worrying About the Frame Problem Even Though It's Computationally Insoluble”, in PYLYSHIN Z.E. (ed.), *The Robot's Dilemma : the Frame Problem in Artificial Intelligence*, 1987, pp. 95–111.
- EDELMAN G.M., *Neural Darwinism*, New York, Basic Books, 1987.
- ENGEL P., “La philosophie de l'esprit”, in MEYER M. (ed.), *La philosophie anglo-saxonne*, Paris, P.U.F., 1994, pp. 529–564.
- ENGEL P., *Philosophie et psychologie*, Paris, Gallimard, 1996.
- FELTZ B., *Croisées biologiques*, Louvain-la-Neuve, Ed. Ciaco, 1991.
- FEYERABEND P., “Problems of Empiricism”, in COLODNY R.G. (ed.), *Beyond the Edge of Certainty*, Englewood Cliffs (N.J.), Prentice-Hall, 1965, pp. 145–260.
- FISSETTE D. et POIRIER P., *Philosophie de l'esprit. Etat des lieux*, Paris, Vrin, 2000.
- FODOR J., *The Language of Thought*, Cambridge (MA), Harvard University Press, 1979.
- GÉLY R., *La genèse du sentir*, Bruxelles/Paris, Ousia, 2000.
- GLYMOUR C., “On the Methods of Cognitive Neuropsychology”, in *The British Journal for the Philosophy of Science*, 45 (1994), pp. 815–835.
- GOULD S.J., *Le pouce du Panda. Les grandes énigmes de l'évolution*, trad. par J. Chabert, Paris, Grasset, 1982.
- HABERLANDT K., *Cognitive Psychology*, Cambridge (MA), Paramount, 1994.
- HABERMAS J., “What is Universal Pragmatics ?”, in HABERMAS J., *Communication and the Evolution of Society*, trad. par Th. McCarthy, London, Heinemann, 1979, pp. 8–14.

- HAUGELAND J., "Representational Genera", in RAMSEY W. *et al.* (eds.), *Philosophy and Connectionist Theory*, Hillsdale, Erlbaum, 1991, pp. 61–89.
- HEBB D., *The Organization of Behavior*, New York, Wiley, 1949.
- HEIMS S. J., *The Cybernetics Group*, Cambridge (MA), MIT Press, 1991.
- HENRY M., *Incarnation. Une philosophie de la chair*, Paris, Seuil, 2000.
- HÉRAULT J., JUTTEN C. et ANS B., "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé", in *Actes du Xe Colloque Gretsi*, Nice, 1985, pp. 1017–1022.
- HILDITCH D., *At the Heart of the World : Merleau-Ponty and the Existential Phenomenology of Embodied and Embedded Intelligence in Everyday Coping*, Diss., St. Louis (MO), Washington University, 1995.
- HOLENSTEIN E., *Phänomenologie der Assoziation. Zu Struktur und Funktion eines Grundprinzipes der Passiven Genesis bei E. Husserl*, Den Haag, Martinus Nijhoff, 1972.
- HUANG K., *Statistical Mechanics*, New Delhi, J. Wiley Eastern, 1988.
- HUSSERL E., *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*, rédigé et édité par L. Landgrebe, Hamburg, Claassen & Goverts, 1954.
- HUSSERL E., *Husserliana XI : Analysen zur passiven Synthesis. Aus Vorlesungs- und Forschungsmanuskripten 1918–1924*, édité par Margot Fleischer, Den Haag, Martinus Nijhoff, 1966.
- HUSSERL E., *Husserliana XVII : Formale und transzendente Logik. Versuch einer Kritik der logischen Vernunft. Mit ergänzenden Texten*, édité par P. Janssen, Den Haag, Martinus Nijhoff, 1974.
- HUSSERL E., *Logique formelle et logique transcendantale. Essai d'une critique de la raison logique*, trad. par Suzanne Bachelard, Paris, P.U.F., 1965.
- HUSSERL E., *Expérience et jugement. Recherches en vue d'une généalogie de la logique*, trad. par Denise Souche-Dagues, Paris, P.U.F., 1970.
- HUSSERL E., *De la synthèse passive. Logique transcendantale et constitutions originaires*, trad. par Bruce Bégout et Jean Kessler avec la collaboration de Natalie Depraz et Marc Richir, Grenoble, Jérôme Millon, 1998.
- HUTCHINS E., *Cognition in the Wild*, Cambridge (MA), MIT Press, 1995.
- JACQUINOT-DELAUNAY G. et MONNOYER L. (coord.), "Le dispositif. Entre usage et concept", in *Hermes*, 25 (1999), pp. 5–244.
- JOU D., CASAS-VAZQUEZ J. et LEBON G., *Extended Irreversible Thermodynamics*, Berlin/Heidelberg, Springer, 1993.

- KARMILOFF-SMITH A., *Beyond Modularity : A Developmental Perspective on Cognitive Science*, Cambridge (MA), Bradford/MIT Press, 1992.
- KEIJZER F. et BEM S., "Behavioral Systems Interpreted as Autonomous Agents and as Coupled Dynamical Systems : a Criticism", in *Philosophical Psychology*, 9 (1996), pp. 323–346.
- KITCHER P., "Marr's Computational Theory of Vision", in *Philosophy of Science*, 55 (1988), pp. 1–24.
- KOHONEN T., "Self-organized Formation of Topologically Correct Feature Maps", in *Biological Cybernetics*, 43 (1982), pp. 59–69.
- KOKOSZKA V., "La conception husserlienne de la temporalité entre 1905 et 1910", in *Tijdschrift voor Filosofie*, 58 (1996), pp. 314–341.
- KOKOSZKA V., *Les conditions transcendantales du devenir commun. Ontologie et monadologie chez Husserl*, Diss., Louvain-la-Neuve, Université catholique de Louvain, 2001.
- KÜHN R., *Leben als Bedürfen. Eine lebensphänomenologische Analyse zu Kultur und Wirtschaft*, Heidelberg, Physica/Springer, 1996.
- KÜHN R., *Husserls Begriff der Passivität. Zur Kritik der passiven Synthesis in der genetischen Phänomenologie*, Freiburg/München, Karl Alber, 1998.
- LADRIÈRE J., "Les limites de la formalisation", in PIAGET J. (dir.), *Logique et connaissance scientifique (Encyclopédie de la Pléiade)*, Paris, Gallimard, 1967, pp. 312–333.
- LADRIÈRE J., "La cybernétique", in *Encyclopaedia Universalis*, Vol. 5, Paris, 1970, pp. 256–258.
- LADRIÈRE J., "Le formalisme et le sens", in MOULOUD N. (coord.), *Les langages, le sens, l'histoire*, Villeneuve d'Ascq., Publications de l'Université de Lille III, 1975, pp. 241–277.
- LADRIÈRE J., *Les enjeux de la rationalité. Le défi de la science et de la technologie aux cultures*, Paris, Aubier-Montaigne/Unesco, 1977.
- LADRIÈRE J., *L'articulation du sens*, 2 tomes, Paris, Cerf, 1984.
- LADRIÈRE J., *L'éthique dans l'univers de la rationalité*, Namur, Artel/Fides, 1997.
- LAVE J., *Cognition in Practice : Mind, Mathematics and Culture in Everyday Life*, Cambridge, Cambridge University Press, 1988.
- LENOBLE J., *Droit et communication*, Paris, Cerf, 1994.
- LENOBLE J., "Law and Undecidability : Towards a New Vision of the Proceduralisation of Law" in ARATO A. et ROSENFELD M. (eds.), *Habermas on Law and Democracy : Critical Exchanges*, Berkeley, University of California Press, 1998, pp. 37–81.

- LENOBLE J., “La théorie de la loi et l'hypothèse de la procéduralisation contextuelle”, in *Les Carnets du Centre de Philosophie du Droit*, n°70 (1999), 28 pp.
- LENOBLE J., “La procéduralisation contextuelle du droit”, in COPPENS PH. et LENOBLE J. (eds.), *Démocratie et procéduralisation du droit*, Bruxelles, Kluwer, 2000, pp. 97–124.
- LENOBLE J. et MAESSCHALCK M., *The Action of Norms*, London/New York, Kluwer International, à paraître.
- LOBELLE M., *Systèmes informatiques*, syllabus dactylographié, Louvain-la-Neuve, Faculté des Sciences Appliquées, 1992.
- MAC LANE S., *Mathematics : Form and Function*, New York, Springer, 1986.
- MAESSCHALCK M., “Pour une critique du formalisme en éthique. En hommage à Jean Ladrière”, in *Science et Esprit*, 48 (1996), pp. 135–152.
- MAESSCHALCK M., “Compétences langagières et ressources compréhensives”, in *Science et Esprit*, 49 (1997), pp. 259–280.
- MAESSCHALCK M., “Typique transcendantale et typique phénoménologique, I. Les enjeux d'une recherche sur la typique de la raison pratique”, in *Les Carnets du Centre de Philosophie du Droit*, n°60 (1998), 21 pp.
- MAESSCHALCK M., “Habitus et lien social, IV. L'apport de Sartre à une typique du contexte”, in *Les Carnets du Centre de Philosophie du Droit*, n°54 (1998), 22 pp.
- MAESSCHALCK M., “Une éthique des styles de vie ? Questions au schématisme de Bourdieu”, in *Science et Esprit*, 50 (1998), pp. 155–176.
- MAESSCHALCK M., “Ressources compréhensives et contextualisation des normes. Une limitation du formalisme procédural”, in *Ethica*, 11 (1999), pp. 33 à 58.
- MAESSCHALCK M., “La conscience absolue comme éveil à la déterminabilité du monde”, in GODDARD J.-C. (ed.), *Le transcendantal et le spéculatif dans l'idéalisme allemand*, Paris, Vrin, 1999, pp. 77–96.
- MAESSCHALCK M., *Normes et contextes. Les fondements d'une pragmatique contextuelle*, Hildesheim/Zürich/New York, Olms, 2001.
- MAESSCHALCK M. et KOKOSZKA V., “Phénoménologie et auto-organisation”, in FELTZ B., CROMMELINCK M. et GOUJON PH. (eds.), *Auto-organisation et émergence dans les sciences de la vie*, Bruxelles, Ousia, 1999, pp. 405–420.
- MARGENAU H. et LINDSAY R.B., *Foundations of Physics*, Woodbridge, Ox Bow Press, 1981.
- MARR D., *Vision*, San Francisco, Freeman, 1982.

- MATARIC M., "Navigating with a Rat Brain : a Neurobiologically Inspired Model for Robot Spatial Representation", in MEYER J.-A. et WILSON S. (eds.), *From Animals to Animats I*, Cambridge (MA), MIT Press, 1991.
- MATURANA H. et VARELA F., *De boom der kennis. Hoe wij de wereld door onze eigen waarneming creëren*, trad. par T. Maas et C. Oks, Doordrecht, D. Reidel, 1988.
- MCCULLOCH W.S. et PITTS W., "A Logical Calculus of the Ideas Immanent in Nervous Activity", in *Bulletin of Mathematical Biophysics*, 5 (1943), pp. 115–133.
- MERLEAU-PONTY M., *Phénoménologie de la perception*, Paris, Gallimard, 1945.
- MÉTRAUX A., "Die Mikrophysik der Wahrnehmung und des Gedächtnisses in der französischen Aufklärung", in FLOREY E. et BREIDBACH O. (eds.), *Das Gehirn, Organ der Seele ?*, Berlin, Akademie Verlag, 1993, pp. 129–150.
- MISSA J.-N., *L'esprit-cerveau. La philosophie de l'esprit à la lumière des neurosciences*, Paris, Vrin, 1993.
- MITTELSTRAß J. (ed.), *Enzyklopädie Philosophie und Wissenschaftstheorie*, Mannheim/Wien/Zürich, Bibliographisches Institut, 1984.
- MONOD J., *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*, Paris, Seuil, 1970.
- MORMANN T., "Husserl's Philosophy of Science and the Semantical Approach", in *Philosophy of Science*, 58 (1991), pp. 61–83.
- NAGEL E., *The Structure of Science*, Indianapolis, Hackett, 1979.
- PAILLARD J., "Réflexions sur l'usage du concept de plasticité en neurobiologie", in *Journal de Psychologie Normale et Pathologique*, 73 (1976), pp. 33–47.
- PETERSEN S.E., FOX P.T., POSNER M.I., MINTUN M. et RAICHLE M.E., "Positron Emission Tomographic Studies of the Processing of Single Words", in *Journal of Cognitive Neurosciences*, 1 (1989), pp. 153–170.
- PLATTEN J. et LEGROS J., *Convection in Liquids*, Berlin, Springer, 1984.
- PRIGOGINE I. et STENGERS I., *La nouvelle alliance*, Paris, Gallimard, 1986.
- PUTNAM H., *Reason, Truth and History*, Cambridge, Cambridge University Press, 1981.
- PUTNAM H., *Representation and reality*, Cambridge (MA), MIT Press, 1989.
- PUTNAM H., *The Threefold Cord : Mind, Body and World*, New York, Columbia University Press, 1999.
- QUINE W.V.O., "Les deux dogmes de l'empirisme", in JACOB P., *De Vienne à Cambridge, l'héritage du positivisme logique*, Paris, Gallimard, 1980, pp. 87–112.

- QUINN R. et ESPENSCHIED K., "Control of a Hexapod Robot using a Biologically Inspired Neural Network", in BEER R. *et al.* (eds.), *Biological Neural Networks in Invertebrate Neuroethology and Robotics*, Boston, Academic Press, 1993.
- RAMACHANDRAN V.S., "Interactions between Motion, Depth, Color and Form : the Utilitarian Theory of Perception", in BLAKEMORE C. (ed.), *Vision : Coding and Efficiency*, Cambridge, Cambridge University Press, 1990, pp. 346–360.
- ROCK I., *The Logic of Perception*, Cambridge (MA)/London, MIT Press, 1983.
- RORTY R., "Consciousness, Intentionality, and Pragmatism", in SCOTT M., CHRISTENSEN S. et TURNER D. (eds.), *Folk Psychology and the Philosophy of Mind*, New Jersey, Erlbaum, 1993, pp. 338–404.
- ROSENBERG A., *The Structure of Biological Science*, Cambridge, Cambridge University Press, 1985.
- RUMELHART D. et MCCLELLAND J. (eds.), *Parallel Distributed Processing : Explorations in the Microstructure of Cognition. Volume 1 : Foundations, Volume 2 : Psychological and Biological Models*, Cambridge (MA), MIT Press, 1986.
- RUMELHART D., SMOLENSKY P., MCCLELLAND J. et HINTON G., "Schemata and Sequential Thought Processes", in RUMELHART D. et MCCLELLAND J. (eds.), *Parallel Distributed Processing*, Cambridge (MA), MIT Press, 1986.
- SCHIEBER M., "How Might the Motor Cortex Individuate Movements ?", in *Trends in Neuroscience*, 13 (1990), pp. 440–444
- SCHIEBER M. et HIBBARD L., "How Somatopic is the Motor Cortex Hand Area ?", in *Science*, 261 (1993), pp. 489–492.
- SCHÜTZ A., "Typus und Eidos in Husserls Spätphilosophie", in *Alfred Schütz. Gesammelte Aufsätze III. Studien zur Phänomenologische Philosophie*, Den Haag, Martinus Nijhoff, 1971, pp. 127–152.
- SEARLE J., *L'intentionnalité. Essai de philosophie des états mentaux*, trad. par Claude Pichevin, Paris, Ed. de Minuit, 1985.
- SÉGAL L., *Le rêve de la réalité. Heinz von Foerster et le constructivisme*, trad. par Anne-Lise Hacker, Paris, Seuil, 1990.
- SEJNOWSKI T. et ROSENBAUM C., "NETtalk : A Parallel Network that Learns to Read Aloud", in *John Hopkins University Electrical Engineering and Computer Science Technical Report*, n° 86/01 (1986).
- SERON X., "La neuropsychologie cognitive", in FRANCK R. (ed.), *Faut-il chercher aux causes une raison ? L'explication causale dans les sciences humaines*, Paris, Vrin, 1994, pp. 344–366.

- SHALLICE T., *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press, 1988.
- SMITHERS T., "Why Better Robots Make It Harder", in CLIFF D. *et al.* (eds.), *From Animals to Animats III*, Cambridge (MA), MIT Press, 1994, pp. 64–66.
- SNEED J., *The Logical Structure of Mathematical Physics*, Dordrecht, D. Reidel, 1971.
- SOULEZ A. (ed.), *Manifeste du cercle de Vienne et autres écrits*, Paris, P.U.F., 1985.
- STEELS L., "The Artificial Life Roots of Artificial Intelligence", in *Artificial Life*, 1 (1994), pp. 75–110.
- STUFFLEBEAM R.S. et BECHTEL W., "PET : Exploring the Myth and the Method", in *Philosophy of Science*, supplement (1997).
- SUPPE F., *The Structure of Scientific Theories*, Chicago, University of Illinois Press, 1977.
- SUPPE F., *The Semantic Conception of Theories and Scientific Realism*, Chicago, University of Illinois Press, 1989.
- SUPPES P., *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*, Dordrecht, D. Reidel, 1969.
- THELEN E. et SMITH L., *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge (MA), MIT Press, 1994.
- TRIANTAFYLLOU M. et TRIANTAFYLLOU G., "An Efficient Swimming Machine", in *Scientific American*, 272 (1995), pp. 64–71.
- UEXKÜLL J.V., *Mondes animaux et monde humain*, Paris, Denoël, 1965.
- ULLMAN S., *The Interpretation of Visual Motion*, Cambridge (MA), MIT Press, 1979.
- VAN DE VIJVER G., *Van cybernetica naar connectionisme. Een epistemologische studie van doelgerichtheid*, Gent, Academia Press, 1991.
- VAN DE VIJVER G., "Self-organization and Naturalism : Perspectives and Deadlocks of the New Connectionism of Atlan and Varela", in *The Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology*, 6 (1989), pp. 177–198.
- VAN ESSEN D. et GALLANT J., "Neural Mechanisms of Form and Motion Processing in the Primate Visual System", in *Neuron*, 13 (1994), pp. 1–10.
- VAN ESSEN D., ANDERSON C. et OLSHAUSEN B., "Dynamic Routing Strategies in Sensory, Motor and Cognitive Processing", in KOCH C. et DAVIS J. (eds.), *Large-scale Neuronal Theories of the Brain*, Cambridge (MA), MIT Press, 1994.
- VAN FRAASSEN B., *The Scientific Image*, London, Oxford University Press, 1980.

- VARELA F., THOMPSON E. et ROSCH E., *L'inscription corporelle de l'esprit. Sciences cognitives et expérience humaine*, trad. par Véronique Havelange, Paris, Seuil, 1993.
- VAUCLAIR J., *L'intelligence de l'animal*, Paris, Seuil, 1992.
- VERLEYSSEN M. et BLAYO F., *Les réseaux de neurones artificiels*, Paris, P.U.F., 1996.
- VOGEL S., *Life in Moving Fluids*, Princeton, Princeton University Press, 1983.
- VYGOTSKI L., *Pensée et langage*, trad. du Russe par Françoise Sève, Paris, Editions Sociales, 1985 (éd. originale en 1934).
- WIENER, N., *Cybernetics or Control and Communication in the Animal and the Machine*, New York, John Wiley and Sons, 1948.
- WINOGRAD T. et FLORES F., *L'intelligence artificielle en question*, trad. de J.-L. Peytavin, Paris, P.U.F., 1989.
- WUNDT W., *Vorlesungen über die Mensch und Tierseele*, Hamburg/Leipzig, Leopold Voss, 1906.
- ZAWIDZKI T. et BECHTEL W., "Gall's Legacy Revisited : Decomposition and Localization in Cognitive Neuroscience", in ERNELING C.E. et JOHNSON D.M. (eds.), *Mind as a Scientific Object : Between Brain and Culture*, Oxford, Oxford University Press, sous presse.

Index onomastique

- ANDERSON C., 141.
ANDLER D., 24.
ANS B., 40.
ARISTOTE, 210.
ATLAN H., 12, 14, 15, 16, 17, 18, 19,
143, 147, 152, 153, 155, 185, 187,
211, 213.
AUSTIN J., 216.
BALLARD D., 65.
BARLOW D.H., 102.
BATESON G., 143.
BECHTEL W., 112, 132.
BECKERMANN A., 42.
BÉGOUT B., 186.
BEM S., 3.
BERK L., 87.
BERTEN A., 6, 7.
BERTHOZ A., 70, 71, 72, 73, 81, 84,
148, 214.
BETH E., 161, 167.
BLAYO F., 26, 34, 35, 40.
BOISACQ-SCHEPENS N., 137, 140,
147.
BORING E., 22.
BOURDIEU P., 43.
BRISART R., 186.
BROCA P., 128, 130.
BROOKS R., 1, 80, 81, 93, 119, 148,
175, 181, 183.
BUNGE M., 154, 160, 163, 171, 172,
174, 176, 177.
CASAS-VAZQUEZ J., 160.
CÉLIS R., 186.
CHOMSKY N., 106, 107.
CHURCHLAND P.M., 5, 30, 31, 32, 33,
34, 36, 37, 102, 106, 174.
CHURCHLAND P.S., 5, 61, 64, 65, 66,
106, 111, 135, 136, 141.
CLARK A., 26, 38, 50, 51, 52, 61, 73,
76, 77, 78, 79, 80, 82, 83, 84, 85,
86, 87, 88, 89, 90, 93, 95, 96, 119,
122, 123, 126, 129, 137, 138, 140,
149, 174, 183.
COHEN J.D., 133.
COLTHEART M., 127, 128, 129.
COTTRELL G., 32, 34, 35, 36, 37, 38,
39, 41, 102.
CROMMELINCK M., 137, 140, 147,
185.
CUMMINS R., 21.
DAMASIO A., 46, 141.
DAMASIO H., 141.
DAVIDSON D., 13.
DAVIES J., 127, 128, 129.
DENNETT D., 12, 13, 106, 107.
DEPRAZ N., 186.
DESCARTES R., 21, 47, 188, 196, 204,
209.
DREYER W., 92.
DREYFUS H.L., 28, 55, 56.
DREYFUS S.E., 55.
EDELMAN G.M., 174.
ENGEL P., 11, 106, 107.
ESPENSCHIED K., 50, 79, 149.
EUCLIDE, 160.
FARAH M.J., 133.
FELTZ B., 163, 185.
FEYERABEND P., 157, 159.
FICHTE J.G., 206.
FISETTE D., 2.
FODOR J., 2, 106, 107, 108, 109.
FOX P.T., 132.
GALLANT J., 140, 141.
GARVIN R., 87.
GÉLY R., 53.
GIBSON J., 214.
GLYMOUR C., 129, 130, 131.
GÖDEL K., 167.

- GOUJON PH., 185.
 GOULD S.J., 69, 111.
 HABERLANDT K., 27.
 HABERMAS J., 4, 6, 106, 107.
 HANSON N.R., 159.
 HAUGELAND J., 122.
 HEBB D., 34, 174.
 HEIMS S.J., 143, 175.
 HEISENBERG W., 159.
 HENRY M., 203, 204, 205, 206, 207.
 HÉRAULT J., 40, 93.
 HIBBARD L., 138, 141.
 HILBERT D., 160.
 HILDITCH D., 45, 46, 47, 51, 52, 53, 55.
 HINTON G., 85.
 HOLENSTEIN E., 201.
 HUANG K., 120.
 HUBEL D.H., 102.
 HUSSERL E., 159, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 197, 199, 200, 201, 202, 203, 204, 206, 207, 209, 216.
 HUTCHINS E., 1, 95, 96, 174.
 JACOB P., 158.
 JACQUINOT-DELAUNAY G., 86.
 JAMES W., 216.
 JANICAUD D., 12.
 JOU D., 160.
 JUTTEN C., 40, 93.
 KANT I., 118.
 KARMILOFF A., 53.
 KEIJZER F., 3.
 KITCHER P., 102, 103, 105, 107, 108, 110, 126.
 KOHONEN T., 39, 40, 41, 68, 93.
 KOKOSZKA V., 185, 205, 207.
 KÜHN R., 199, 205.
 LADRIÈRE J., 4, 7, 8, 90, 118, 119, 154, 165, 167, 168, 209.
 LAVE J., 54, 55, 58.
 LEBON G., 160.
 LEGROS J., 91.
 LEIBNIZ G.W., 172.
 LENOBLE J., 6, 8, 218.
 LINDSAY R.B., 159.
 LIVET P., 7, 9.
 LOBELLE M., 24.
 LUHMANN N., 143.
 MAC LANE S., 160.
 MAESSCHALCK M., 4, 5, 6, 7, 9, 43, 167, 169, 185, 206, 215, 216, 217, 218.
 MARGENAU H., 159.
 MARR D., 19, 60, 64, 68, 100, 101, 102, 103, 104, 105, 107, 110, 111, 112, 113, 121, 125, 126, 142, 143, 152, 178, 212.
 MATARIC M., 73.
 MATURANA H., 143, 145, 146, 147, 148.
 MCCLELLAND J., 86.
 MCCULLOCH W.S., 33, 122.
 MERLEAU-PONTY M., 45, 46, 47, 48, 49, 52, 53, 90.
 MÉTRAUX A., 21.
 METZINGER TH., 42.
 MINTUN M., 132.
 MISSA J.-N., 128.
 MITTELSTRAß J., 155.
 MONOD J., 117, 119, 121.
 MORMANN T., 159, 162.
 NAGEL E., 12, 119, 120.
 NEWTON I., 104, 166, 172.
 OLSHAUSEN B., 141.
 PAILLARD J., 113.
 PASK G., 143.
 PETERSEN S.E., 132, 133.
 PITTS W., 33, 122.
 PLATTEN J., 91.
 POIRIER P., 2.
 POSNER M.I., 132.
 PRIGOGINE I., 120.
 PUTNAM H., 3, 5, 8, 12, 13, 215, 216.
 QUINE W.V.O., 158, 166.
 QUINN R., 51, 79.
 RAICHLE M.E., 132.
 RAMACHANDRAN V.S., 61, 65, 67, 68, 69, 111, 134, 135, 136, 141, 142, 181, 182, 183.

- RAMSEY W., 122.
 ROCK I., 43.
 ROMERO R.D., 133.
 RORTY R., 12, 13, 14.
 ROSCH E., 27, 80, 148, 183, 213.
 ROSENBAUM C., 36.
 ROSENBERG A., 5, 114, 115, 118.
 RUMELHART D., 85, 88.
 RYLE G., 12.
 SCHELLING F.W.J., 206.
 SCHIEBER M., 138, 141.
 SCHRÖDINGER E., 159.
 SCHÜTZ A., 207.
 SCHWARZ G., 21.
 SEARLE J., 12, 106.
 SÉGAL L., 143, 144, 146.
 SEJNOWSKI T.J., 36, 61, 111, 135.
 SERON X., 109, 113, 114, 129.
 SERVAN-SCHREIBER D., 133.
 SHALLICE T., 129.
 SMITH L., 72, 82, 83.
 SMITHERS T., 84.
 SMOLENSKY P., 85.
 SNEED J., 161.
 SOULEZ A., 156.
 SPERRY R.W., 145.
 STEELS L., 93.
 STENGERS I., 120.
 STRUCHTRUP H., 92.
 STUFFLEBEAM R.S., 132.
 SUPPE F., 153, 154, 155, 157, 161,
 169, 171, 177, 179, 180.
 SUPPES P., 161.
 TARSKI A., 168, 169.
 THELEN E., 72, 82, 83, 214.
 THOMPSON E., 27, 80, 148, 183, 213.
 TOULMIN S., 159.
 TRIANTAFYLLOU G., 75.
 TRIANTAFYLLOU M., 75.
 UEXKÜLL J.V., 63, 64.
 ULLMAN S., 68, 69.
 VAN DE VIJVER G., 18, 118.
 VAN ESSEN D., 140, 141, 142.
 VAN FRAASSEN B., 154, 155, 157,
 161, 169.
 VARELA F., 1, 18, 27, 80, 143, 145,
 146, 147, 148, 183, 213, 214.
 VAUCLAIR J., 148.
 VERLEYSEN M., 26, 34, 35, 40.
 VOGEL S., 76.
 VON FOERSTER H., 143, 144, 145,
 146, 147, 176.
 VON NEUMANN J., 24.
 VYGOTSKI L., 86, 87.
 WERNICKE C., 128, 130.
 WIENER N., 77, 176.
 WIESEL T.H., 102.
 WINOGRAD T., 25.
 WITTGENSTEIN L., 7, 216.
 WUNDT W., 21, 44.
 ZAWIDZKI T., 112.

Table des matières

Avant-propos	1
1. Première perspective : l'action et le contexte dans les sciences cognitives	1
2. Deuxième perspective : la théorie de l'action	4
2.1. La réversibilité entre l'action et le contexte	4
2.2. Limitations internes et limitations contextuelles de la modélisation.....	7
3. Perspectives de recherche	9
Introduction.....	11
1. La problématique des limitations de la naturalisation de l'intentionnalité en philosophie de l'esprit	12
2. L'approche des limitations à partir de la problématique de la modélisation	14
3. Aperçu de la structure de l'ouvrage.....	18
Chapitre 1 : Du computationnalisme aux modèles écologiques de la cognition.....	21
1. Introduction.....	21
2. Le modèle de l'ordinateur	23
3. De l'intelligence habituelle à l'intelligence productive	28
3.1. Les modèles connexionnistes : des mémoires associatives à l'auto-organisation.....	29
3.1.1. Les mémoires associatives	29
3.1.1.1. Un réseau de reconnaissance de visages.....	30
3.1.1.2. Le neurone binaire et l'apprentissage associatif	33
3.1.1.3. Des propriétés inattendues	35
3.1.2. L'apprentissage par auto-organisation.....	39
3.1.3. L'intelligence habituelle	42

3.2. L'extension des modèles connexionnistes.....	44
3.2.1. Le corps simplement habituel : le cas Schneider	45
3.2.2. L'adaptation contextuelle et l'activité en situation	50
3.2.2.1. La transformation de nos habitudes.....	50
3.2.2.2. La négociation active en situation	53
3.3. Conclusion :	
le connexionnisme et les formes d'intelligence.....	56
4. Au-delà des mécanismes internes : l'écologie de la pensée	59
4.1. L'écologie subjective	62
4.1.1. Les représentations partielles	64
4.1.2. Les représentations pratico-sociales.....	69
4.2. La détermination écologique	74
4.2.1. Limiter la charge représentationnelle :	
le couplage dynamique.....	77
4.2.1.1. Les modèles cybernétiques.....	77
4.2.1.2. Les systèmes décentralisés sous contrôle	
de l'environnement	79
4.2.1.3. Les boucles de rétroaction élargies.....	82
4.2.2. La structuration active de l'environnement :	
les dispositifs.....	85
4.2.2.1. Les dispositifs.....	85
4.2.2.2. La conscience réflexive	88
4.2.3. Les systèmes émergents	90
4.2.3.1. L'émergence directe et l'émergence indirecte.....	91
4.2.3.2. Les variables globales.....	94
Chapitre 2 : Les conséquences méthodologiques.	
Du fonctionnalisme à la modélisation contextuelle	99
1. Introduction.....	99
2. Le défi de la cognition : modéliser les computations.....	101

2.1. La théorie computationnelle de David Marr et ses critiques	101
2.1.1. La computation comme traitement d'information	102
2.1.2. Les critiques de l'approche computationnelle	105
2.1.3. Le problème du fonctionnalisme	110
2.2. L'articulation entre la computation et l'implémentation	112
2.2.1. L'explication fonctionnelle en biologie	113
2.2.2. Le concept de système téléonomique	117
2.2.3. L'émergence des capacités de traitement d'information à partir du cerveau	121
3. Le défi de la modélisation contextuelle : modéliser les représentations et les fonctions	125
3.1. L'interaction entre les représentations et la critique fonctionnaliste	125
3.1.1. La décomposition modulaire dans les études structure - fonction	127
3.1.1.1. La méthode de la dissociation en neuropsychologie clinique	127
3.1.1.2. La méthode de la soustraction en imagerie cérébrale	131
3.1.2. Les modules interactifs	133
3.1.2.1. L'interaction des modules dans le système visuel	134
3.1.2.2. Les solutions globales de coordination du mouvement	136
3.2. Le couplage fonctionnel et la critique par les théories de la clôture opérationnelle	142
3.2.1. Définir l'hypothèse de la clôture opérationnelle	143
3.2.2. La clôture dans les systèmes cognitifs	145
Chapitre 3 : Limitations internes et limitations contextuelles de la modélisation de l'intentionnalité	151
1. Introduction	151
2. La critique du traitement standard de la limitation	155

2.1. Du modèle standard à la conception sémantique.....	155
2.2. La prise en compte de la dimension sémantique des modèles.....	161
2.2.1. Les conditions d'élaboration des modèles particuliers ...	163
2.2.2. Les conditions de la productivité sémantique des formalismes.....	165
3. Du formalisme aux modèles particuliers	170
3.1. Les limitations internes de la productivité sémantique du formalisme en sciences cognitives.....	172
3.2. Modèles et interprétation dans la modélisation de l'intentionnalité.....	177
3.2.1. Le caractère contextuel de la modélisation de l'intentionnalité	178
3.2.2. Le rôle de l'interprétation	180
4. Pour un traitement réflexif de la contextualisation : la solution phénoménologique et sa radicalisation	184
4.1. Le dépassement de la question des limitations dans la phénoménologie génétique	186
4.1.1. L'horizon eidétique de la logique formelle moderne.....	188
4.1.1.1. La manifestation effective du sens : les orientations fondamentales de la logique formelle moderne	188
4.1.1.2. Les présuppositions idéalisantes de la logique formelle moderne.....	190
4.1.2. La réduction phénoménologique transcendantale	193
4.1.2.1. De l'horizon de la logique formelle à l'horizon du monde de l'expérience	193
4.1.2.2. L'expérience perceptive comme donation des choses dans leur être même	195

4.2. La structure d'éveil de la conscience comme forme de vie originaire.....	198
4.2.1. Les synthèses passives dans le flux de la conscience.....	199
4.2.2. La vie originaire comme vie en éveil dans le présent vivant du temps	201
4.3. L'habitus pré-réflexif et la radicalisation de la méthode phénoménologique.....	203
4.3.1. L'ouverture à la sphère de déterminabilité de la conscience	204
4.3.2. Une théorie radicale de la connaissance.....	208
Conclusion	211
1. Le premier axe : l'opération de modélisation de l'intentionnalité dans les sciences cognitives	211
2. Le deuxième axe : les limitations contextuelles de l'opération de modélisation	213
3. Contribution de la théorie écologique à l'hypothèse de la réversibilité asymétrique	214
Bibliographie	219
Index onomastique	229
Table des matières.....	233