L'intentionnalité et les modèles artificiels
Émergence et réalisation : deux côtés de la même pièce

THÈSE

présentée à la Faculté des sciences, pour obtenir
le grade de Docteur ès sciences, par

Miriam Edith Scaglione

UNIVERSITÉ DE NEUCHÂTEL
Institut d'Informatique et
d'Intelligence Artificielle
rue Émile-Argand 11
2007 Neuchâtel
Suisse

le 20 septembre 1996

IMPRIMATUR POUR LA THÈSE

L'intentionalité et les modèles artificiels émergence et réalisation : deux côtés de la même pièce

de Mme Miriam Scaglione

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel sur le rapport des membres du jury,

Messieurs J.-P. Müller, P.-J. Erard, D. Miéville, J.-M. Besnier (Paris) et M. Imbert (Toulouse)

autorise l'impression de la présente thèse.

Neuchâtel, le 9octobre 1996

Le doyen:

R. Dändliker

R. Dändliker

À mon père Juan Carlos Scaglione et en memoire de ma mère Lidia Santamaria à qui je dois tout.

On raconte qu'un jour un moine salésien et un jésuite vinrent consulter le souverain pontife à Rome pour soumettre à son infaillible jugement une brûlante mais fumeuse question.
"Très Saint Père, puis-je fumer quand je prie? " s'enquit le salésien. La réponse du pape fut prompte et sans équivoque; il n'était pas question d'autoriser la pratique d'un vice durant l'oraison car c'eût été rien moins qu'un sacrilège
Tout autre semblait le souci du disciple de Saint Ignace.
"Est-il permis de prier pendant qu'on fume? " demanda-t-il. "Le Seigneur ne peut que trouver agréable qu'on se tourne vers lui avec piété, quelle que soit l'activité du moment" répondit le vicaire du Christ.

Remerciements

Lorsque j'était petite pendant que tous mes camarades de l'école maternelle de Buenos Aires pensaient à jouer, j'étais obsedée par une question: Comment fabrique-t-on un robot qui soit un clône des humoins? Evidemment, je ne posais pas la question ainsi, je voulais seulement construir pour de vrai *mon héro télévisuel nommé Astraboy. Par la suite, j'ai fait de l'informatique, de l'IA et des sciences cognitives et trente années après j'oi reposé la question d'une manière plus savante. Dane la réalisation de ce réue d'enfance beaucoup m'ont aidée. J'aimerais exprimer ma gratitude en quelques lignes, ce qui est bien peu à côté de ce que j'ai reçu.*

Je remercie les institutions qui ont financé mes études. Par ordre chronologique, le IGSC World Laboratory, la Fondation Tissot pour la pramotion de l'économie, l'Institut d'Informatique et Intelligence Artificielle de l'Université de Neuchâtel qui m'o permis de faire aussi l'expérience de l'enseignement, le Fonds Nationol de la Hecherche Scientifique suisse qui m'a donné une bourse pour assister au First International Summer School of Cognitives Sciences à Buffalo aux Etats Unis. L'année passée j'ai eu l'immense satisfaction de passer six mois comme chargée de recherche à l'Ecole Polytechnique en France et de travailler dans l'unité de psychiatrie de l'Hôtel-Dieu auprès du grand specialiste qu'est le docteur Henri Crivois. Je me suis initiée à d'autres domaines de la cognition comme la psychologie pathologique. Je ueux exprimer toute ma gratitude aux Professeurs Henri Grivois et Joëlle Proust, responsables de cette aventure. J'ai aussi été invitée à l'unité de recherche "Cerveau et Cognition" à la Faculté de Medecine de Rangueil de Toulouse et tiens à remercier son directeur le Prof. Michel Imbert, qui m'a toujours encouragée, ainsi que tous les checheurs et collaborateurs pour leur accueil et pour les discussions auxquelles j'oi pu participer.

Il y a aussi des personnes qui m'ont aidé et soutenue eans rêlache. Je pense au premier chef au docteur Pierre Rossel et à toute sa famille, à tous mes collègues de l'Institut d'Informatique de Neuchâtel, particulièrement Miguel Hodriguez, Françoís Sprumont, Pontien Déguénon, Philippe Blache, Olivier von Doch, Sylvain Nogues et à la secrétaire Mme Gianfranca Cerrito. Les professeurs Pierre-Jean Erord et Jacques Savoy que je remercie tout spécialement de cette année que j'ai passée comme leur assistante, monsieur Hans Nägeli et monsieur Denis Miéville.

Je remercie aussi mes collègues de l'IREC à l'EPFL, surtout mes amis Martin Schuler, Jean-Claude Bolay, Anne Danton et Thèrese Huissoud; aussi tous mes amis en Europe qui ont suivi soit la totalité, soit une partie de mon parcours: Gerard et Lucienne Chevallier, Alain et Marie-Claude Garnier, M. et Mme. Staudenmann, Jocelyne Tissot-Daguette, Victor Tribelli, M. et Mme. Amaudruz, Luc Tissot, Côme Carpentier de Gourdon, Ines Ortega, Jean et Françoise Erceau.

Les chereheurs du CREA à l'Ecole Polytechnique de Fronce pour ses indications bibliogrophiques et ses conseils: Elisabeth Pacherie, Roberto Casati, Pierre Jacob, Pascal Engel, Paul Bourgine, Jean Petitot, Adriano Palma, Françoís Recanati et les professeurs George Rey, Paul Smolensky et John Searle. Je dois une grotitude particulière aussi au Professeur Jean-Michel Besnier pour m'avoir suggeré des améliorations à la première lecture du manuscrit, de même qu'au Professeur Denis Miéville. Je remercie également le Prof. Barry Smith de l'Université de New York à Buffalo qui a été porticulièrement généreux dans ces conseils bibliographiques et philosophiques.

Les professeurs et collègues du Séminaire de Philosophie de l'Université de Neuchâtel, Richard Glauser, Daniel Schultheas et plus encore Christine Tappolet pour nos discussions de presque tous les soirs et François Girard. Toutes les imprécisions et incorrections que peut contenir ce travail ne viennent que de ma maladresse et ne sont en aucun cos le reflet de ces conseils.

Je remercie ma famille et mes amis de Buenos Aires qui m'ont toujours encourogée. Tout d'abord mon père. J'ai une pensée très émue pour ma mère et ma grande mère qui nous ont quitté avant que je finisse ce travail. La famille Creco-Cillis, la famille Faggi, la famille Wiszniacki, Fernando y Ramón Gonzalez, Maria del Carmen Tisi, Ines, ainsi que le profeseeur Hernan Santiago Nottoli et le professeur Enrique Rabossi et la Sociedad Argentina de Anâlisis Filosofico.

Enfin, je suis infiniment reconnaissante au Professeur Jean-Pierre Müller qui a eu le courage d'accepter de diriger une thèse interdisciplinaire comme celle-ci, pour la liberté de recherche qu'il m'a donnée et pour les discussions motivantes que nous avons eues toutes ces années.

Avant propos

Le présent travail concerne le problème de la naturalisation de l'intentionnalité, c'est à dire l'explication en termes non–mentaux des phénomènes psychiques d'un système physique.

Mon objectif est de justifier les deux propositions suivantes. Premièrement, les concepts de réalisation et d'émergence ne sont pas contradictoires mais inverses. Je vais démontrer qu'une propriété est émergeote des propriétés d'un niveau inférieur si elle est réalisée par ces dernières. Secondo, je montrerai les limites de la définition standard du concept de multiréalisation pour conclure à l'aide de la première proposition qu'il existe une possibilité d'échapper à ces critiques lorsqu'on se pose le problème de la caractérisation des états physiques d'une manière pertinente. Les deux propositions vont justifier la nécessité de tenir plus grand compte des résultats des neurosciences de préférence aux modèles artificiels.

Ce travail est composé de trois parties. La première partie pose le problème corps–esprit. C'est aussi une "boîte à outils" pour des concepts qui deviendront centraux dans la suite du texte.

La deuxième partie présente et critique les théories cognitives suivantes: le monisme anomal de Donald Davidson et les théories fonctionnalistes, c'est à dire ces théories qui tout en admettant l'existence de corrélations entre états mentaux et états physiques (identité des types ou identité occasionnelle), cherchent à décrire les états mentaux en termes des leur rôle fonctionnel en faissant l'économie d'une description précise du substrat nerveux. Je vais conclure qu'elles n'arrivent pas à donner une solution satisfaisante de ce problème.

La troisième partie a un double objectif. En premier lieu il s'agit d'une analyse historique du concept d'émergeoce dans le but de le dégager de toutes les caractéristiques qui le rendent incompatible avec une démarche scientifique. Ensuite, tout en présentant et en critiquant l'élan phénoménologique des sciences cognitives j'y tente une description historique des modèles de morphodynamique pour introduire les concepts d'invariants dynamiques nécessaires à la justification de ma conclusion.

J'explique en détail comment une méthode qui tient compte de la dynamique du substrat nerveux peut néanmoins rester non–réductionniste. Cela a comme conséquence directe le repositionnement des neurosciences au centre de la scène en lieu et place des modèles artificiels. Il s'agit en fait d'une démarche *qui veut replacer l'église au milieu du village.*

Table des matières

Partie I

Introduction

Chapitre 1

L'Origine des Sciences cognitives

The question which troubles laymen, and which has long troubled philosophers, even if it is somewhat disguised by today's analytic style of writing philosophy, is this: are we made of matter or soul-stuff? To put it as bluntly as possible, are we just material beings, or are we 'something more'?

Putman, Hilary *Philosophy and Our Mental Life*

1.1 Introduction

Le concept d'intentionnalité, c'est-à-dire la relation entre l'esprit et les objets du monde extérieur, a été et demeure le carrefour des sciences cognitives pendant ces cinquante dernières années. Pour ces sciences *connaître* équivaut à produire un modèle du phénomène et à effectuer sur lui les manipulations réglées [Dupuy, 1994, cf. page 93]. Ce postulat constitue à la fois une règle pour la démarche scientifique et une description de l'activité de tout système cognitif. Mais si la connaissance implique une modélisation, cette dernière nécessite un substrat ou un soutien qui nous permette d'opérer ces transformations réglées, ce qui revient à dire que l'on a besoin d'un système représentationnel.

Il existe donc un parallélisme entre la démarche des sciences cognitives et leur objet : le(s) système(s) cognitif(s) est (sont) bâti(s) sur le concept de renvoi. Les sciences cognitives –comme toutes les sciences– sont renvoyées à leur objet d'étude et celui-ci se rapporte aussi au monde.

Ce même parallélisme fait du concept d'intentionnalité la pierre d'angle de la recherche des sciences cognitives. Bien que nées d'une réflexion interdisciplinaire d'un groupe de sciences, les sciences cognitives possèdent des caractéristiques communes : l'effort épistémologique et la quête constante d'un modèle de cognition. Cette interrogation épistémologique constante due à la nouveauté de ces sciences en formation explique la grande influence de la tradition analytique anglo-saxonne avec son bagage de rigueur et son exigence de clarification de la philosophie des sciences. Je vais montrer que la tradition analytique est peuplée des efforts qui tendent à compatibiliser le physicalisme avec une théorie de l'intentionnalité et que cette interrogation est centrale dans la philosophie de l'esprit.

Les questions qui se posent dans ce domaine se réfèrent à l'étude de phénomènes mentaux, à la nature des facultés de l'esprit et à l'intuition et au raisonnement entre autres. Comment est-il possible que des événements purement mentaux puissent avoir un rôle causal dans nos actions étant donné que ces dernières sont des événements physiques? Autrement dit, est-il possible que des

événements de l'esprit, donc non physiques ou immatériels, déclenchent des phénomènes physiques (c'est-à-dire des actions) sur notre corps, sans être soumis aux lois de la physique (comme par exemple le principe de conservation de la matière), de la chimie ou de la biologie?

Jusqu'à l'arrivée des sciences cognitives sur le devant de la scène, la philosophie a tenu le rôle central dans cette recherche. Après la Seconde Guerre mondiale, l'ensemble des sciences qui composent les sciences cognitives (notamment la linguistique, les neurosciences, l'intelligence artificielle et la psychologie entre autres) se sont alignées sur la philosophie de l'esprit. Ainsi, la philosophie, et en particulier la philosophie de l'esprit, a tranformé son rôle puisqu'elle a abandonné son statut de fondatrice pour remplir une mission à mon avis plus importante: elle est devenue l'arbitre de controverses qui s'éveillent à l'intérieur de ces sciences. Jean-Pierre Dupuy va même plus loin:

> Les sciences cognitives se présentent volontiers comme la reprise à nouveaux frais par la science des questions philosophiques les plus anciennes concernant l'esprit humain, son organisation, sa nature, les relations qu'il entretient avec l'organisme (le cerveau) avec autrui et avec le monde. Mais l'identité de ce qui se donne pour science de l'esprit reste profondément philosophique. Cette science qui parle au nom des sciences et des techniques qui composent le domaine (encore une fois, principalement les neurosciences, l'intelligence artificielle, la psychologie dite cognitive et la linguistique), et auxquelles elle apporte ce supplément d'âme (ou d'esprit) qui les réunit les unes aux autres, est en réalité une philosophie. C'est la philosophie qui se glisse à l'intérieur du cheval de Troie des sciences et des techniques pour investir le domaine de l'esprit, et en chasser les intrus qui occupaient encore la place: d'autres philosophies – principalement les philosophies de la conscience, la phénoménologie, l'existentialisme–, d'autres psychologies – comme le behaviorisme et la psychanalyse –, d'autres sciences – singulièrement les science sociales et les sciences de l'homme de type structuraliste.[Dupuy, 1994, page 93]

1.2 La philosophie analytique et la philosophie de l'esprit

La philosophie de l'esprit, qui est l'animatrice et l'arbitre des sciences cognitives, a ses racines dans la philosophie analytique. Mais qu'est-ce que la philosophie analytique?

Le début du XXè siècle est caractérisé par une rupture entre la philosophie et la psychologie. La philosophie en vogue en Allemagne vers la moitié du XIXè siècle est l'idéalisme hégélien et le romantisme. En réaction à l'hégélianisme dominant, Jakob Friedrich Fries (1798–1844) et Friedrich Eduard Beneke (1798–1854) prônaient une position philosophique connue sous le terme de psychologisme, entièrement basée sur la psychologie. Ils soutenaient que l'unique outil valable pour la quête philosophique est l'introspection et qu'on ne peut déterminer la vérité qu'en utilisant des éléments subjectifs issus de celle-là.

Vers la moitié du XIXè siècle, John Stuart Mill (1806–1873) préconise l'introspection comme unique base des axiomes mathématiques et des principes de la logique. Dans son texte *Examination of Sir William Hamilton's Philosophy*, la logique est considérée comme une branche de la psychologie. Gottlob Frege (1848–1925) a soutenu quelques années plus tard la position opposée à celle de Mill en relation aux mathématiques. Il soutient avec d'autres tel Rudolf H. Lotze (1817–1881) quelques années auparavant qu'il faut faire la différence entre l'acte psychologique de penser et le contenu de cet acte. Le premier existe seulement comme un phénomène temporellement déterminé, tandis que la caractéristique de l'être du second est sa validité.

Edmund Husserl (1859–1938) fait aussi une critique systématique du psychologisme dans les champs des mathématiques et de la logique. Il soutient que, si les lois logiques étaient basées sur les lois de la psychologie, alors elles seraient très vagues et n'auraient pas la validité qu'elles revendiquent, étant alors basées sur l'induction à la façon des lois empiriques.

Or, Gottlob Frege et Edmund Husserl, deux fondateurs de la philosophie de notre siècle, ont récusé le psychologisme dans le cadre des recherches en logique et mathématique. Cette récusation de la psychologie dans le cadre de la logique a aussi enchaîné la réfutation de cette discipline dans d'autres domaines de la philosophie. Ceci fait dire à Pascal Engel que la philosophie de l'esprit, comme la partie de la philosophie qui "analyse des concepts mentaux, c'est-à-dire comme une enquête conceptuelle", a pris un faux départ [Engel, 1994b, page 7]. Le caractère mystérieux qu'elle attribue aux phénomènes mentaux parce qu'ils sont fondamentalement subjectifs lui fait croire que toute philosophie basée sur eux est destinée à l'échec.

La tradition analytique[1] née à partir de Frege, Bertrand Russell (1872–1970) et George Eduard Moore (1889–1951) connaît deux grandes écoles inspirées toutes les deux par Ludwig Wittgenstein. D'un côté le *Tractatus* qui a influencé le *Cercle de Vienne*[2] qui est à l'origine du Positivisme logique ou Néopositivisme. De l'autre, les *Investigations Philosophiques* qui est à l'origine de la tradition analytique de la philosophie du langage avec deux grandes tendances, celle d'Oxford et celle de Cambridge. La première tendance a comme représentants Gilbert Ryle (1900–1982), John Langshaw Austin (1911–1960) et Peter Frederick Strawson (1919–), et est inspirée des travaux de G.E. Moore (orienté vers l'étude du langage ordinaire). La seconde a comme représentants John Wisdom (1904–1993), qui est un élève de Wittgenstein et lui succède en 1952 dans sa chaire à Cambridge, Norman Malcom (1911–1990) et G. A. Paul. Wisdom soutient que la démarche philosophique consiste à donner un éclairage (*illumination*) de la structure ultime des faits. Il se différencie de son maître par son attitude envers les théories philosophiques. Pour Wittgenstein, elle n'étaient qu'un symptôme de confusion linguistique, tandis que Wisdom soutient que la démarche philosophique a toujours fait appel à l'usage linguistique, et tente d'établir les similarités entre la philosophie linguistique et les formes de spéculation faites auparavant. Certains de ses textes sur l'analogie entre la philosophie et la psychanalyse ont fait croire qu'il pourrait considérer la philosophie comme une thérapie des faits du langage. Mais cette interprétation a paru excessive à plusieurs de ses critiques, et il ne faut certainement considérer sa démarche que dans le cadre de l'effort de compréhension du travail philosophique dans sa quête de la structure ultime des choses.

1.3 L'empirisme logique et le néo–positivisme

Bien que l'empirisme logique et le néo-positivisme ou positivisme logique soient parfois considérés comme synonymes, il existe des différences. L'empirisme logique trouve son origine dans les travaux de Russell et Moore, qui prônaient une certaine continuité entre le sens commun et les sciences tout en préconisant l'expérience comme base de toute argumentation scientifique. Le néo–positivisme inspiré par le Cercle de Vienne essaie de rapprocher les sciences de la philosophie tout en exorcisant des faux problèmes posés par la métaphysique. Le nom de positivisme est dû au fait que cette doctrine considère que la science est l'unique forme de connaissance et qu'il n'existe aucune chose dans l'univers qui ne puisse, en principe, être scientifiquement connue. Alors, les questions métaphysiques sont récusées comme ne répondant pas aux méthodes scientifiques, étant donné leur caractère transcendantal. Dans la même ligne d'argumentation, les énoncés formulés à la première personne, typiques de la méthode introspective en vogue dans la philosophie et la psychologie à cette époque, sont aussi récusés, étant donné qu'ils ne peuvent pas être soumis à l'observation publique. Ainsi, les positivistes logiques ramènent la philosophie, qui était jusqu'alors

[1] Pourquoi la tradition de Frege, Russell et Moore est-elle qualifiée d'analytique? Parce que Russell et Moore ont brandi l'analyse contre deux prémisses du monde de pensée néo-hégélien, qu'ils avaient fini par abhorrer: l'idée que la réalité authentique est toujours formée de totalités organiques (à moins qu'elle ne forme qu'un gigantesque tout); et l'idée que l'abstraction est une falsification ou que la décomposition d'une totalité organique est toujours une abstraction illégitime (cf. Russell, 1900, chapitre 9). En 1903, Russell et Moore dirigent leurs coups contre ces deux propositions. Selon Russell [Russell, 1903, page 133], " à moins qu'elles ne servent qu'à déguiser la paresse, en apportant une excuse à ceux qui n'aiment que les labeurs de l'analyse, elles ne peuvent avoir d'autre sens que le suivant: bien que l'analyse nous procure la vérité, elle ne peut jamais nous la procurer toute entière. Désormais, Russell et Moore ne se départiront plus, contre leurs adversaires néo-hégéliens, de ce style d'argumants: ou bien ce que vous dites est vrai mais trivial, ou bien c'est faux. Or, à l'analyse, cette hantise de l'abstraction se révèle elle-même ou triviale ou fausse" [Jacob, 1980, pages 34–35].

[2] Le cercle de Vienne (*Wiener Kreis*) s'est formé autour de Moritz Schlick en 1929. En 1922, Schlick prend la Chaire de Philosophie des Sciences inductives à la suite du physicien Ernest Mach, et, dans ce cadre, un groupe de savants se réunit. Il y dans ce groupe des mathématiciens comme Kurt Gödel, Gustav Bergman et Hans Hahn, des philosophes comme Rudolf Carnap et Victor Kraft, des physiciens comme Philipp Frank et aussi des représentants des sciences sociales comme Otto Neurath. En 1951, Carnap, Hahn et Neurath intitulent le manifeste du Cercle de Vienne: Conception scientifique du monde. "Ils y mentionnent l'influence de cinq domaines scientifiques sur le nouvel empirisme qu'ils défendent: le positivisme et l'empirisme plus anciens (de Comte et de Mach); l'étude des fondements, des buts et des méthodes des sciences empiriques; la logique et ses applications à la réalité; les axiomatiques; enfin, l'hédonisme et la sociologie positiviste. Par-dessus tout, ils placent leur entreprise sous l'égide de trois représentants de la conception scientifique du monde: Albert Einstein, Bertrand Russell et Ludwig Wittgenstein" [Jacob, 1980, pages 99].

considérée comme la reine des sciences, au niveau d'une science parmi les autres ou plus exactement à une métathéorie des sciences. Elle devient, comme l'a suggéré Rudolf Carnap (1891–1970) en 1934, une syntaxe logique du langage de la science.

L'analyse logique devient le principal outil dans la démarche d'abolition de la métaphysique. Si les énoncés métaphysiques semblent avoir de la signification, c'est en partie parce qu'il sont énoncés en langage naturel. Celui-ci étant ambigu, on peut lui imputer la responsabilité de ce mirage. Ainsi, toute une série d'activités s'organisent dans le but de formaliser la logique de façon à pouvoir traduire des énoncés du langage naturel en des énoncés logiques qui seront eux, exempts de toute ambiguïté.

En effet, le but était d'édifier un système de concepts tel que toute proposition se rapportant au monde puisse y être traduite par une proposition ne se rapportant qu'aux concepts du système. Russell et Frege créent la logique des propositions et l'analyse de la quantification, convaincus que ceci va aider à supprimer les ambiguïtés des langages naturels. L'analyse des énoncés réalisée sur la base des relations grammaticales est refusée, il faut d'abord traduire les énoncés en un langage logique, étant donné que la simple relation grammaticale sujet - prédicat n'est pas assez précise. En effet, soit l'énoncé suivant de la langue française: "Chaque homme aime une femme" . Cette phrase accepte deux interprétations possibles:

- (1) Chaque homme appartenant à un groupe aime l'une des femmes de l'autre groupe.
- (2) Il existe une femme unique que chaque homme aime.

L'utilisation de quantificateurs résoud le problème. En effet, dans le langage canonique du calcul de prédicat les deux interprétations sont traduites respectivement comme suit:

- $\forall x \exists y(x \text{ aime } y)$
- $\exists x \forall y(y \text{ aime } x)$

Russell développa deux paradigmes d'analyse: la *théorie des descriptions* et la *théorie simple des types* . La première théorie essaie de résoudre des problèmes de dénotation causés par des phrases comme "L'actuel Roi de France est chauve". En effet, ce type de phrases est dénué de référence et leur signification dépend du fait que ces entités ont un être[3] sans pour autant exister. Frege, qui prônait la distinction entre sens et référence, disait que la phrase avait un sens mais qu'elle était dénuée de référence. Comme pour Frege la référence d'une phrase est sa valeur de vérité, la phrase n'était ni vraie ni fausse. Cependant Russell, qui tenait au principe du tiers-exclus, propose une analyse basée non pas sur la relation sujet–prédicat mais sur la transcription de cette même proposition en termes de quantificateurs. L'idée sous-jacente est de réfléchir sur des ensembles: on ne trouve l'actuel Roi de France ni dans l'ensemble de choses chauves ni dans son complément. Donc, il y a deux propriétés qui représentent à la fois deux ensembles différents: l'une est d'être l'actuel Roi de France, l'autre d'être chauve. Russell propose de jouer avec les quantificateurs de la façon suivante:

$$\exists x \forall y[(\Phi(y) \equiv (y = x)) \wedge \Psi(x)] \tag{1.1}$$

Où $\Phi(x)$ veut dire x est l'actuel Roi de France et $\Psi(x)$ veut dire x est chauve.

Donc la proposition ci-dessus se lit:

Il existe un x tel que, pour tout y, y a la propriété $\Phi(y)$ si et seulement si y est égal à x et x a la propriété $\Psi(x)$.

Comme aucun élément ne satisfait l'opérateur existentiel, la proposition est fausse. La théorie simple des types a comme objectif de résoudre des paradoxes qui avaient été trouvés dans la théorie des ensembles. Tous ces paradoxes ont une caractéristique commune: les éléments d'une totalité sont définis dans les termes de la même totalité (par exemple, l'ensemble de tous les ensembles est

[3]Dans le chapitre concernant Brentano, je vais traiter la théorie de Meinong, qui faisait une différence entre *être* et *exister*.

un ensemble). C'est à cause de ce type de paradoxe que Russell écrit une lettre a Frege, à la suite de laquelle ce dernier ajoute un appendice à ses *Grundgesetze der Arithmetik*, en 1903.[4]

Dans le système fregéen original, chaque attribut F avait une classe comme extension, cette classe étant notée $\hat{x}(Fx)$ et la propriété étant

$$\hat{x}(\Phi x) = \hat{x}(\Psi x) \supset (x)(\Phi x \equiv \Psi x)$$

Cela voulait dire que si les classes déterminées par deux attributs sont identiques, les extensions déterminées par ces attributs sont les mêmes. Cependant, que se passe-t-il avec l'extension des classes déterminées par des attributs, c'est-à-dire par $\hat{x}(Fx))$ et $\hat{x}(yx))$? Autrement dit, est-ce que l'implication suivante sera vraie?

$$\hat{x}(\Phi x) = \hat{x}(\Psi x) \supset (\Phi(\hat{x}(\Phi x))x \equiv \Psi(\hat{x}(\Psi x)))$$

Bref, est-ce que l'extension du premier attribut $(\hat{x}(Fx))$ appartient à l'extension du second et vice versa? Et voilà comment nous nous trouvons dans les conditions mêmes du paradoxe de Russell, puisque, si cette implication était vraie, l'extension serait dès lors auto-référentielle. Cependant, tous les objets qui sont dans le premier attribut sont aussi dans le second, sauf l'extension $(\hat{x}(Fx))$ même. La théorie des types, proposée par Russell, et qui accepte l'existence d'un niveau de langage différent de celui auquel appartiennent l'intension et l'extension, permet d'éviter ce type de situations. Si Frege avait accepté cette solution, il aurait bâti une théorie analogue à la théorie de type russellien. Mais il a préféré seulement opposer l'exception suivante:

> If a is not the extension of a concept, then a falls under that concept if and only if it falls within its extension. [...] The extension of the concept does not fall within itself.[Dummet, 1954, page 103]

L'ambiguïté octroyée aux langages naturels a poussé Rudolf Carnap à bâtir une théorie syntaxique de la signification cognitive (par opposition à une signification émotive ou poétique), dans le cadre de son entreprise consistant à faire de la philosophie une logique de la science. Sur ces bases, il pensait pouvoir montrer que le mystère sur lequel reposent les mauvaises propositions métaphysiques (qu'on peut définir comme n'ayant pas de signification cognitive) est dû à la confusion des objets du mondes et des expressions se référant aux propriétés mêmes du langage. Dans la tradition wittgenstienne du *Tracatus*, une proposition a une signification cognitive si, d'abord, elle est bien formée, et ensuite elle a un contenu informationnel sur la réalité, ceci en opposition à deux autres types de propositions: premièrement les tautologies, qui n'ont aucun contenu informatif puisque n'étant pas des modèles de la réalité, et deuxièmement les phrases logiquement mal formées cf. [Wittgenstein, 1921, cf. Tr. 6.1 - 6.2 et Tr 4.461 - 4.4611].

Pour Carnap, il existait deux types d'énoncés métaphysiques: les bons et les mauvais. Les mauvais énoncés métaphysiques sont des phrases intrinsèquement absurdes parce qu'elles contiennent des erreurs logiques (cf. Tr. 4.003). Il pensait qu'il pouvait faire la part des choses en utilisant la théorie des types de Russell, tout en se besant seulement sur des critères syntaxiques. Ainsi, il pensait pouvoir démontrer que les bons énoncés métaphysiques, c'est-à-dire ceux qui ne sont pas intrinsèquement absurdes et qui ont la prétention de parler de la réalité, ne sont en fait que des recommandations linguistiques déguisées. Pour y arriver il a voulu

> ... produire un système de règles de formation des phrases à partir d'un alphabet de base et de règles de transformation (ou d'inférence) de théorèmes à partir des axiomes, grâce auquel on aurait pu déterminer le statut cognitif de n'importe quel énoncé. L'espoir étant de montrer d'une part que les pires énoncés de la métaphysique spéculative violent une de ces règles (par exemple, la théorie des types), et, d'autre part, que les thèses métaphysiques opposées [au positivisme logique], sur le fondement des mathématiquee, ou sur la théorie de la connaissance, étaient mal formulées. [Jacob, 1980, page 105])

Pour y parvenir, Carnap distingue deux types différents d'idiomes qui servent à exprimer les énoncés d'une science: l'idiome matériel qui correspond au langage-objet et l'idiome formel qui correspond au métalangage. Il existe trois types d'énoncés: les énoncés-objet , par exemple "cinq est un nombre premier"; les énoncés syntaxiques , par exemple "cinq n'est pas un terme d'objet,

[4] A ce propos, une intéressante discussion se trouve dans [Quine, 1955].

mais un terme numérique"; et une classe intermédiaire d'énoncés quasi-syntaxiques en mode matériel, à savoir des énoncés syntaxiques déguisés en énoncés–objets, comme par exemple "cinq n'est pas une chose, mais un numéro". Mais pour trouver les différences entre les deux idiomes, on doit fait appel au rapport entre les langages et des entités extra–linguistiques. Or il est faux qu'on puisse déterminer le statut d'énoncé cognitif sur des bases purement syntaxiques. Carnap ne le voyait pas avec la même clarté, parce que la frontière entre la sémantique et la syntaxique, vers 1934, n'était pas encore aussi distincte qu'aujourd'hui. La parution en 1935 de l'ouvrage d'Alfred Tarski *La Syntaxe logique du langage* sur la sémantique oblige Carnap à élargir le point de vue syntaxique et à tenir compte des concepts et des significations.

C'est seulement avec l'arrivée du la linguistique transformationnelle de Noam Chomsky (1928–) qu'on a pu faire une différence tranchée. Ce courant a montré qu'il existe des règles de transformation, par exemple sur les structures de surface qui ne sont en rien solidaires des concepts sémantiques comme la référence ou la vérité.[5] Mais la démarche de Chomsky consiste en l'étude des langage naturels, à la différence des années trente, où les logiciens pensaient que ces derniers étaient dénués de règles et n'étaient intéressés qu'à dissiper leurs ambiguïtés et imprécisions. Cependant, le *concept de contenu cognitif* tel qu'il a été compris par le Cercle de Vienne excède celui du *Tractatus* même. En effet, si une proposition, pour être cognitivement signifiante, doit avoir un contenu informationnel de la réalité, et si toutes ces propositions sont réductibles à d'autres, rapportant elles des perceptions directes ou des résultats immédiats de l'expérience, alors l'interprétation donnée par les membres du Cercle de Vienne de l'aphorisme suivant du *Tractatus* n'a rien de surprenant: *Comprendre une proposition, c'est savoir ce qui arrive, quand elle est vraie* (Tr. 4.024.); ce qui a été interprété ainsi: *La signification d'un énoncé c'est sa méthode de signification.* Cette proposition est à la base de la théorie connue comme la théorie vérificationniste de la signification, mais elle trahit l'esprit du *Tractatus*. Si la base de la vérification est l'expérience, il faut trouver un langage scientifique qui soit approprié. Ce langage devra permettra non seulement d'arriver à un accord intersubjectif, mais devra encore être un langage qualitatif qui porte sur les choses. La vérification intersubjective est devenue le critère de la preuve scientifique, mais ceci demande que le langage utilisé relève uniquement des choses ou événements publiquement observables. Deux types différents de termes étaient possibles pour bâtir ce langage basé sur des données publiquement observables: d'une part ceux qui se rapportent exclusivement aux objets physiques, d'autre part ceux qui se réfèrent aux données sensorielles ou *sense data*.

Le terme de *sense data* a été introduit par Russell et Moore et désigne les objets que l'on perçoit (*objects of perceptual awareness*) comme la couleur de patches et des formes qui sont différentes des surfaces des objets physiques mêmes. Les qualités des *sense–data* sont supposées être différentes des objets physiques mêmes parce que leur perception dépend des conditions du champ visuel dans lequel l'objet se trouve. Les *sense data* change dans la mesure où le champ perceptif change, tandis que l'objet physique d'origine reste constant. Carnap, dans son livre *La Construction logique du monde*, soutient un programme d'un second type puisqu'il est d'inspiration phénoménologiste[6]. Il prend comme modèle la réduction logiciste faite par Russell et Frege de la théorie élémentaire des nombres à la théorie logique. Il est persuadé de pouvoir bâtir un système constructif basé sur des unités non analysables, ces dernières étant semblables au concept proposé par la phénoménologie de Husserl et la *Gestalttheorie* . Il existerait donc des termes primitifs désignant des tranches d'expériences élémentaires ou des sensations formant une totalité. Mais ces termes primitifs basés sur des *sense–data* posent des problèmes au concept de vérification intersubjective. Si ce qu'on perçoit ne sont pas les objets physiques mêmes mais leurs apparences dans le champ visuel, la reconstruction des qualités réelles de ces objets est le fruit d'une reconstruction mentale et privée de celui qui perçoit. Dans ce cas, le principe de vérification intersubjective ne peut demeurer valable. Dans ce sens, les objections du Cercle de Vienne, notamment celles dues à Otto Neurath (1882–1945), ne vont pas tarder.

Si les énoncés décrivant individuellement les sensations pures bénéficient de la confiance indéfectible que chaque locuteur du langage phénoménaliste accorde à ses sensations, le phénomène d'inter-

[5] J'ai déjà discuté ce sujet dans [Scaglione, 1989].

[6] Une approche phénoménologique soutient que la perception ne se fait pas dans le cadre d'une relation directe entre le percevant et la chose perçue, mais au moyen d'une autre entité: le *sense–data*.

> **subjectivité serait un miracle inexplicable. Si l'intersubjectivité est au contraire un fait observable, alors il s'explique simplement par le fait que tout locuteur psychologiquement réel parle un langage dont les termes désignent non des événements ou des qualités sensoriels mais des objets physiques. Les langues naturelles possèdent effectivement des mots, qui mentionnent des objets sensoriels purs de toute interprétation [...] Neurath en conclut qu'un langage physicaliste, composé de mots désignant des objets physiques et leurs propriétés, est plus réaliste, d'un point de vue psychologique et épistémologique: désormais, les phrases de base d'un langage physicaliste s'appelleront de phrases ou énoncés protocolaires. Ce sont des phrases admises par le langage physicaliste, et non construits. [Jacob, 1980, page 121]**

Carnap s'incline face aux critiques et accepte d'adopter un langage physicaliste, langage que préféraient aussi les positivistes logiques sur un autre front: leur opposition au dualisme existant entre les sciences de l'esprit (*Geisteswissenschaften*) et les sciences de la nature (*Naturwissenschaften*) dans les universités allemandes dans ce temps. Ce combat pour l'unification des sciences au domaine purement physique est connu sous le nom de physicalisme. La langage physicaliste se présente comme le langage unificateur de toutes les sciences, biologie et psychologie comprises; ceci aura d'ailleurs en psychologie une importance centrale puisque, jusqu'à l'arrivée du positivisme logique, celle–là étudiait les problèmes de la conscience, de l'esprit, en acceptant les éooncés à la première personne. Or, avec le positivisme, cela ne sera plus considéré comme en ayant une valeur scientifique. En effet, les énoncés basés sur des rapports à la première personne, n'étant pas publiquement observables, ne respectent pas la règle de validation intersubjective; ces énoncés sont donc considérés comme dépourvus de toute valeur scientifique. La perspective de la première personne est donc abandonnée au bénéfice de la troisième personne. Les conditions de vérification des attributions mentales seront donc basées sur la conduite des agents, étant donné que celle–ci est publique et observable. Le behaviorisme est la théorie du mental qui a adopté ce paradigme.

1.4 Physicalisme versus dualisme

Le réductionnisme et/ou l'unité des sciences Le dualisme cartésien est la théorie philosophique qui accepte l'existence de deux substances: celle matérielle ou physique et celle immatérielle ou de l'esprit. Cette version du dualisme est la version métaphysique. Il existe aussi une autre version, le dualisme méthodologique ou linguistique qui soutient que les événements mentaux (que l'on peut éventuellement considérer comme étant physiques) sont très complexes, que l'observateur n'a pas le détachement nécessaire pour les étudier, et qu'en conséquence des termes n'appartenant pas au domaine purement physique sont nécessaires pour leur description.

En opposition à ceci se trouve la thèse du physicalisme qui est une version moderne du matérialisme traditionnel et qui affirme que tout ce qui existe est matière en mouvement. Alors, tous les processus, événements et états des choses peuvent être décrits par les sciences physiques. On associe au physicalisme deux thèses très souvent confondues (cfr [Pacherie, 1993]). La première est la thèse de la généralité de la physique et la seconde, propulsée par les positivistes et qui est une conséquence de la première: l'unité de la science. Selon la première, tous les événements physiques tombent sous les lois de la physique, selon la seconde, les autres sciences doivent être réductibles à des théories physiques. Les empiristes logiques élaborent deux types de théories réductionnistes ou métathéories. La *théorie de l'explication déductive–nomologique* est due à Carl Hempel (1905–) et Paul Oppenheim , tandis qu'Ernest Nagel (1937–) expose la *théorie! de la réduction nomologique.*

La *théorie de l'explication déductive–nomologique* soutient l'existence de deux types d'énoncés différents dans toute explication: l'*explicans* . et l'*explicandum.* Le premier est composé d'une ou plusieurs lois (généralisation nomologique), qui sont des énoncés existentiels mais aussi des énoncés universels décrivant les conditions initiales. L'ensemble des énoncés de l'*explicans* doit non seulement être vrai et contenir au moins une loi, mais aussi il doit avoir un contenu empirique vérifiable. L'*explicandum* est un énoncé existentiel ou une loi qui est une conséquence logique de l'*explicans.* Mais cette théorie n'a pas été exempte de critiques. Une de ces critiques vise la confusion entre des énoncés nomologiques et des énoncés conditionnels universaux, puisque la notion de loi n'est pas expliquée mais supposée. Une autre critique fait remarquer la nécessité d'une théorie de

l'explication qui tient compte des intérêts pragmatiques dans le cadre de l'explication est proposée.

> **Supposons que, dans le dortoir d'une école de filles, une surveillante, lors de sa ronde, découvre un professeur masculin nu comme un ver essayant de se faufiler dans les toilettes. La surveillante, si elle s'intéresse à la physique, pourrait déduire le phénomène observé, en conformité avec le modèle [déductif- nomologique] d'explication, de principes physiques, par exemple la loi affirmant qu'aucun corps ne peut se mouvoir avec une vitesse supérieure à la vitesse de la lumière. Mais cette explication serait dépourvue d'intérêt pragmatique. [Jacob, 1980, page 242]**

La théorie de la réduction de Ernest Nagel (1961) fait la différence entre deux types de réductions: les réductions homogènes et les réductions hétérogènes, en se basant sur le vocabulaire descriptif de chacune. Lorsque les deux théories, celle à réduire et celle réductrice, ont le même vocabulaire descriptif, il s'agira d'une réduction homogène, tandis qu'elle sera hétérogène dans le cas contraire. Un exemple de réduction homogène est la réduction des lois galiléennes sur les trajectoires des projectiles à la surface de la Terre à la théorie newtonienne de la gravitation universelle. En revanche, la réduction de la thermodynamique classique à la théorie cinétique des gaz est considérée comme hétérogène. En effet, dans ce cas, quelques traductions de termes sont nécessaires. Ainsi, par exemple, le terme "température" contenu dans la loi de Boyle de la thermodynamique classique est traduit par les termes d'énergie cinétique moyenne des molécules de gaz. Mais, pour que la réduction hétérogène aboutisse, il faut proposer des "lois ponts" (*bridge laws*). Une loi-pont est le résultat d'une expression que nous décrivons ci-dessous. Supposons que nous ayons une théorie $T1$ qui est réductible de façon hétérogène à une autre théorie $T2$. Soit (a) une loi de la théorie $T1$, C et D étant des prédicats simples ou complexes en $T1$:

(a) Tous les C sont (produisent, causent) des D.

Pour pouvoir réduire (a) à une loi de la théorie $T2$, il est nécessaire et suffisant que (b) et (c) soient aussi des lois, P et Q étant des prédicats simples ou complexes en $T2$:

(bi) Tous les C sont P.

(bii) Tous les D sont Q.

(c) Tous les P sont (produisent, causent) des Q.

Pour que toute la théorie $T1$ soit réductible à $T2$, il est nécessaire de trouver des lois-pont de ce type pour toutes les lois appartenant à $T1$.

On voit bien, donc, que le but du réductionnisme est clair: d'un côté, il y a une économie lexicale, mais de l'autre, la théorie réduite ne serait ainsi qu'un cas particulier d'une théorie plus générale. Ceci a pour corollaire deux conséquences très chères aux positivistes logiques: d'une part, il s'opère une réduction ontologique des termes descriptifs de la théorie réduite et de l'autre, il serait possible d'avoir un programme dans lequel toutes les sciences de la nature seraient réduites aux théories de la physique et de donner ainsi raison à la position physicaliste forte, puisqu'on aurait montré que la matière est l'unique substance existante.

Malgré le fait que l'optimisme réductionniste des positivistes les pousse à croire au triomphe du programme physicaliste même en ce qui concerne les sciences humaines ou sociales, ceci apparaissait plus difficile. Nous verrons plus tard que les efforts faits en psychologie dans le cadre réductionniste ont eu pour résultat les théories behavioristes.

1.5 La Seconde Guerre, la révolte contre le néo-positivisme et les sciences cognitives

La fin de la Seconde Guerre est la date symbolique de la révolte contre le néo-positivisme, dont deux principes sont au centre de la controverse: d'une part la foi sans limite dans le formalisme logique et le mépris pour l'étude des langues naturelles, et d'autre part l'autorité absolue donnée au rôle de l'expérimentation dans la découverte scientifique.

Nous avons déjà exposé brièvement la position du groupe de Cambridge à l'égard de l'étude des langages naturels et de son interprétation des *Investigations philosophiques* de Wittgenstein.

Néanmoins, le groupe d'Oxford a été encore plus important pour le développement que les sciences cognitives ont connu plus tard. Ce groupe soutient la défense d'une aire qui a été laissée de côté par les logiciens : la pragmatique. Les outils formels développès par Russell, Tarsky et Carnap permettent l'analyse de phrases affirmatives seulement. Cependant, il existe dans la vie de tous les jours d'autres énoncés, comme les interrogations ou des énoncés impératifs, dont l'analyse en termes de valeur de vérité n'est pas pertinente. Lorsqu'on veut développer des critères d'évaluation qui leur soient adéquats, on s'aperçoit très vite que le contexte d'émission et la stratégie du locuteur sont des données incontournables dans cette démarche. Les néo-positivistes ont montrés des réticences à reconnaître l'importance du contexte, parce que ceci doit nécessairement faire appel à des traits proches d'une position psychologiste qu'ils récusaient, comme par exemple assigner un ensemble d'intentions au locuteur d'une phrase.

Dans la sémantique fregéenne, un énoncé déclaratif a trois rôles : tout d'abord, en tant que signification linguistique conventionnelle dans la langue, ensuite en tant que proposition parce qu'il est porteur d'une valeur de vérité, et finalement comme contenu cognitif de la croyance qu'entretient l'émetteur de l'énoncé. Soit l'énoncé "La neige est blanche". La valeur de vérité de cette proposition dépend uniquement de sa signification linguistique ou conventionnelle. Par contre, dans le cas de l'énoncé "Je suis française", bien que la signification conventionnelle soit toujours la même, sa valeur de vérité dépend de la nationalité de l'émettrice. Cela veut dire que, dans ce cas, nous avons un énoncé pour lequel, bien qu'il possède une signification conventionnelle constante, la valeur de vérité de la proposition qu'il représente varie selon le contexte de son émission. Ce dernier énoncé est dite une expression indexicale. Traditionnellement, la signification des expressions indexicales est composée par des règles de référence du type: "*Je* désigne le locuteur", "*Tu* désigne le destinataire", "*maintenant* désigne le moment (ou une plage temporelle incluant le moment) où l'on parle". [Récanati, 1992, cf. page 240] Or, cette analyse montre premièrement que l'hypothèse du contexte nul dans l'interprétation des énoncés, comme les néo-positivistes ont bien voulu le croire, fait abstraction de traits importants comme les indexicaux, et deuxièmement que la frontière entre la sémantique – l'étude du sens des phrases – et la pragmatique – l'étude du sens que les phrases prennent dans le contexte d'énonciation – n'est qu'un mirage. [Proust, 1982, page 22]

De tous les partisans de la logique formelle de son époque, Frege est le seul à avoir pris en compte, bien que de façon un peu périphérique, un concept voisin de celui de stratégie du locuteur : la coloration d'un énoncé. En effet, bien que Frege considère les concepts de sens et référence comme seuls pertinents pour la signification, il a esquissé une différence entre sens et coloration (*Färbung*). Soit l'énoncé "Paul est un homme très brillant", qui a une coloration différente, tout en ayant la même signification linguistique et valeur de vérité que "Paul est un homosexuel très brillant" ou que "Paul est un mec très brillant". Ces trois énoncés marquent des différences dans la stratégie du locuteur. Dans cette ligne d'analyse, John Austin (1911–1960) proposa son programme, qui est l'un des plus importants dans la théorie pragmatique et dont le but principal est de caractériser la stratégie du locuteur. Austin vise à caractériser les types d'actions qu'on peut accomplir soit en émettant une phrase, soit comme résultat de son émission. Cette théorie connue sous le nom d'actes de langage (*speech acts*) montre que le langage ne sert pas seulement à décrire l'état du monde, mais également à essayer de le changer. Elle affirme aussi qu'il est plus pertinent de parler de directions d'ajustement possibles entre ces énoncés et le monde que de valeur de vérité.

La signification d'une phrase sera, dans ce cadre, équivalent au but dans lequel l'acte de langage a été accompli. Le but d'un acte de locution peut être soit de rendre les mots (la proposition qu'elle représente) conformes au monde comme dans le cas des assertions, soit de rendre le monde conforme aux mots comme dans les cas de promesses ou de demandes. Dans la théorie originale d'Austin (1962), ces actes de langage sont répartis en trois catégories.

- Actes locutoires : c'est l'acte de dire quelque chose. Ex : Il a dit "Tue-la! " en exprimant par "Tue" tue et en se référant à elle par "la". Ex: Il m'a ordonné (conseillé) de la tuer.
- Actes illocutoires : c'est l'action accomplie en disant quelque chose, c'est l'action de poser

une question, de donner un conseil, de faire une prédiction.

- Actes perlocutoires : c'est l'acte accompli par le fait de dire quelque chose. Il s'agit de cas où l'action est de dire quelque chose qui vise à persuader quelqu'un d'agir, ou à le faire revenir à la raison. Ex.: Il m'a persuadé de la tuer.

La caractérisation des locutions d'Austin implique que deux phrases peuvent accomplir le même acte de locution seulement si elles ont la même signification (sens) . Donc "...deux phrases ont la même signification si elles peuvent être utilisées pour accomplir le même acte de locution. Mais, malgré cette liaison, personne, même Austin, ne peut donner comme explication de la signification des actes de locutions, parce que la théorie devient étroitement circulaire" [Fodor, 1977, page 22]. Voilà pour la réaction des linguistes à la formalisation logique des énoncés et la remise en valeur des langages naturels. Maintenant, je vais exposer brièvement la réaction contre les principes épistémologiques de la philosophie positiviste : l'unification des sciences dans le physique, le rôle, excessif selon les détracteurs, attribué à l'expérimentation dans le cadre de la découverte scientifique.

Paul Feyerabend (1924-1994) et Thomas Kuhn (1922-) , qui sont habités d'un esprit anti-positiviste, sont deux des principaux acteurs de cette réaction. Ils affirment que l'intuition et la création jouent aussi un rôle dans la découverte scientifique, qui ne peut se contenter de la seule expérimentation. Ils s'efforcent aussi de montrer que le progrès des sciences est loin d'être cet *in crescendo* basé sur une accumulation invariante de données observables. Thomas Kuhn publie en 1962 *La Structure des révolutions scientifiques*, qui est considéré comme le manifeste anti-empiriste. Mais un mouvement comme celui-ci n'a été possible que grâce au renouveau de la sociologie des sciences opéré dans l'après-guerre aux États-Unis. C'est Alexandre Koyré qui attira l'attention sur l'importance des présupposés métaphysiques sans lien avec l'expérience. L'étude des sciences ne se fait plus dans un cadre abstrait, et prend en compte le contexte des découvertes qui font l'objet de l'analyse.[7]

Il ne serait pas juste de dire que le néo-positivisme n'était pas du tout intéressé aux problèmes ou concepts mentaux malgré l'esprit qui mettait d'avance en échec toute tentative de fonder des distinctions ou des concepts philosophiques importants sur des notions psychologiques.

> Au contraire, l'analyse de la psychologie et la question entre les sciences morales et les sciences de la nature étaient parmi les plus importantes dans l'agenda des positivistes, et les écrits de Wittgenstein, de Ryle et d'Austin ont donné lieu à des travaux remarquables sur le problème des autres esprits, l'imagination, la sensation, la croyance, l'intention ou la volonté. [Engel, 1994b, page 7]

De cette façon-là, la philosophie de l'esprit, considérée comme la partie de la philosophie qui analyse des concepts mentaux, c'est-à-dire comme une recherche conceptuelle, *a priori* et 'en fauteuil', détachée des travaux des psychologues et des scientifiques" est née.

Les progrès de l'informatique, et notamment de l'intelligence artificielle, des neurosciences, de la linguistique et l'importance des écoles psychologiques comme le behaviorisme ont fait se rassembler ces domaines sous l'unique dénomination de sciences cognitives . En effet, suite au behaviorisme, considéré jusqu'alors comme la version scientifique de la psychologie, les investigations deviennent – comme nous le montreront au cours de ce travail– *cognitifs* parce que des concepts comme *représentation*, *image*, *croyance* et *désir*, *but* s'avèrent centraux dans le cadre explicatif. Ces concepts sont souvent empruntés aux autres sciences appartenant aux sciences cognitives. Mais il ne faut pas oublier ce que relève remarquablement Hilary Putnam (1926-) :

> Les tentatives de Frege, Russell, Carnap et du premier Wittgenstein ont été considérées comme des attaques contre la métaphysique, mais, en fait, elles comptent parmi les plus ingénieuses, les plus profondes et techniquement les plus brillantes constructions de systèmes métaphysiques jamais développées. Même si elles ont échoué, la logique symbolique moderne, une grande partie de l'actuelle théorie du langage et une partie de la science cognitive contemporaine proviennent toutes de ces tentatives.[Putnam, 1985, page 29]

[7] Comme exemple, citons Pierre Jacob, qui montre l'image que ce révisionnisme donne d'Isaac Newton : "à l'image d'un Newton soucieux jusqu'à l'obsession de ne pas proposer d'hypothèses . . . léguée par Ernest Mach, les nouveaux historiens substituent une image de Newton névrotique, passionné d'alchimie, empêtré de croyances religieuses et fasciné par le problème de la Trinité" [Jacob, 1980, page 233].

Le but des sciences cognitives est de fonder les bases d'une psychologie scientifique comme une science naturelle de l'esprit. La tâche de ces sciences sera la naturalisation [8] de l'intentionnalité, c'est-à-dire en expliquant l'esprit humain comme un système matériel et en rejoignant ainsi la tradition analytique anglo-saxonne dont elles sont nées. Or, l'intentionnalité étant "la propriété en vertu de laquelle toute sorte d'états et d'événements mentaux renvoient à ou concernent ou portent sur des objets et des états de choses du monde" [Searle, 1983, page 15], les sciences cognitives ont pour objet de montrer comment un système matériel peut d'abord représenter la réalité et agir selon les représentations ainsi formées. Pour aboutir à une explication de ces phénomènes, deux vocabulaires sont possibles : le vocabulaire physicaliste et le vocabulaire intentionnel.

Le premier est celui utilisé par les sciences naturelles comme la physique ou la biologie. En revanche, le second est celui que nous utilisons pour exprimer des propositions et concepts psychologiques et qui nous sert à expliquer, décrire et prédire – au moins de façon naïve – les comportements chez les humains en termes – entre autres – de croyances et de désirs. L'opposition de ces deux types de vocabulaires va au-delà d'un choix sans importance. Bien au contraire, elle nous renvoie à un problème classique dans la tradition philosophique, le problème de la relation corps – esprit (*the mind - body problem*). En effet, l'existence de ces deux catégories de vocabulaire signale-t-elle l'existence de deux substances différentes auxquelles on fait référence, ou est-ce que ce sont deux façons différentes d'exprimer des phénomènes d'un seul type? Et s'il s'avère que ce sont véritablement deux substance différentes, comment peut-on alors expliquer l'interaction de l'une sur l'autre? Nous allons, dans le chapitre suivant, exposer les antécédents de ce sujet qui nous permettra après d'aborder le problème de l'intentionnalité.

[8] Le naturalisme est la doctrine philosophique qui soutient que les seules choses existantes sont des choses naturelles, c'est-à-dire des particules et des propriétés naturelles. Cette doctrine est donc très proche du physicalisme. Le naturalisme comporte deux versions : celle qui considère la catégorie du naturel comme étant intuitivement donnée, et celle qui est basée sur l'idéalisation des sciences naturelles. C'est cette seconde version qui est l'objet de ce texte.

Chapitre 2

Le problème corps – esprit

Car le monde changera si vous élevez vos enfants dans la liberté du behavioriste.

John Watson

2.1 Introduction

Le programme de naturalisation des sciences humaines comme la psychologie – entre autres – tel qu'il est proposé par les positivistes a mis l'accent sur un problème traditionnel en philosophie : le problème corps–esprit.

Dans le chapitre précédent nous avons signalé l'existence de deux types de phénomènes, les pbénomènes physiques et les pbénomènes mentaux et nous nous nous sommes demandés si l'existence de différents vocabulaires pour leur description est due au fait que ce sont des phénomènes de natures différentes ou si par contre, ces langages ne sont que des variations explicatives de réalités toutes de même substance – qu'elle soit mentale ou physique.

L'analyse du problème de la relation corps–esprit admet deux méthodes possibles qui ne sont pas mutuellement exclusives: celle qui vise à déterminer la nature des états mentaux et celle qui aborde la question de la pertinence causale des états mentaux. La première correspond au débat ontologique sur l'existence de deux substances – la matérielle et la mentale – ou simplement d'une seule; la deuxième approche tente de déterminer si le mental peut avoir un rôle causal dans le comportement. Évidemment la réponse à la question ontologique première restreint le nombre des réponses possibles à la deuxième interrogation.

Il y a quatre grandes catégories de réponses: *le dualisme*, *le matérialisme*, *l'idéalisme* et *le fonctionnalisme*.

Commençons par le dualisme qui généralement admet deux types de formulations. D'un coté le dualisme classique de René Descartes (1596–1650) qui soutient l'existence de deux substances : la matérielle qui est l'objet des science naturelles et la spirituelle qui compose nos états de conscience. D'autre part, le dualisme des propriétés qui peut rester agnostique vis à vis du problème ontologique de la substance.

La position matérialiste postule en général qu'il n'existe que la substance matérielle et que tous les états mentaux comme la douleur, les croyances, le désir ne sont que des états physiques.

L'idéaliste soutient aussi qu' il existe une unique substance mais pour lui elle est d'essence mentale; un champion de cette doctrine est George Berkeley (1685–1753). Le point central de la théorie de Berkeley n'est pas de mettre en doute la réalité des objets physiques mais d'affirmer que seuls les esprits existent de façon indépendante alors que les objets physiques sont inertes et dépendent totalement des esprits qui les perçoivent. Ainsi il conclut qu'il n'y a que des esprits et des impressions sensorielles causées par des objets matériels dont nous n'avons aucune évidence directe.

Le fonctionalisme se situe à mi-chemin du dualisme et du matérialisme. Il s'oppose au dualisme parce qu'il nie l'existence d'une substance mentale et il contredit le matérialisme en niant que les états mentaux soient identiques aux états matériels. Le point de vue de cette dernière théorie est que l'important n'est pas la substance matérielle en soi mais la façon dont elle est organisée. Ainsi une des idées maîtresses du fonctionnalisme réside dans la *multiplicité d'implantations* ce qui veut dire qu'un même état mental peut être implanté dans différentes matières physiques i.e. un cerveau, un ordinateur; ce que nous appelerons le problème de la multiréalisabilité.

Ce que je donne ci-dessus n'est qu'une définition schématique des positions sans prétention à décrire chacune des catégories, ce qui exigerait des critères difficiles à établir pour classifier les états ou propriétés dans l'ordre mental ou physique en un cadre épistémologiquement neutre comme celui utilisé jusqu' ici.

Il existe différents critères pour séparer les aspects mentaux des aspects physiques. Donald Davidson, par exemple a postulé que les états mentaux sont ceux qui admettent une description mentale alors que les états physiques ont une description physique. Mais qu'est-ce qu'une description physique? [1] Il y a plus d'une réponse. D'un côté, une description peut être considérée physique si elle se conforme aux lois de la physique mais on peut exiger de surcroît que le vocabulaire qui l'exprime contienne uniquement des termes publiquement observables. En relation au critère déterminant du mental, la tradition analytique veut considérer comme d'ordre mental les énoncés qui *portent sur ou sont à propos des* objets, propriétés ou relations autres que physiques. Ainsi, pour cette école, les croyances, les désirs, les intentions, en général tous ces énoncés qui sont groupés sous le nom d'*attitudes propositionnelles* seront définis comme mentaux.

2.2 Le "trilemme" classique

La difficulté dans la résolution du problème corps-esprit tient aussi au fait que quelque soit la réponse qu'on veuille donner, on se heurte à l'impossibilité de concilier les trois propositions contradictoires de la triade suivante [2]:

(1) Il y a des états (propriétés, processus, etc.) physiques qui sont connus par le sens commun et avec l'aide des sciences physiques.

Il existe aussi des états (propriétés, processus, etc.) mentaux qui sont révélés tant par l'introspection que par la psychologie cognitive.

Mais les propriétés et faits mentaux sont distincts des propriétés et des faits physiques;

(2) Les propriétés et faits mentaux ont une efficacité causale, et ne sont pas de purs épiphénomènes;

(3) *Principe de l'interaction causale* : les propriétés ou faits physiques sont seuls suffisants pour causer l'occurrence des mouvements physiques et des actions.

Etant donné qu' il s'agit d'un *trilemme* il est nécessaire de réfuter au moins une des proposition pour conserver la cohérence de l'ensemble. Dans le contexte du dualisme, les propositions (1) et (2) sont acceptées mais par contre la proposition (3), le principe de l'interaction causale, doit être récusée, sous peine d'accepter l'epiphénomènalisme du mental. Mais si on veut concilier (3) tout en soutenant (1) et (2) nous nous trouvons face au problème de l'interaction causale: comment les états mentaux peuvent-ils interagir de façon causale sur les entités physiques si on viole le principe de l'interaction causale? Une solution élégante est celle de Donald Davidson bien qu'elle menace de relèguer le mental au rang d'un épiphènomène.

Par contre, un physicaliste va accepter (1) et (3) mais il doit rejeter (2) dans le sens qu'il ne peut concéder une efficacité causale aux états mentaux.

L'idéaliste tiendra a soutenir du (1) l'existence des états mentaux mais il rejettera probablement (3) et bien entendu, il acceptera *mordicus* (2).

[1] cf. [Davidson, 1980b] aussi voir le chapire 6 de ce texte.
[2] cfr. [Engel, 1994b, page 20]; [Jacob, 1992a, page 321]

La classification des catégories qu'on vient de proposer est plus ou moins schématique parce qu'il y a des nuances entre les théories qui la sous–tendent. L'existence de ce *puzzle* à trois pièces tient à l'impossibilité de concilier les intuitions que nous avons tous au sujet du mental avec le principe d'interaction causale.

Ces intuitions ou expériences sont en relation avec plusieurs aspects de notre vie de tous les jours:

- *La valeur explicative de la psychologie naïve / ordinaire.* Premièrement, les explications que nous donnons de nos actes ou de ceux des autres, de même que les prédictions des actions futures des autres nous semblent d'une valeur irréfutable, et ont à nos yeux un pouvoir de prédiction tel que la vie en société est inconcevable autrement.

  Par exemple, situons–nous dans une station de sports d'hiver dans les Alpes vaudoises où des amis de Jean l'attendent en faisant des commentaires du type: "Il est tombé 40 cm de neige ici, à Villars-sur- Ollon. Bien que Jean vienne d'acheter une voiture à double traction, il ne va pas arriver au chalet parce que l'année passée, lorsqu'il essayé de monter, il a eu un grave accident qui a failli le tuer. Et d'ailleurs c'est pour cela qu'il a acheté cette voiture à double traction. Je crois que ce qui l'a poussé à un tel investissement n'est pas son désir d'arriver jusqu'au chalet en voiture mais le fait qu'il se sente plus en sécurité s'il conduit ce type d'engin".

  Ce genre d'explication sur notre comportement et sur celui d'autrui, nous y avons recours tous les jours et on considère que cela ressort de la *psychologie populaire ou naïve ou ordinaire* [3] et on peut observer que ces prédictions basées sur l'interaction des états mentaux (croyances, désirs) que nous nous attribuons régulièrement aux uns et aux autres sans effort apparent ne sont pas si mauvaises, puisqu'elles suffisent en général à régler nos relations interpersonnelles quotidiennement. Le terme *psychologie ordinaire* est inspiré de l'expression *physique naïve* qui est "la *théorie* du comportement des corps solides, liquides, gazeux de taille macroscopique à la surface de la Terre grâce à laquelle tout membre de l'espèce humaine s'oriente avec succès dans son environnement local".

- *Les expériences qualitatives ou* qualia. En deuxième lieu, divers états de la conscience tels que les états qualitatifs ou subjectifs sont difficiles à concevoir comme résultant ou émanant des états purement physiques. Les états qualitatifs mentaux sont appelés *qualia* et sont caractérisés par les qualités phénoménales des expériences qui les accompagnent. Une autre caractéristique est qu'ils ne sont accessibles qu'introspectivement et ne peuvent être rapportés qu'à la première personne. Des exemples des qualia sont *voir quelque chose de rouge, avoir mal au dos.* En général, les sensations et les douleurs brutes sont des exemples de *qualia.*

  Les propriétés des *qualia* sont les suivantes: ils sont *ineffables* du fait que même douée d'un pouvoir d'éloquence magnifique, je ne pourrais guère les décrire; mais ils sont *ineffables* également parce qu'ils sont *intrinsèques* dans le sens qu'ils sont atomiques et non-analysables parce qu'ils sont simples et homogènes, d'où la difficulté à les décrire et à le partager avec autrui. Ils sont donc *privés.* Ils sont *directement ou indirectement appréhendés par la conscience* parce qu'ils font partie des propriétés de mon expérience. [Dennett, 1990, cf. page 519–520]

  Pourquoi ces expériences conscientes qualitatives rendront–elles plus difficile la résolution du *trilemme classique*? Pour répondre à cette question complexe il faudra tenir compte de différents aspects du problème.

  Lorsqu'on accepte que ces expériences qualitatives sont ineffables et intrinsèques nous avons des difficultés à les naturaliser parce qu'il n'est pas possible de les décrire, encore moins si on veut user d'un vocabulaire physicaliste et à la troisième personne. Lorsqu'on se place dans

[3] Pierre Jacob (1992) Nous signale les différences entre les trois termes qui peuvent être choisis pour traduire l'expression anglaise *folk psychology* . Le terme *naïve* souligne le caractère tacite de cette manière de raisonner. Le terme *populaire* fait référence au fait que cette connaissance empirique relève d'une tradition non–scientifique et n'est pas enseignée en faculté. Finalement le terme *ordinaire* conserve selon Jacob la double acception. Nous allons choisir, avec Jacob le terme *ordinaire* à partir d'ici pour traduire le mot anglais *folk.* (cf. [Jacob, 1992a, page 316])

une perspective physicaliste, nous avons aussi des difficultés parce que bien qu'on puisse les expliquer du point de vue neurologique, c'est–à–dire du point de vue des états physiques tels que la perception du rouge ou la saveur salée, il n'est pas vrai que nous puissions expliquer en disant par exemple *le rouge tel que je le vois* ou *tel que vous le voyez* selon ce qui ressort de nos cerveaux individuels ou de nos systèmes nerveux. Pour éclaircir cette dernière argumentation nous allons emprunter un exercice à la pensée de Frank Jackson [Jackson, 1982, page 128] connu sous le nom d'argument de la connaissance.

Supposons que nous soyons en train de faire une expérience pour cataloguer la capacité des gens à discerner les couleurs et que, pendant ce test, un certain individu nommé Fred soit découvert. Fred a une capacité hors pair pour discerner les couleurs comme on n'en avait jamais enregistré auparavant. Il est capable de choses inconcevables pour nous tous. En effet lorsqu'on lui présente une corbeille pleine de tomates mûres, il les classe en deux groupes inégaux et ceci d'une manière complètement cohérente. Cela veut dire que si vous lui bandez les yeux, mettez les même tomates dans la corbeille, enlevez le bandeau et lui demandez de refaire la même opération avec cette corbeille–là, il va former exactement les mêmes groupes qu'auparavant. Si vous lui demandez comment il a réussi une telle performance, Fred va vous répondre que pour lui les tomates de la corbeille étaient de deux rouges différents, disons $rouge_1$ et $rouge_2$. Il va vous dire qu'il a essayé d'apprendre à ses amis à reconnaître cette différence mais qu'il a toujours échoué, ce dont il avait déduit que la plupart des gens sont aveugles à la différence entre ces deux rouges. Cependant pour lui $rouge_1$ et $rouge_2$ sont des couleurs aussi distinctes que le bleu et le jaune par exemple, ce qui veut dire que Fred peut voir au moins une couleur de plus que nous. Une étude physiologique a montré que Fred pouvait diviser en deux groupes différents les ondes de lumière appartenant au spectre du rouge aussi nettement qu'il peut le faire entre le spectre du bleu et celui du jaune.

Supposons qu'on ait fait une étude approfondie du système optique et du cerveau de Fred et imaginons qu'on ait découvert que ses cônes oculaires répondent différemment à un groupe des ondes de la section rouge (ou peut–être a–t–il un cône additionnel) et que de ce fait Fred puisse avoir un registre plus étendu d'états cérébraux concernant les comportements relatifs aux perceptions de la couleur.

Mais on peut pousser encore plus loin notre hypothèse ; supposons que nous arrivions à connaître tout sur le corps de Fred. Non seulement ses dispositions de comportement, sa physiologie mais aussi toute son histoire en relation avec les autres gens pour autant que cela puisse s'exprimer selon un langage exclusivement physicaliste. Malgré cette connaissance exhaustive que nous parviendrions à accumuler, nous ne saurions répondre à la question fondamentale concernant le type d'expérience que Fred éprouve quand il perçoit la différence entre le $rouge_1$ et le $rouge_2$?

Qu'est–ce que percevoir une nouvelle couleur? Même si on possède la description physique complète et exhaustive de tous les faits relatifs à la performance de Fred, on n'a pu répondre aux questions fondamentales; le physicalisme n'arrive pas à nous satisfaire; quelque chose d'important est ignoré, l'espect subjectif des expériences conscientes n'étant pas explicité à partir de ce point de vue.

Maintenant je vous propose d'imaginer qu'avec tous les données obtenues à partir de Fred, je puisse fabriquer un agent doué de la même structure, des mêmes rôles fonctionnels, de la même histoire, entre autres que Fred, mais qu'une fois finie l'expérience reconstructive, ce nouveau Fred s'avère n'avoir aucun des états qualitatifs qu'avait l'individu nommé Fred au départ. Dès lors on peut conclure que le nouveau Fred n'est qu'un robot et que la description physique n'est pas suffisante pour permettre la reconstruction de l'original.

2.3 Stratégies du physicalisme

La solution du problème corps–esprit réside en la réconciliation de l'intuition très répandue que le mental a un rôle causal d'une part avec le principe de l'itéraction causale d'autre part. Les straté-

gies employées par le physicalisme sont principalement trois: le réductionnisme, l'éliminativisme et la survenance (*surpervenience*). Le réductionnisme prend diverses formes mais en href il peut être caractérisé comme une thèse prônant que les états mentaux *ne sont que* des états physiques. L'éliminativisme considère que les énoncés mentaux sont vides de signification tandis que la psychologie naïve est loin d'être une théorie qui puisse acquérir une valeur prédictive ou normative quelconque. La survenance est une stratégie non–réductioniste qui postule que les états mentaux peuvent être considérés comme survenant sur des états physiques.

2.3.1 La réduction

Les positions physicalistes modernes résultent d'un mouvement qui a son origine dans le positivisme logique. Dans le chapitre précédent nous avons exposé les techniques réductrices utilisées par les partisans de cette doctrine dans le but de faire aboutir le projet d'unité des sciences. Cependant, il n'existe pas un consensus unanime, même quant au concept de réduction, et encore moins sur la manière dont cette réduction doit être faite.

La réduction peut être considérée comme un type spécial de relation d'identité appelée la relation *n'est rien que* (*nothing–but*) [Searle, 1994b, page 112] qui en général soutient que la les items X peuvent être réduits aux items Y si et seulement si X *n'est rien que* Y. Il y a deux grandes tendances; la première est le réductionnisme éliminativiste selon lequel un item X se trouve totalement éliminé quand il est réduit par ou réduit à un item Y. Selon la seconde tendance, l'item réduit X n'est pas totalement éliminé mais continue à jouer un rôle ou à avoir une place dans la description de l'item réducteur Y.

Le concept de réduction est appliqué dans divers domaines; en général on en reconnaît cinq différents :

- *La réduction ontologique*

  Une entité X est réduite a une entité Y. C'est la forme plus importante de réduction : les objets d'un certain type sont considérés comme n'étant rien d'autre que des objets d'un autre type. Tous les autre types des réduction tendent à celle–ci. Par exemple, les gènes ne sont rien d'autre que des molécules de DNA.

- *La réduction ontologique des propriétés*

  Une propriété P est réduite à une autre propriété P'.

  Ceci est aussi une réduction mais qui concerne seulement les propriétés et en général elle procède par une réduction de termes théoriques. Ainsi, par exemple la chaleur n'est rien que l'énergie cinétique moyenne des molécules.

- *La réduction dans le domaine théorique*

  Une théorie T est réduite à une autre théorie T'. Nous avons déjà exposé ce type de réduction dans le chapître précédent. Par exemple j'ai cité le cas de la réduction des lois de gaz aux lois statistiques de la thermodynamique.

- *Réduction logique ou de définition*

  Des énoncés ou mots se référant à un type d'entités peuvent être traduits sans aucun résidu à ceux qui se réfèrent à d'autre types d'entités. Par exemple, les propositions se référant aux nombres peuvent être traduites à des propositions sur la théorie des ensembles. Ce type de réduction suppose une réduction ontologique; dans l'exemple que nous venons de donner les nombres ne sont rien que des ensembles d'ensembles.

- *La réduction causole*

  Des entités du type X et des entités de type Y ont toutes les deux un potentiel causal mais les pouvoirs des entités du type X sont explicables en termes des pouvoirs causaux des entités de type Y.

En général la réduction causale amène à la réduction ontologique. Par exemple, les objets solides sont impénétrables par d'autres objets solides; ils sont aussi résistants à la pression, etc. Mais ces pouvoirs causaux sont explicables à partir des mouvements vibratoires des molécules aux structures en treillis.

La réduction ontologique peut être *forte* ou *faible*. Dans le premier cas non seulement l'existence de l'entité réductrice est présupposée mais aussi celle de l'entité réduite . Si on accepte la réduction des entités mentales à des entités physiques, alors les deux types sont censés exister effectivement. Mais cette existence peut être interprétée de deux façon différentes, ce qui permet d'identifier deux versions du réductionnisme fort. Une des ces versions soutient que l'entité réduite est contenue d'une certaine façon dans l'entité réductrice. L'autre version, par contre maintient que les deux entités appartiennent à différents mondes ou domaines ontologiques. Dans ce cas, la réduction opérée entraîne l'abandon ou l'ignorance d'un type d'entité au bénéfice de l'autre qui appartient à un domaine ontologique différent.

La réduction ontologique faible implique un changement, non pas du point de vue de la dimension ontologique mais du point de vue linguistique ou selon la théorie de la connaissance en rapport avec l'entité à réduire. En définitive, plutôt qu'un changement ontologique, c'est un changement épistémologique qui admet trois interprétations ou versions différentes.

La première version est celle qui considère les concepts ou entités à réduire faux ou erronés; la réduction consiste à remplacer une conception erronée par une autre correcte. Dans cette version, réduction équivaut à élimination. Il ne semble pas toute–fois pertinent de parler de réduction dans ce cas parce que la différence entre ce dernier concept et élimination est réduite à néant.

La deuxième version soutient l'arbitrarité de la réduction; elle considère que les deux items, ce qui est réduit aussi bien que le réducteur, ont le même degré d'acceptation et que la réduction est dictée par des circonstances externes ou contingentes. Dans cette deuxième version, réduire équivaut à choisir.

La troisième prône que la réduction se base sur des critères d'adéquation, ce qui veut dire le remplacement d'une conception par une autre plus adéquate. Cependant le critère d'adéquation ontologique croissante suppose une révision ontologique parce que si un terme s'avère n'être plus adéquat il faut se demander si cette situation ne tient pas au fait que son référent est vide dans le nouveau cadre.

2.3.2 L'éliminativisme

L'éliminativisme est très proche de la conception faible du réductionnisme et il consiste à remplacer les termes mentaux par des termes physiques. Par exemple, dans le cas de l'éliminativisme neurologique, les concepts mentaux sont éliminés au bénéfice des concepts et des généralisations des neurosciences. Il existe une autre théorie éliminativiste syntaxique du mental proposée par Steven Stich dans laquelle les concepts mentaux sont remplacés par des généralisations syntaxiques caractéristiques des solutions représentationnelles cognitives. Ces deux versions d'éliminativisme considèrent que les énoncés de la psychologie ordinaire sont vides de signification. Ils considère que la conception des états ou processus auxquels ces énoncés font référence comporte une erreur ontologique, d'où le besoin de remplacer cette conception erronée par une autre qui soit correcte.

2.3.3 Corrélation et dépendance

Les théories invoquées dans le fonctionnalisme computationnel[4] qui veulent conserver la pertinence du vocabulaire intentionnel sont obligées de récuser toute démarche réductionniste ou éliminativiste. Néanmoins, ces mêmes théories ont l'ambition de rester des théories physicalistes.

Elles se trouvent dans l'obligation de proposer à la place de l'identité une autre relation entre les états physiques et les états mentaux. Traditionnellement deux réponses qui ne s'excluent pas nécessairement entre elles ont été données comme alternatives à l'identification. Une de ces réponses est la *survenance*, l'autre est le concept de *réalisation physique*.

[4] Voir chapitre 7.

Les différences entre ces deux positions sont les suivantes. La *survenance* peut être conçue comme une corrélation systématique entre les entités ou propriétés mentales et physiques sans qu'on soit obligé de décrire le mécanisme de base de cette corrélation. Dans la survenance les propriétés physiques et les propriétés mentales peuvent être considérées comme séparées et distinctes.

La relation de *réalisation* est plus forte que celle de survenance car elle n'affirme pas seulement l'existence d'une corrélation ou d'une relation causale du physique au mental; elle affirme en outre que les entités ou les propriétés mentales sont réalisées par les entités et les propriétés physiques. En définitive, elle affirme que toute propriété mentale est une instance d'une réalisation physique sur laquelle la propriété mentale repose.

La relation de réalisation consiste à démontrer explicitement le type de mécanisme de base de la corrélation. Je vais discuter maintenant les différentes conceptions de survenance et de réalisation.

Le concept de survenance

L'idée centrale de la survenance est la suivante:

> [...]no difference of one kind without differences of another.[Kim, 1990, page 23].

Dans la philosophie de l'esprit un énoncé équivalent sera: il n'est pas possible d'avoir un duplicata physique, molécule–à–molécule d'un agent sans avoir, à la fois, un duplicata psychologique. Toutefois cette idée peut être expliquée de différentes manières qui ne sont pas forcément équivalentes les unes aux autres. Mais d'abord un peu d'histoire de ce concept.

Le concept de survenance (*supervenience* du latin "supervenire") n'est pas exclusif à la philosophie de l'esprit. Selon Jaegwon Kim [5] le concept a été utilisé par les émergentistes du début de siècle. La thèse de l'émergence soutient que lorsqu'un processus physico–chimique de base atteint un certain niveau de complexité d'un type approprié, de véritables caractéristiques nouvelles, dont les mentales font surface en tant que propriétés émergentes. Lloyds Morgan qui est considéré comme un des plus importants représentants de cette théorie utilise le concept de *survenance* comme une variante du terme *émergence*, ce dernier étant considéré comme le terme de base de la théorie. Plus tard, en 1950 R. M. Hare utilise le terme survenance en un texte non–publié dans le cadre de la philosophie morale. En ce domaine, le concept de survenance sert à exprimer la nécessaire covariation (non systématique selon certaines auteurs comme G. E. Moore) entre les caractéristiques morales de correction ou incorrection des actions – propriétés morales évaluatives – et les autres caractéristiques – propriétés naturelles ou descriptives –, ces dernières amorales mais donnant les raisons de la correction ou de l'incorrection de l'action. Ainsi, Hare (1952) introduit le terme survenance pour la première fois dans la philosophie morale:

> Supposons que nous disions que 'Saint François était un homme bon'. Il est logiquement impossible de dire cela et de maintenir en même temps qu'il aurait pu exister un autre homme placé exactement comme saint François, et qui serait différent de lui en ceci seulement qu'il ne serait pas bon. ([Hare, 1952, page 145] traduction française dans [Jacob, 1992a, page 298])

Le concept de survenance exprime cette covariation nécessaire entre les propriétés morales et les propriétés descriptives non–morales et non–evaluatives. L'idée sous–jacente est la suivante: des propriétés d'un type donné doivent covarier avec les propriétés d'un autre type d'une certaine façon. Or l'idée de la survenance repose sur la relation entre deux ensembles des propriétés. Cette version que nous avons esquissée est la version de survenance relative à une notion constitutive, mais Donald Davidson [Davidson, 1980a, *Mental Events*] a proposé une version de survenance en rapport avec la causalité dans la philosophie de l'esprit au début des années soixante–dix en étant probablement le premier à utiliser ce concept dans ce cadre. Selon [Kim, 1990] le concept de survenance que prône Davidson est proche des idées de Moore dans le cadre de la théorie de l'émergence: les phénomènes mentaux sont survenants aux phénomènes physiques sans être pour autant réductibles au moyen des définitions ou des lois à ces derniers.[6]

La démarche de Davidson est un démarche non–réductioniste du mental comme nous allons le voir dans le chapitre 5. Trois idées sont étroitement associées au concept de survenance:

[5] cf. [Kim, 1990]

[6] Voir chapitre sur l'émergence dans la troisième partie

a *La covariation des propriétés*: Si deux propriétés de base sont indiscernables alors elles seront indiscernables au niveau des propriétés dont elles surviennent.

b *Dépendance*: Les propriétés survenantes sont déterminées ou sont dépendantes des propriétés de base dont elles surviennent.

c *Non – réductibilité*: La covariation et la dépendance de propriétés peuvent être obtenues sans que la réductibilité des propriétés survenantes aux propriétés de base soit nécessaire.

De toutes les propriétés qu'on vient de citer, la covariance constitue la principale composante.

Covariance, dépendance et survenance. Lorsque l'on regarde les trois conditions qui définissent la survenance, on peut s'interroger sur la différence entre les concepts de survenance et de covariance. Il s'agit de savoir si le concept de covariance n'aurait pas suffi à exprimer ce que désigne le concept de survenance. Dans le problème corps–esprit par exemple, lorsque l'on applique le concept de survenance aux propriétés physiques sur des propriétés mentales on veut dire que les propriétés physiques déterminent les propriétés mentales, ou qu'il est impossible d'avoir des différences entre les propriétés mentales sans qu'il y ait aussi des différences physiques. Néanmoins l'inverse n'est pas vrai.

La survenance est donc une relation *asymétrique* tandis que la covariance n'est ni symétrique ni *asymétrique*.[7]

En effet, selon [Kim, 1990, cf.page 148][8] la covariance n'est ni symétrique ni *asymétrique* car c'est essentiellement une relation de nécessité ou d'implication (*entailment*) qui exprime *une détermination mutuelle* ou une *dépendance mutuelle* entre propriétés.

La covariance n'établit pas donc une dépendance métaphysique d'un élément de la relation par rapport à l'autre élément comme l'implique la survenance. Voilà pourquoi il faut ajouter la deuxième condition qui fait de la dépendance une caractéristique de la relation de survenance.

Selon [Kim, 1990] les relations de dépendance permettent l'utilisation des clauses explicatives du type "parce que" ou "en vertu de" pour signaler les propriétés physiques déterminantes des propriétés mentales. Une relation de corrélation ne permettra donc pas de donner un compte–rendu explicatif semblable. Selon Kim :

> Property covarition per se is metaphysically neutral; dependence and other such relations, suggest ontological and explanatory directionality –that upon which something depends is ontologically and explanatorily prior to, and more basic than, that which depends on it. In fact, we can think of dependency relation as explaining or grounding property covariations; e. g. one might say that mental properties covary with physical properties because the former are dependent on the latter. [Kim, 1990, page 148 dans Kim(1993)]

Finalement, on voit que le concept de survenance ne peut pas être amalgamé au concept de covariation. Bien que le concept de covariation implique le concept de dépendance il n'y a pas d'emblée de relation asymétrique et donc il n'est pas apte à dénoter la relation corps–esprit que l'on veut définir.

Types de survenance

Je vais exposer les différentes variantes du concept de survenance, mais étant donné qu'il s'agit de covariance des propriétés il faut fixer un ensemble sur lequel ces propriétés vont s'appliquer. Disons que cet ensemble soit la totalité des systèmes cognitifs, bien que la définition que je donne soit assez vague intentionnellement pour permettre au lecteur de composer cet ensemble à sa guise (c'est à dire qu'il peut considérer que les systèmes artificiels sont concernés par cette intention ou aussi les animaux); cela ne devrait pas poser de problème pour atteindre notre but final. Au contraire, je

[7] Il est simple de trouver des exemples de covariance qui ne sont pas asymétriques. Par exemple, l'aire de la surface de la sphère covarie avec le volume de cette sphère mais l'inverse se vérifie aussi. On ne peut donc pas dire que l'aire de la surface de la sphère *détermine* le volume pas plus qu'on ne peut soutenir l'inverse. D'où le relation de covariance qui résulte de ce cas n'est pas asymétrique.

[8] Ce numéro de page correspond à la réédition du texte dans [Kim, 1992c].

pense que cela permet une plus grande généralisation des définitions. Je vais appeler cet ensemble d'application le domaine d'application. Dans ce domaine il y a deux ensembles de propriétés: l'ensemble M qui est composé par les propriétés mentales et l'ensemble P des propriétés physiques (comme par exemple des propriétés physico–chimiques ou biologiques entre autres). Est–ce que l'ensemble des propriétés mentales M survient à l'ensemble P des propriétés physiques dans le domaine de référence?

Nous avons (au moins) les réponses suivantes [Kim, 1994, cf. page 577]:

- La survenance faible: Nécessairement (cela veut dire dans tous les mondes possibles) si X et Y appartenant au domaine sont indiscernables en vertu de P (P–indiscernables), alors X et Y sont aussi M–indiscernables. La survenance des propriétés M des propriétés P est assurée seulement à l'intérieur de chaque monde possible. Selon la survenance faible, dans un même monde il ne peut pas exister de duplicata physiques sans qu'il soient aussi des duplicata mentaux. Cependant rien n'empêche d'avoir des duplicata physiques qui soient M–discernables dans des mondes différents. En effet, rien n'empêche dans cette définition d'avoir par exemple un monde physique exactement semblable au nôtre mais totalement dépourvu de propriétés mentales. Ceci est évident lorsqu'on observe que la définition se vérifie trivialement dans le cas où les propriétés mentales n'existent pas, c'est à dire lorsque l'ensemble des propriétés mentales est vide. En effet, si deux éléments appartenant au domaine sont P–indiscernables, il seront aussi de ce fait M–indiscernables puisqu'il n'existe aucune critère de M–discernabilité. Aussi on peut avoir le complément de l'exemple précédent, des mondes physiquement tels que celui que nous connaissons mais où tous les éléments appartenant au domaine aient les mêmes propriétés mentales.

  S'il y a des éléments P–indiscernables, ils seront aussi M–indiscernables parce que n'importe quel élément du domaine est M–indiscernable. De ce qu'on vient d'exposer on peut voir les limites de la survenance faible car en effet, elle ne permet pas de rendre compte d'une position physicaliste forte soutenant que les propriétés physiques déterminent tous les faits du monde. Comme le dit Kim de manière synthétique mais éloquente: Une fois qu'on a distribué toutes les propriétés physiques aux individus, on aimerait qu'il existe *une façon unique* de faire la distribution des propriétés mentales. Ce problème sera réglé par les deux autres conceptions des survenance.

- La survenance globale: étant donnés deux mondes P–indiscernables (cela veut dire que les propriétés sont distribuées de la même façon à tous les individus du domaine) alors ils résulte aussi qu'ils sont M–indiscernables (ils ne peuvent pas être différents quant à la distribution des propriétés mentales). La survenance globale concerne les mondes (plutôt que les individus) auxquels on applique les propriétés physiques. Cela veut dire que les termes de substitution dans chaque propriété sont les mondes mêmes et non les éléments du domaine. Mais deux mondes seront P–impossibles à distinguer (M–indistinguibles) si les extensions de chacune des P–propriétés (M–propriétés) sont les mêmes dans chacun des mondes. [9].

  Si les mondes sont P–indistinguibles alors il seront M–indistinguibles. Ceci donne l'assurance que le caractère mental du monde dépend de son caractère physique. Cependant, la survenance globale a aussi ses limites; à première vue on a l'impression qu'elle n'est pas assez fine puisque elle ne tient pas compte du niveau individuel.

  Premièrement la survenance globale permet l'existence de mondes où les propriétés physiques sont presque identiques, excepté pour un petit détail, –par exemple que dans un des mondes il existe un atome de plus d' hydrogène– mais différant radicalement du point mental (par exemple un monde peut être démuni de toutes propriétés physiques). Ceci veut dire qu' étant donné deux mondes presque P–indiscernables, il puissent être radicalement M–discernables

[9] La question qui vient à l'esprit est: que se passe-t-il lorsque les deux domaines ne contiennent pas les même individus? Comme par exemple, en méréologie, la survenance des propriétés et relations qu'entretiennent les trous avec les parts. Cette question cependant n'est pas pertinente dans le domaine qui nous occupe, on peut considérer les domaines d'étude dans la philosophie de l'esprit comme homogènes. Néanmoins Kim a traité de ce problème . Voir [Kim, 1993d].

[Kim, 1990, cfr. page 23]. Néanmoins, je considère que cette situation, bien qu' étrange n'est pas très grave parce que les mondes dont on parle sont P–indiscernables et ceci n'est pas un problème de granulité. Or le fait d'avoir un atome de plus d'hydrogène est un indice de discrimination des propriétés et que les propriétés survenantes soient proches ou non ne diminue en rien la vérité de la proposition: "ces deux mondes ne sont pas globalement survenants".

Deuxièmement, l'autre critique que l'on fait est que la survenance globale n'implique pas la survenance faible et ceci me semble poser un problème parce que cela montre que la vraie limite tient à un problème de granularité. En effet la survenance faible permet une analyse plus fine que dans le cas de la survenance globale. Par exemple, s'il existait deux mondes qui fussent des duplicata physiques mais dans l'un desquels les êtres humains eussent troqué leurs capacités mentales avec celles des escargots, la survenance globale continuerait de se vérifier tandis que ce ne serait pas le cas pour la survenance faible.

La troisième thèse postule la survenance forte qui consiste en la vérification des propriétés sur des individus d'un domaine mais en les comparant d'un monde à l'autre.

- La survenance forte: soient deux individus X et Y, et deux mondes possibles V et W, si X qui appartient à V est P–indiscernable avec Y qui appartient à W (ceci veut dire que X contient en V exactement les mêmes propriétés que Y possède en W) alors X de V est M–indiscernable de Y de W. Comme on peut voir et à la différence de ce qu'autorise la survenance faible, les individus peuvent être choisis parmi divers mondes possibles mais s'il sont P–indistinguibles ils doivent s'avérer aussi être M–indistinguibles. Il est facile de voir que la survenance forte implique la survenance faible (rien n'interdit de choisir les individus dans le même monde) et la survenance globale (il suffit de définir que le monde possible W vérifie la propriété physique P si et seulement s'il existe un individu appartenant à W tel quel X vérifie P, la conséquence de l'implication précédente étant une proposition inférable à partir de la survenance forte). La survenance globale n'implique ni la survenance faible, comme on a déjà démontré ni la survenance forte (pour cette dernière implication voir [Kim, 1993c]). Cependant, selon Kim [Kim, 1994, cf. pages 578 et 580] lorsqu'on ajoute les prémisses métaphysiques pertinentes la survenance globale impliquerait la survenance forte. Plus exactement les deux types de survenance sont équivalentes quand on restreint les propriétés pris en compte aux propriétés intrinsèques et il semble que la survenance globale n'implique pas la survenance dans le cas où les propriétés mentales (et non les propriétés de base ou physiques) incluent des propriétés non intrinsèques. Finalement, l'introduction du concept de survenance dans la pbilosophie de l'esprit a été motivée par le fait qu'il semble conférer une plausibilité au physicalisme non réductioniste. L'idée que la notion de survenance n'implique pas la réduction du mental au physique est actuellement débattue. La question est de savoir si de ces concepts possibles de survenance que l'on propose il n'en existe pas au moins un qui soit compatible avec au moins un des concepts de réduction (soit–elle analytique–logique ou nomologique) des états mentaux aux états physiques. Selon [Kim, 1994] un certain énoncé de survenance forte sera cohérent avec le réductionnisme, mais la question reste ouverte et controversée.

> As we saw, a supervenience claim consists of a claim of covariance and a claim of dependence (leaving aside the controversial claim of non–reductibility). This means that the thesis that the mental supervenes on the physical amounts to the conjonction of the two clairos: (1) the mental covaries with the physical (*à la* strong or global supervenience), and (2) the mental depends on the physical. Notice, however, the fact that *the thesis says nothing about just what kind of dependence is involved in mind–body supervenience.* When you compare the supervenience thesis with the standard positions on the mind – body problem, you are struck by what the supervenience thesis doesn't say. For each of the classic mind–body theories has something to say, not necessarily anything plausible, about the kind of dependence that characterizes the mind–body relationship. [Kim, 1994, page 582]

2.3.4 Le concept de réalisation

Le concept de réalisation s'applique aussi bien aux propriétés qu'aux lois. Dans ce chapitre je vais traiter de la réalisation des propriétés tandis que j'aborderai la réalisation des lois dans le chapitre 7 (§7.5.1) en relation avec les conséquences que ce dernier concept peut avoir sur le rôle causal des états mentaux.[10]

La littérature spécialisée fait état de deux concepts de réalisation des propriétés que je vais appeler fort et faible.

Dans [LePore and Loewer, 1989] ces auteurs soutiennent un concept faible de réalisation. Il justifient le rôle que ce concept a dans la philosophie de l'esprit de nos jours.

> It is practically received wisdom among philosophers of mind that psychological properties (including content properties) are not identical to neurophysiological or other physical properties. The relationship betwren psychological and neurological properties is that the latter *realise* the former. Furthermore, a single psychological property might (in the sense of conceptual possibility) be realised by a large number, perhaps in infinitely many, of different pbysieal properties and even by non–physical properties.
>
> Exactly what is it for one of an event's properties to *realise* another? The usual conception is that e's being P realisrs e's being F iff e is P and e is F and there is a strong connection as a necessary connection which is *explanatory*. The existence of an explanatory connection between two properties is stronger than the claim that $P \rightarrow M$ is physically necessary since not every physically necessary connection is explanatory. [LePore and Loewer, 1989, page 179, les italiques font partie du texte original]

Il y a deux caractéristiques importantes à noter dans la définition précédente. La première est que la connexion demandée est du type explicatif. La seconde est que la connexion requise entre les propriétés réalisées et celles qui réalisent (de niveau plus bas) est une condition seulement suffisante.[11]

Le caractère explicatif de la relation de réalisation indique qu'elle est relative à une théorie ou à un mécanisme d'implantation. Par ailleurs, on dira que si une propriété M est réalisée par la propriété P, alors la propriété P implante la propriété M, aussi on dira que la propriété M est instantiée par la propriété P.[12] Lepore et Loewer ne le disent pas en ces mots mais il parlent d'un système de connexions entre les propriétés réalisées et réalisatrices.

Le caractère explicatif que l'on demande à la relation de réalisation sera mieux compris si l'on se tourner vers le concept de théories des propriétés et de théories de transformation dû à Robert Cummins.

Dans son livre [Cummins, 1983] cet auteur établit une différence entre deux types différents de théories scientifiques. D'un côté on trouve les théories qui visent à expliquer les événements, voire les changements et de l'autre, les théories destinées aux explications de propriétés.

Selon [Cummins, 1983, cf. page 1] les premières sont des théories de transition qui ont comme but fondamental la réponse à la question *quand*.

Les autres théories qui tendent à expliquer des propriétés ne doivent pas être comprises comme répondant à la question "Pourquoi le système S a–t–il acquis la propriété P? " mais plutôt "En vertu de quoi S possède–t–il maintenant la propriété P? "

Cummins signale que les théories de transformation n'ont comme énoncés que des lois causales mais il signale deux faits à prendre en compte. Le premier est que les énoncés nomologiques causaux ne sont pas forcement explicatifs, l'autre est qu'il existe d'autres énoncés nomologiques qui ne sont point causaux et dont certains sont explicatifs.

---

[10] Le concept de multiréalisation sera aussi traité dans le chapitre 7 (§7.5.1.).

[11] Il faut noter que la condition de *nécessité physique* invoquée dans la citation de Lepore et Loewer signifie qu'il existe une connexion physique nécessaire de la loi énoncée par la proposition suivante $P \rightarrow M$ que j'appellerai désormais proposition L. L'existence de L est nécessaire physiquement dans le cadre d'un réseau de relations ou d'une théorie et doit aussi avoir des pouvoirs explicatifs pour satisfaire à la définition de réalisation.

Par contre, à l'intérieur, pour ainsi dire, de la loi L il s'avère que P est *une condition suffisante* pour M, si l'on a P alors on a M autrement dit, il n'est pas possible d'avoir P sans avoir M. [Kim, 1992b, cf. page 7]

[12] Je préfère traduire le mot anglais *implementation* par *implantation* à l'instar d'une grande partie de la communauté en informatique et contrairement à la plupart des philosophes français qui ont en général choisi de la traduire par *implémentation*.

Dans [Cummins, 1983, cf. page 7–8] cet auteur donne une liste des énoncés nomologiques non causaux; parmi eux on trouve, par exemple, les lois qui se réfèrent aux corrélations nomiques, par exemple la loi qui établit la corrélation entre la conductivité électrique et la conductivité thermique. Cependant ce type de loi n'a pas de rôle explicatif car bien au contraire, celle–ci se trouvent dans la catégorie des faits qu'une théorie doit expliquer.

Les lois d'instantiation ou de réalisation appartiennent à un autre type de relation nomologique non causale et elles ont des pouvoirs explicatifs.

> These are lawlike statements specifying the (or an) analysis of a specified type of system. An example is the statement that temperature is instantiated in a gas as the average mean kinetic energy of the molecules in the gas. [Cummins, 1983, page 7]

En définitive, selon Cummins une théorie qui a des pouvoirs explicatifs doit expliquer non pas pourquoi un système a passé de l'état e_1 à l'état e_2, mais en vertu de quoi s'est produit ce changement d'état.

Un exemple intéressant de cette situation est discuté par [Berckermann, 1992, cf. page 110]. Il s'agit des molécules de cristaux violets qui ont une couleur bleue–violette en conditions normales. Lorsqu'on leur ajoute une certain quantité d'acide hydrochlorique, leur couleur tourne au vert. Une explication causale de cet événement répondra à la question "Pourquoi le cristal violet a–t–il viré au vert? " La réponse sera: "Parce que l'on a ajouté une certaine quantité d'acide chlorhydrique".

Une explication en termes des propriétés répondra à la question: "En vertu de quoi la couleur a–t–elle viré du bleu–violet au vert? " Alors la réponse sera (plus ou moins) la suivante:

"Lorsque l'on expose les molécules de cristal violet à la lumière pour qu'elles absorbent les fréquences du jaune–orange du spectre responsable de la couleur bleue –violette du cristal, il faut que le niveau d'énergie de certains de leurs électrons monte. Pour que ce soit le cas, il faut que trois des atomes présents dans la molécule soient capables d'acquérir des charges positives. Alors lorsque l'on ajoute une certaine quantité d'acide hydrochlorique, un de ces atomes prend un proton en annulant par ce même fait sa capacité de prendre une charge positive. La couleur vire au vert parce que les molécules ne sont plus en mesure d'absorber la fréquence jaune–orange du spectre de lumière. Voilà pourquoi la couleur tourne au vert."

En définitive, selon Cummins, une théorie a des pouvoirs explicatifs lorsqu'elle est capable d'expliquer les capacités dispositionelles de façon telle qu'elle arrive à établir des relations entre des *types* généraux d'événements. Au contraire, les théories de transformation citent tout simplement les capacités dispositionelles, et les lois qui en découlent servent à établir des relations entre événements particuliers.

Une relation de réalisation requiert le cadre d'une théorie explicative des propriétés dans le sens de Cummins, et c'est cela qui est impliqué par la condition de pouvoir explicatif énoncé par Lepore et Loewer.

Revenons maintenant à la seconde condition de la définition de réalisation.

La caractéristique de *suffisance* des propriétés réalisatrices de niveau plus bas dans $P \rightarrow M$ (où P est une propriété réalisatrice et M une propriété réalisée)[13] justifie le caractère non–réductionniste que l'on pourrait attribuer à la relation de réalisation.

En effet, si l'on demande aussi le caractère nécessaire alors on tombe dans une position qui équivaudrait à une réduction. Si l'on a une relation nécessaire et suffisante entre les propriétés réalisées et leurs réalisatrices alors on aura une coextension nomologique[14] entre les deux ensembles de propriétés qui serait équivalente à une réduction. La relation de réalisation ne sera, dans ce cas, qu'une définition d'une propriété de niveau supérieur en fonction des autres de niveau fondamental. Les définitions ainsi obtenues seront équivalentes à des lois–ponts dans le cadre d'une réduction à la Nagel[15] et elles comprendront une vraie réduction.

Il est intéressant de noter que rien n'empêche la condition suffisante d'être aussi nécessaire. La définition de la réalisation n'assure pas par elle–même le caractère non–réductionniste de la relation, elle en laisse seulement la possibilité.

[13] Pour le moment je laisse volontairement de côté la possibilité que P soit une propriété ou une disjonction des propriétés.

[14] qui repond à une loi naturelle

[15] cf. chapitre 1 §1.4

Le concept de réalisation se trouve à la base du concept de multi–réalisation que je discuterai plus tard. Le concept de multiréalisation assure le caractère non–réductionniste de la relation.

En effet, dans le cas du problème corps–esprit, la possibilité logique de multiréalisation de certaines propriétés de la cognition que l'on utilise pour soutenir une position non–réductionniste du mental sera incompatible avec l'exigence d'une double conditionnelle.[16]

Je qualifie la position de LePore et de Loewer de *faible* étant donné qu'elle n'exige pas un niveau d'implantation de base. Le concept de réalisation de ces deux auteurs peut être appliqué entre n'importe quelle paire de propriétés appartenant à deux niveau consécutifs (sans que le plus bas soit basique). Cette conception de la réalisation sera reprise par Fodor pour expliquer les propriétés des entités ou des contenus mentaux en tant que (multi)réalisées non pas directement par le niveau de base mais par le niveau computationnel.[17]

Outre le concept de LePore et de Loewer que je viens de discuter, il est utile d'analyser le concept de réalisation de Jaegwon Kim.

Jaegwon Kim accepte, en principe, la conception de la réalisation de LePore et de Loewer mais il établit certaines nuances.

Pour Kim, le rôle explicatif que LePore et Loewer attribuent à la définition est en fait une simple relation epistémique.

Cette interprétation ne me semble pas faire justice à ce que LePore et Loewer veulent définir.

Je suis d'accord avec Kim qui, à la place d'une relation explicative demande une relation métaphysique objective entre les propriétés P et M car une telle relation ne peut pas être un fait brut [Kim, 1993b, cf. page 196] mais ceci n'est pas en contradiction avec la conception de LePore et Loewer. Je reviendrai plus tard sur les différences entre Kim et les deux autres auteurs.

Kim propose comme concepts–clés pour les explications des relations de réalisation le *mécanisme causal* et la *microstructure*.

> When P is said to 'realize' M in the system s, P must specify a micro–structural property of s that provides a causal mechanism for the implementation of M in s; moreover, in interesting cases – in fact, if we are speaking meaningfully of 'implementation' of M – P will be a member of a family of physical properties forming a network of nomological connected micro–structural states that provides a micro–causal mechanism, in systems appropriately like s, for the nomological connections among a broad system of mental properties of which M is an element. This underlying micro–states will form an explanatory basis for the higher properties and the nomic relations among them; but realization relation itself must be distinguished from explanatory relation. [Kim, 1993b, page 197]

Kim met l'accent sur le fait qu'une relation de réalisation doit être explicative mais cette condition ne doit pas s'ajouter explicitement à sa définition, elle est forcement dérivée des autres conditions que l'on impose au concept de réalisation.

Je pense que la définition de Kim est équivalente à celle de LePore et de Loewer seulement dans le cas où le niveau qui réalise les propriétes est le niveau de base ou élémentaire. En effet, dans le cas que je viens de citer les mécanismes de réalisation s'appliquent à la microstructure et comme on se trouve au niveau de base, la relation de réalisation sera d'emblée explicative. [18]

J'ai décidé de qualifier le concept de réalisation de Kim de *fort* pour souligner l'exigence que le niveau d'implantation soit celui de base.

2.3.5 Conclusion sur les stratégies du physicalisme

J'ai exposé les stratégies utilisées par le physicalisme pour résoudre le problème de la relation corps–esprit. De tout ce qui vient d'être exposé, on doit conclure que ces stratégies, bien que paraissant disjointes à première vue se recoupent en fait comme c'est le cas entre un certain type de réductionnisme et l'éliminativisme; de même on vient de montrer que selon Kim la survenance forte peut être considérée comme compatible avec des positions réductionnistes.

[16] Voir chapitre 7 et la conclusion pour une critique des limites du concept de multiréalisation.

[17] Voir chapitre 7.

[18] Traditionnellement, dans la philosophie des sciences on considère les relation nomologiques données en termes de microstructure comme explicatives d'emblée. (Par exemple, en relation avec l'*hypothèse de l'unité de la science* voir [Oppenheim and Putnam, 1958]. D'ailleurs j'y reviendrai dans le chapitre 8 concernant le concept d'émergence.) Ainsi tous les énoncés nomologiques qui font référence au micro–niveau sont toujours explicatifs.

En effet, une démarche typiquement réductionniste consistera à trouver des corrélations entre les *types* des états mentaux et les physiques; plus concrètement elle consistera à établir pour chaque propriété mentale une propriété physique co–extensionnelle (le fait que la propriété soit co–extensionnelle en vertu d'une définition, c'est à dire analytiquement ou sur la base des lois empiriques, donc nomologiquement importe peu à cet égard). La condition pour effectuer une réduction est appelée *la condition de connectibilité forte* qui affirme:

> Pour toute propriété mentale m il existe une propriété physique *p* telle que *nécessairement*, si le système vérifie m à l'instant *t* alors p est aussi vérifiée par lui à l'instant *t*.

Maintenant nous pouvons comparer la condition de connectibilité forte avec la version alternative suivante de la survenance forte :

> Nécessairement, pour toute propriété *m* dans *M*, pour tout *x* qui vérifie m alors il existera une propriété p dans P telle quelle *x* vérifie aussi p, et *nécessairement* pour toute chose qui vérifie p alors elle vérifie aussi *m*.

Comme on peut le voir, la première partie de cet énoncé équivaut à la condition de connectibilité sauf pour la référence à l'indice de temps, mais dans la deuxième partie de l'énoncé de la version de survenance forte l'opérateur modal *nécessairement* sert à stabiliser la relation dans tous les mondes possibles.

Cependant, le travail n'est pas fini parce qu'on cherche pour chaque propriété m (mentale) une propriété *p* (physique). Selon la théorie de survenance forte, pour chaque propriété mentale m il existe un répertoire des propriétés physiques p telles que la vérification d'une d'elles suffit pour affirmer la vérification de m. Alors la propriété qu'on recherche ne sera que la disjonction de toutes les propriétés *p* associées au répertoire de m. Il est évident que dans ce cas, l'extension de la propriété résultante de cette disjonction sera co–extensionnelle avec m. Ceux qui tiennent le concept de survenance pour un concept non– réductionniste ne sont pas d'accord avec cette interprétation. Ils nient que la propriété obtenue par une disjonction puisse être considérée comme base de réduction des propriétés mentales, étant donné son artificialité et sa complexité et ils préfèrent utiliser la survenance globale pour assurer une corrélation stable au sein des mondes possibles afin d'échapper justement à la menace du réductionnisme.

J'exposerai tout au long de ce travail les différentes théories de la relation corps – esprit et je ferai allusion à ces diverses stratégies.

2.4 L'approche dualiste

Le dualisme cartésien postule non seulement l'existence de deux substances, l'une physique et l'autre immatérielle ou spirituelle mais aussi les relations causales dans les deux sens. La version cartésienne du dualisme reçoit le nom d'*interactionisme* car René Descartes (1596–1650) soutient la pertinence causale du mental, tantôt comme relation causale du mental au physique, tantôt dans le sens inverse.

Le dualisme a d'autres versions, ces positions sont les suivantes: l'occasionalisme, le parallélisme, l'épiphénoménalisme et le dualisme des propriétés.

L'occasionalisme prôné par Nicolas de Malebranche (1638–1715) et Arnold Geulincx (1624–1669) entre autres, pour qui Dieu est l'intermédiaire qui établit les connexions entre les événements mentaux et physiques. Ainsi dans cette perspective, il n'existe pas de connexion causale entre le corps et l'esprit; seul Dieu est la vraie cause; seule la providence divine est responsable des régularités de nos expériences.

En effet, Malebranche écrivait

> Nous avons que deux sortes d'idées, idées d'esprit, idées de corps; et ne devant dire que ce nous concevons, nous ne devons raisonner que suivant ces deux idées. Ainsi, puisque l'idée que nous avons de tous les corps nous fait connaître qu'ils ne se peuvent remuer, il faut conclure que ce sont les esprits

que les remuent. Mais quand on examine l'idée que l'on a de tous les esprits finis, on ne voit point de liaison nécessaire entre leur volonté et le mouvement de quelque corps, que ce soit, on voit au contraire qu'il n'y en a point, et qu'il n'y en peut avoir. On doit aussi conclure, si on veut raisonner selon ses lumières, qu'il n'y a aucun esprit créé qui puisse remuer quelque corps que ce soit comme cause véritable ou principale, de même que l'on a dit qu'aucun corps ne se pouvait remuer soi-même.

Mais lorsqu'on pense à l'idée de Dieu, c'est-à-dire d'un être infiniment parfait et par conséquence tout-puissant, on connaît qu'il y a une telle liaison entre sa volonté et le mouvement de tous les corps, qu'il est impossible de concevoir qu'il veuille qu'un corps soit mû, et que ce corps ne le soit pas. Nous devons donc dire qu'il n'y a que sa volonté qui puisse remuer les corps, si nous voulons dire les choses comme nous les concevons, et non pas comme nous les sentons. La force mouvante des corps n'est donc point dans les corps qui se remuent, puisque cette force mouvante n'est autre chose que la volonté de Dieu. Ainsi les corps n'ont aucune action: et lorsqu'une boule qui se remue en rencontre et en meut une autre, elle ne lui communique rien qu'elle ait: car elle n'a pas elle-même la force qu'elle lui communique. Cependant une boule est cause naturelle du mouvement qu'elle communique. Une cause naturelle n'est donc point une cause réelle et véritable, mais seulement une cause occasionnelle, et qui détermine l'auteur de la nature à agir de telle manière en telle et telle rencontre. [Malebranche, 1963, Livre *VI*, 2^e partie, chapitre 3]

Le parallélisme a été originairement proposé par Gottfried Wilhelm Leibniz (1646–1716) qui, tout en soutenant le dualisme des substances arguait que les événements mentaux et physiques ont une corrélation régulière mais ne relevant d'aucune relation causale directe ou indirecte. Dans le parallélisme l'on admet que les causes des événements physiques se trouvent en d'autres événements physiques, celles des événements mentaux en d'autres événements mentaux mais l'on nie l'existence des relations causales entre les deux substances. Leibniz trouve que la solution proposée par l'occasionalisme n'était pas une bonne stratégie pour un Dieu intelligent. Il propose, en revanche que Dieu a établi une harmonie depuis le début, ainsi chaque objet agit de la façon appropriée dans le moment opportun. Le parallélisme était si évident pour Leibniz qu'il voit même en lui une preuve de l'existence de Dieu.

A new proof of the existence of God can also be found here, one of surprising clarity. For the perfect agreement of so many substances which have no communication whatever with each other can come only from a common source [Leibniz, 1992, page 120]

Cette harmonie s'exprime sans arrêt dans la vie de tous les jours. Nous croyons qu'une balle de billard produit le mouvement d'une deuxième par une collision entre elles, mais en fait le mouvement de la deuxième balle est dû à la programmation que Dieu a fait des choses; dans cette programmation il est établi que la deuxième balle doit bouger au moment exact où que la première entre en contact avec elle.

Il en va de même pour l'interaction corps–esprit, lorsque mon esprit cause le mouvement du bras il bouge parce que Dieu à dessiné le bras de manière telle qu'il fait le mouvement au moment exact où je vais le bouger. Les deux exemples expliquent l'illusion que nous avons, selon Leibniz de l'existence de causes directes, mais ceci n'est qu'une illusion.

La métaphore proposée par Leibniz est celle de deux montres créées par Dieu, marchant en parfaite harmonie depuis le commencement et restant synchronisées à jamais. Antérieurement à la partie du texte que je vais citer, il récuse l'idée qu'il existe une influence entre les deux montres pour la synchronisation, de même qu'il récuse la possibilité d'une assistance quelconque entre les deux montres comme le suppose l'occasionnalisme puis il continue les critiques de ces positions comme suit:

But I hold that this [l'occasionalisme] makes a *deus ex machina* intervene in a natural and ordinary matter where reason requires that God should help only in the way in which he concurs in all other natural things. Thus there remains only my hypothesis, that is, the *way of prestablished harmony*, according to which God has made each of the two substances from the beginning in such a way that, though each follows only its own laws which it has received with its being, each agrees throughout with the other, entirely as if they were mutually influenced or as if God were always putting both his hand, beyond his general concurrence. [Leibniz, 1992, page 121]

Une autre théorie dualiste est l'épipbénoménalisme. L'épiphénoménalisme est à l'origine une thèse dualiste qui reconnaît que le monde physique est un système autonome et qu'il ne peut pas exister des relations causales allant du mental au physique. Néanmoins, l'epiphénoménalisme

soutient l'existence des états mentaux, dont l'occurrence est totalement déterminée par des états physiques. Ainsi, ces derniers sont des phénomènes primaires tandis que les états mentaux ne sont que des conséquences secondaires. Or les états mentaux, bien qu'existants, n'ont aucun rôle causal. L'exemple donné par Jérôme Schaffer[19] est celui de l'ombre de la main; si la main bouge son ombre aussi va changer de façon solidaire mais sans utiliser d'énergie pour opérer ce changement.

On peut voir que dans ces deux cas, le principe de l'interaction causale est sauf. Mais dans le parallélisme on ajoute la possibilité de relations causales entres des états mentaux tout en préservant le principe de causalité physique. Le quête de cohérence entre les énoncés (2) et (3) du trilemme se fonde sur la constance de la corrélation entre les états ou événements mentaux et physiques. Cependant, cette constance ne peut être expliquée au moyen d'aucune des méthodes de la science ou des statistiques modernes. Il faut donc admettre que cette occurrence est accidentelle, mais accepter une position semblable peut nous amener à soutenir que les résultats des autres domaines de la sciences sont, eux aussi les fruits du hasard plutôt que le résultat de régularités statistiques ou de corrélations entre les relations causales et les inférences des sciences.

L'épiphénoménalisme, par contre, tout en respectant le principe de l'interaction causale nie toute pertinence causale aux états mentaux et fait donc l'abandon d'une partie de la deuxième proposition de notre trilemme. A partir de là, l'existence effective des états mentaux est secondaire puisqu'ils peuvent être totalement assimilés aux état physiques dont ils dépendent et l'on peut ainsi opérer une réduction sur eux.

Dans la philosophie de l'esprit contemporaine on trouve une nouvelle version dualiste qui ne vise pas les substances mais les propriétés. Généralement les dualistes de propriétés admettent que les phénomènes mentaux sont des phénomènes physiques ou cérébraux mais affirment qu' ils possèdent aussi des propriétés mentales non-physiques. C'est dans ce sens que vont les arguments de la connaissance comme celui de Frank Jackson cité plus haut. Il y a aussi dans cette même ligne de pensée Thomas Nagel et Saul Kripke. Tous ces raisonnements maintiennent l'irréductionnisme des propriétés mentales. Lorsqu'on donne un compte rendu à la troisième personne de faits tels que "voir une couleur de plus que le reste des mortels" ou simplement "avoir mal à un orteil", quelque chose est omis comme je l'ai déjà expliqué plus haut. Or, pour les partisans du dualisme de propriétés, il existe des propriétés mentales que ne sont pas réductibles aux phénomènes physiques et ce point de vue oppose le dualisme des propriétés aux théories physicalistes éliminativistes ou réductionnistes que je présente dans les pages suivantes.

2.5 Les positions matérialistes réductionistes

Je vais commencer pour expliquer les positions matérialistes réductionnistes et ensuite j'exposerai la position éliminativiste neurologique

2.5.1 Le behaviorisme

Le behaviorisme fut la principale théorie psychologique jusqu'au milieu du siècle et il résulte de l'application à la psychologie du programme du positivisme logique selon son euphorique fureur réductrice et unificatrice des sciences. En effet, la psychologie était considérée comme une science dont les conceptions et méthodes ne correspondait pas aux postulats positivistes. A cette époque Rudolf Carnap [Carnap, 1932] a signalé cette situation dans un texte paru en 1932 :

> Aujourd'hui diverses sciences sont parvenues à différentes étapes dans le processus de décontamination de la métaphysique. Principalement grâce aux effort des Mach, Poincaré et Einstein et, généralement parlant, presque toute la physique est libre de métaphysique. D'autre part, en psychologie le travail pour arriver à une science libre de métaphysique a déjà commencé. La différence entre les deux sciences est plus claire lorsqu'on observe les attitudes des experts appartenant aux deux domaine vis-à-vis d'une position que nous récusons comme métaphysique et vide de signification. Dans le cas de la physique par exemple, les physiciens pour la plupart considéreraient une telle position comme anthropomorphique, mythologique ou métaphysique. Ils démontrent ainsi leur attitude contraire à la métaphysique qui correspond à la nôtre. D'autre part, dans l'exemple concernant la psychologie . . .,

[19] cf. [Schaffer, 1968]

la plupart des psychologues considérerait cette conception qu'on vient de critiquer comme évidente du point de vue intuitif. Par là on peut voir que l'orientation métaphysique des psychologues et la nôtre sont opposées. [Carnap, 1932, page 28 qui est une réimpression du texte paru en 1933, ma traduction]

Ainsi le behaviorisme a été la réponse inspirée par le positivisme et illustre l'effort de bâtir une théorie scientifique du mental, c'est à dire une théorie vierge de postulats métaphysiques. Le behaviorisme connaît au moins deux versions : le behaviorisme méthodologique en psychologie et le behaviorisme philosophique.

En psychologie, le behaviorisme méthodologique n'a pas pris formellement position sur le statut ontologique des énoncés mentaux. Selon cette version, toutes les explications basées sur des énoncés mentaux doivent être rejetées et ceci pour des raisons purement méthodologiques. En effet, les états internes, n'étant pas publiquement observables, ne sauraient faire partie d'une psychologie scientifique puisqu'ils ne permettent aucun type de vérification intersubjective. Dans ce sens Carl Hempel par exemple, soutenait que la psychologie ne pouvait traiter que du comportement, étant donné que les propositions référentes aux états mentaux sont vides de signification. Néanmoins il affirme que tous les énoncés mentaux ont une traduction douée d'une signification équivalente dans le langage physique. Par contre, Skinner acceptait l'existence de phrases mentales intraduisibles en termes physiques mais elles ne peuvent jouer aucun rôle selon lui dans une explication psychologique.

L'objectif de la psychologie sera donc la prédiction et le contrôle du comportement et l'unique donnée valable pour bâtir une telle théorie sera le comportement publiquement observable.

Le behaviorisme philosophique ou logique est lié à la thèse vérificationiste de la signification et défend des positions ontologiques. Pour le noyau le plus radical de ces théoriciens il fallait réduire toute expression mentale au domaine physique. Les énoncés mentaux sont considérés par Carl Hempel [Hempel, 1949] comme des abréviations des descriptions de certains modes de réponse physique des hommes ou des animaux. Du point de vue du trilemme classique les behavioristes logiques rejettent la proposition (2) de pertinence causale du mental en disant que celle-ci n'est qu'une illusion que la psychologie ordinaire se suscite elle-même. Les explications de la psychologie ordinaire ne sont que des analyses logiques ou conceptuelles entre des propositions mentales et au plus, elles peuvent jouer une rôle dans la compression du comportement mais en aucune façon rendre compte des vraies causes.

Fortement influencés par le Cercle de Vienne, les behavioristes logiques cherchaient à réduire la psychologie à la physique et cette réduction comportait deux phases.

La première phase consiste à formuler des adscriptions mentales qui ne sont autre chose que des dispositions du comportement. Lorsqu'on dit que "Jean a mal au dos" cela ne se réfère pas à un état interne quelconque de Jean mais indique simplement que Jean va se comporter d'une façon déterminée, par exemple va avoir du mal à ramasser quelque chose qui vient de tomber par terre, et sous-entend toute une batterie de comportements. Ainsi, les descriptions mentales sont sémantiquement équivalentes à un ensemble hypothétique de dispositions de comportement; ces dernières étant décrites par des propositions hypothétiques dont l'antécédent est un stimulus et le conséquent un comportement observable. Cette équivalence sémantique entre les deux types d'énoncés a conféré à cette démarche réductrice un caractère analytique. [20]

[20] Locke fut le premier à parler de l'existence de vérités analytiques par opposition aux vérités synthétiques. Néanmoins, on considère Kant comme celui qui a reformulé ce caractère dit analytique dans la philosophie contemporaine.

Dans la section 4 et 5 de l'Introduction à la *Critique de la raison pure* il établit la différence entre les jugements analytiques et les jugements synthétiques. L'analyse kantienne se fonde sur les types de rapports possibles entre le sujet (désormais A) et le prédicat (désormais B). Ces rapports peuvent être de deux types :

> Ou le prédicat B appartient au sujet A comme quelque chose qui est contenu implicitement (= *versteckter Weise*) dans ce concept A, ou B est entièrement en dehors du concept A, quoiqu'il soit, à la vérité, en connexion avec lui. Dans le premier cas, je nomme le jugement *analytique*, dans l'autre *synthétique*. Ainsi les jugements (les affirmatifs) sont analytiques quand la liaison du prédicat au sujet y est pensée par identité; mais on doit appeler jugements synthétiques ceux en qui cette liaison est pensée sans identité. On pourrait aussi nommer les premiers *explicatifs*, les autres *extensifs*, car les premiers n'ajoutent rien au concept du sujet par le moyen du prédicat, mais ne font que le décomposer par l'analyse en ses concepts partiels qui ont été déjà (bien que confusément) pensés en lui; tandis qu'au contraire les autres ajoutent au concept du sujet un prédicat qui n'avait pas été pensé en lui et

Or, l'énoncé "Jean a mal au dos" a pour signification un ensemble de propositions hypothétiques dont voici au hasard quelques unes:

"Si on l'invite a jouer au bowling, il va refuser."

"S'il avait quelqu'un pour lui faire un massage, il le lui demanderait."

---

qu'on n'aurait pu en tirer par aucun démembrement. [Kant, 1975, page 37]

Ensuite, Kant donne comme exemple de jugement analytique l'énoncé "Tous les corps sont étendus"; il signale que l'on n'a pas besoin de sortir du concept de corps "pour trouver l'étendu uni à lui (*sic*)", on ne fait que décomposer le concept du prédicat. Si l'on utilise le langage actuel on dirait que la propriété d'étendue est enchassée dans le concept de corps.

L'énoncé "Tous les corps sont pesants" en revanche, est un jugement synthétique car "ici le prédicat est tout à fait différent de ce que je pense dans le simple concept d'un corps en général (*sic*)". Finalement, les jugements analytiques n'étendent pas du tout nos connaissances; ils rendent intelligibles les concepts que l'on possède déjà. Les jugements synthétiques, par contre,

> ... je dois avoir en dehors du concept du sujet quelque chose encore (X) sur quoi l'entendement s'appuie pour reconnaître qu'un prédicat qui n'est pas contenu dans ce concept lui appartient cependant. [Kant, 1975, page 38]

Un des éléments sur lesquels l'entendement s'appuie pour formuler des jugements synthétiques est l'*expérience*. Dans l'exemple de la pesanteur, c'est l'expérience que l'on fait du corps qui nous permet de lier ce dernier caractère à ceux qui sont analytiquement contenus dans le concept de corps. Ainsi, Kant conclut dans la deuxième édition de la *Critique* que *les jugements de l'expérience, comme tels, sont tous synthétiques*. [Kant, 1975, cf. page 38]

Néanmoins, il y a d'autres jugements qui tout en étant synthétiques ne sont pas appuyés par l'expérience. Ces jugements sont les jugements synthétiques *a priori*. Un des exemples de Kant est "Tout ce qui arrive a sa cause". Ce dernier énoncé n'est pas un énoncé analytique parce que le concept de cause n'est pas contenu dans le concept de *quelque chose qui arrive*; parmi les choses que l'on peut concevoir à partir de ce concept il y a une certaine précedence dans le temps, mais pas le concept de cause. Il nous faut trouver une base d'entendement pour le jugement synthétique mais elle ne peut pas être l'expérience.

Si l'on regarde de près le dernier jugement, le concept du prédicat (cause) doit être ajouté comme une caractéristique "à tout ce qui arrive" et il a aussi un carantère d'appartenance nécessaire mais l'expérience ne peut en aucune manière dicter ce caractère nécessaire.

La compression de la possibilité du jugement synthétique *a priori* requiert, selon Besnier l'introduction d'un troisième terme entre les éléments unis dans un jugement; ce troisième terme sera la notion d'*expérience possible*.

> La notion d'« expérience possible » remplit en effet la fonction tierce exigée par le jugement synthétique *a priori*: c'est elle qui dicte les conditions auxquelles doit satisfaire le sensible en général pour être transformé en objet et c'est elle qui est soumise par Kant à la méthode logico-transcendantale et à une démarche que la *Critique* nomme de roanière un peu ambiguë « déduction ». [Besnier, 1993, page 189]

Dans la première édition de la *Critique* Kant nous fait part du postulat suivant qui se réfère aux jugements synthétiques:

> Dans toutes les sciences théoriques de la raison sont contenus comme principes, des jugements synthétiques *a priori*.
>
> 1) Les jugements mathématiques sont tous synthétiques.
>
> ...
>
> 2) La Science de la nature (*physica*) contient, à titre de principes, des jugements synthétiques *a priori*. [Kant, 1975, page 40–42]

Kant justifie que les jugements mathématiques sont synthétiques de la manière suivante: soit le jugement "Toute ligne droite est la plus courte entre deux points"; il dit que cela est un jugement synthétique

> Car mon concept de ce qui est *droit* ne contient rien de quantitatif, mais seulement une qualité. Le concept du plus court est donc entièrement ajouté et ne peut être tiré par aucune analyse du concept de la ligne droite. Il faut ici de l'intuition qui seule rend la synthèse possible. [Kant, 1975, page 41]

Les jugements mathématiques sont synthétiques mais ils sont aussi *a priori* car ils ne sont pas empiriques et ils comportent une nécessité que l'on ne peut pas tirer de l'expérience.

En général les positivistes logiques s'accordent avec Kant sur le fait les propositions analytiques doivent exprimer les vérités nécessaires que l'on peut considérer comme étant *a priori*, cependant ils ne sont pas d'accord pour considérer les propositions vraies des mathématiques comme synthétiques *a priori*, car au contraire ils les considèrent analytiques.

Néanmoins les positivistes logiques ont contribué à rendre plus claire la définition du caractère analytique des énoncés grâce au tournant linguistique qui a caractérisé ce mouvement. Selon eux, la vérité ou la fausseté d'un énoncé analytique sont déterminées en vertu de la signification des mots de l'énoncé et des règles de grammaire qui gouvernent leur combinaison. (Pour une lecture de la position du positivisme vis-à-vis de l'analycité voir, par exemple [Ayer, 1962]). Cette dernière définition a la vertu de pouvoir s'appliquer à des énoncés qui n'ont pas une structure en termes de sujet et de prédicat.

Une fois finie la première phase de la réduction, la deuxième est censée permettre l'identification des paramètres des *dispositions de comportement* dans les énoncés des propositions hypothétiques pour parvenir à des lois et à des généralisations formulées dans un vocabulaire appartenant à la physique ou à la psychophysique. Nous allons retourner à la notion de disposition du comportement dans le chapitre consacré au fonctionalisme étant donné que cette notion, telle qu'elle est utilisée par Gilbert Ryle (1900–1982) est considérée comme un concept inspirateur du fonctionalisme.

Les critiques adressées au behaviorisme sont de deux types: celles qui sont relatives au sens commun et les critiques techniques. [Searle, 1994b, page 34–35] Les critiques issues du sens commun portent évidemment sur le fait que le behaviorisme laisse sans explication toute la série des expériences subjectives comme penser ou sentir. Les behavioristes ont été accusés sarcastiquement de simuler l'anesthésie comme dans cette blague citée par Searle [Searle, 1994b]: le premier behavioriste dit au second après qu'ils eurent fait l'amour: "C'était très bien pour toi, mais pour moi, c'était comment? "

Du point de vue technique un des arguments les plus souvent invoqués est celui de la circularité de l'analyse du comportement. En effet, la première phase de la réduction consistant à réduire un énoncé mental comme celui de notre exemple "Jean a mal au dos" en une série de dispositions de comportement, exige une série infinie de propositions simultanément vraies pour que la traduction soit sémantiquement équivalente. Or, pour que la traduction de notre exemple en la déduction "Si on l'invite à jouer au bowling, il va refuser" soit vraie, il est nécessaire que "Jean croie que le fait de jouer au bowling aggraverait son état physique", mais également que "Jean considère que la santé vient avant le bowling" , et "Jean croit que s'il décline de jouer au bowling, ses partenaires de jeu vont le considérer comme une *petite nature*"

Toutes ces dernières propositions doivent se vérifier pour que la traduction soit pertinente, mais chacune d'elles doit aussi être traduite en énoncés sur le comportement. Or cette traduction va donner naissance iterativement encore à de nouvelles propositions.

La deuxième critique, en relation avec la précédente, vise l'impossibilité de donner une définition de la notion de *disposition*. Personne ne peut établir clairement les caractéristiques qui doivent avoir les antécédents des propositions hypothétiques déterminantes des dispositions du comportement.

La troisième critique met en relief l'impossibilité de réaliser la deuxième phase de la réduction. Elle consiste à réduire les termes relatifs aux énoncés des dispositions du comportement (comme stimulus–réponse) en termes de quantités physiques en supposant que les relations qu'elles entretiennent sont celles de la physique. Ainsi Noam Chomsky [Chomsky, 1959] dans sa critique de B. F. Skinner soutient que les concepts de stimulus, de réponse et de renforcement qui caractérisent la méthode behavioriste ne sont applicables que dans les conditions très simplifiées du laboratoire et ne sont pas pertinentes dans les conditions normales de la vie réelle. Dans des situation réelles, un même stimulus peut provoquer différentes réponses; Chomsky signale particulièrement les difficultés à prédire les comportements verbaux. La programme behavioriste n'est pas absurde, mais s'il n'avait pas échoué, selon Jacob [Jacob, 1992a], le statut ontologique du mental n'eût pas pour

---

Des critiques de cette conception du caractère analytique ont été formulées par la suite, principalement par Quine et Putnam.

Dans un texte très cité de Quine "One dogma of Empiricism" paru dans [Quine, 1963] l'auteur se pose la question si l'on est capable d'identifier clairement et sans faire des pétitions de principe les propositions analytiques. Le texte de Quine constitue un long travail de démantèlement de tous les arguments qui soutiennent l'existence du caractère analytique d'un énoncé. D'abord, il réfute la définition même de l'énoncé analytique selon laquelle le terme de sujet a la même signification que celui de prédicat. Dans son argumentation il signale l'ambiguïté du terme signification et il remarque qu'il peut être compris de façon extensionnelle (théorie de la référence) comme de façon intensionnelle (théorie de la signification) en évoquant le problème des contextes opaques (problème de Frege). Quine analyse un par un les concepts sur lesquels cette identification entre le concept du sujet et le concept de prédicat peut reposer; ces concepts sont: l'égalité de signification, la définition, ou la synonymie. Ensuite il argumente que cette relation ne peut pas non plus se baser sur des règles sémantiques. Finalement, Quine conclut que la distinction analytique–synthétique bien qu'elle soit intuitivement raisonnable ne peut pas établir une séparation précise entre les énoncés analytiques et les énoncés synthétiques. Il dit, non sans ironie, que cette distinction n'est qu'un dogme non–empirique des empiristes et un article de foi métaphysique.

Des auteurs comme Jerry Fodor et Ernest LePore entre autres nient aussi la validité de cette distinction. Ceci aura des conséquences sur la théorie fodorienne et en particulier, pour la réfutation que Fodor fait du holisme des contenus. (voir chapitre 7)

autant été éclairci. Ceci est dû au caractère ambigu des deux interprétations possibles dans ce cadre: l'interprétation réductioniste ou l'interprétation éliminativiste. Selon la première faudrait-il conclure que le mental existe mais qu'il est identique au comportement, ou d'autre part, que les conditions de vérité des énoncés mentaux dépendent des conditions de vérité des énoncés physiques? Selon la deuxième version, on pourrait déduire que le mental n'*existe pas* puisque tout le contenu des propositions mentales peut être traduit en propositions appliquées au comportement; mais alors, lorsqu'on accepte l'existence des propositions non traduisibles de cette façon –comme c'est par exemple le cas de Skinner– ces dernières doivent-elles être considérées comme vides de sens?

2.5.2 La théorie de l'identité

La théorie de l'identité peut être perçue comme la solution idéale aux problèmes présentés par le behaviorisme et par le dualisme à la fois. Les dualistes ont raison quand il affirment l'existence des états internes, mais ils ont tort quand ils les considèrent comme non–physiques. Les behavioristes par contre ont raison d'être matérialistes mais ils ont tort de nier l'existence des états internes. La théorie de l'identité admet l'existence des états internes, mais considère qu'ils ne sont pas constitués d'une substance immatérielle et qu'ils sont en fait identiques aux états ou événements qui relèvent du système nerveux central.

A la différence du dualisme des propriétés qui explique difficilement la corrélation entre les propriétés psychologiques et les propriétés physiques, le matérialisme de types qui postule leur identité est affranchi de ce problème.

Toutefois l'identité entre les état mentaux et les états physiques dans ce cadre est une identité de type ou générique (*type identity*) par opposition à l'identité particulière ou occasionnelle (*token identity*). Cette distinction entre générique et occasionnel est initialement due à Charles Sanders Pierce (1839–1914) dans le cadre de la sémiotique et fut utilisée pour signaler la différence entre les occurrences concrètes et particulières d'un signe et sa catégorie générique. Par exemple, dans le phrase "*The cat is on the mat*" il y a six mot lorsqu'on décide de compter les occurrences (*les token*) et il y en aura moins si on décide de les compter comme types. En effet, comme la phrase contient des articles, des noms, une préposition et un verbe on pourrait aussi répondre qu'il y existe quatre types de mots.

Contrairement au behaviorisme qui pratique une réduction analytique linguistique car il considère que les énoncés mentaux et les dispositions du comportement entretiennent une relation de synonymie; le physicalisme de type pratique une réduction synthétique *a posteriori*.

Comme l'a signalé U. T. Place [Place, 1956], cela signifie que l'identité entre les états mentaux et les états physiques du cerveau est perçue comme une hypothèse scientifique raisonnable, comparable à l'hypothèse qui affirme que la lumière est un mouvement de charges électriques.

Les postulats de base de la réduction peuvent être résumés en deux propositions selon Élisabeth Pacherie [Pacherie, 1993] :

1. Chaque état mental est identique à un état neurologique.

2. Chaque propriété mentale est identique à une propriété neurologique.

Le premier postulat affirme l'existence d'une identité occasionnelle entre des états particuliers, tandis que le second professe l'identité entre les propriétés mentales/psychologiques et les propriétés physiques/neurologiques.

Ceci cependant pose des problèmes parce que l'état neurologique correspondant à un état mental comme la sensation de chaleur n'aura pas pour autant la nature de la chaleur, et symétriquement, les propriétés chimiques ou électriques des états neurologiques ne peuvent pas être assimilées aux propriétés des états psychologiques correspondants.

Place (1956) compare l'énoncé "la conscience est un processus du cerveau" –(*consciousness is a process in the brain*) qui dans un langage plus à jour sera formulé: "les processus mentaux sont des processus neurologiques"– à l'énoncé "un nuage est une masse de minuscules particules en suspension". En général, l'indépendance logique des deux expressions entraîne l'indépendance

ontologique entre les états des choses auxquelles elles se réfèrent. Cependant dans les deux cas cités plus haut, cette règle ne s'applique pas. Il est clair que dans l'un comme dans l'autre, les paires d'énoncés signifient du point de vue du sens deux choses bien différentes et qu'ils n'impliquent pas une relation logique comme il en est par exemple entre roses et fleurs puisque toutes les roses sont des fleurs. Néanmoins on ne conclut pas pour autant que les deux expressions, "nuage" et "minuscules particules en suspension" sont deux choses différentes du point de vue référentiel. Bien au contraire, il y a identité entre les termes du binôme "nuage"/"minuscules particules en suspension" mais cette identité n'est pas intrinsèque et a besoin d'être vérifiée moyennant une observation visuelle. Ceci explique que dans ce cas l'identité ontologique ne soit pas inférée à partir de l'identité logique.

Ce qu'on vient d'exposer pour les nuages est aussi applicable à l'identité entre les processus mentaux et ceux du cerveau. La différence entre les deux cas est que les opérations requises pour les vérifier respectivement sont des processus fondamentalement différents.

Une des conséquences impliquées par l'identité entre les propriétés des états mentaux et celles des états neurologiques mérite d'être soulignée. La question pourrait être formulée de la façon suivante : Si les propriétés physiques sont identiques aux propriétés mentales, alors lorsqu'on éprouve la sensation de couleur, par exemple le vert, cela voudrait dire, en vertu de ce principe d'identité que l'état de cerveau correspondant est aussi vert? Pour Place ce problème n'est pas insurmontable, il suffit de ne pas commettre l'erreur logique qu'il appelle la "fallacie phénoménologique" et qui consiste à prendre comme vrais les deux présupposés suivants très fortement reliées :

1. étant donné que notre faculté à décrire des choses de notre environnement dépend des événements mentaux, les descriptions que nous faisons des événements se rapportant aux sensations sont de prime abord des descriptions ou propriétés qui caractérisent ces états mentaux et seulement après et de façon secondaire, inférentielle et indirecte des propriétés descriptives de l'environnement et des objets extérieurs.

2. que reconnaissance des propriétés phénomènologiques précède normalement l'apprentissage des propriétés réelles des objets de l'environnement.

Pour Place les deux présuppositions sont fausses. En réalité les caractéristiques réelles des objets ne sont pas dérivées des propriétés phénoménologiques car c'est le contraire qui est vrai. D'abord on apprend à reconnaître les propriétés réelles et ensuite on apprend à les décrire en utilisant des propriétés phénoménologiques. L'exemple cité par Place se réfère à un sujet qui fait un rapport sur la perception d'une *after-image* de couleur verte.

De même qu'il n'y a rien dans l'environnement qui soit vert, il n'y a rien non plus dans le cerveau de l'individu qui soit vert. Les propriétés phénoménales comme la couleur ne peuvent pas être appliquées de façon pertinente aux processus de cerveau. Quand la personne exprime qu'elle perçoit une *after image* verte, elle ne dit pas qu'il existe quelque chose, en l'occurrence une *after image* et que cette dernière est verte, mais par contre, elle dit qu'elle a une expérience qu'on éprouve d'habitude quand on voit une tâche verte et qu'on a appris à décrire de cette façon.

Trois objections ont été émises à l'encontre du matérialisme de types. Deux proviennent surtout des partisans du dualisme des propriétés et signalent premièrement l'existence des propriétés mentales qui n'ont pas de corrélations neurologiques identiques, deuxièmement l'identité des propriétés n'est pas une propriété synthétique comme les physicalistes des propriétés le prétendent mais une propriété nécessaire vérifiable a posteriori.

La troisième objection vient du côté des fonctionnalistes et souligne que le matérialisme des types restreint la cognition aux êtres munis de cerveaux, ce qui veut dire que cette position nie la multi-implantation de la cognition.

La première objection fut élevée par les dualistes de propriétés qui récusent l'identité entre les propriétés mentales et neurologiques. Ils accordent aux physicalistes que chaque état mental est identique à un état neurologique mais que malgré cela, on peut nier l'identité des propriétés. Très schématiquement l'argumentation adopte la forme suivante [Jacob, 1992a, cf.] :

(a) Les caractéristiques qualitatives phénoménologiques de mes sensations me sont directement connues par introspection.

(b) Les propriétés de mes états cérébraux ne me sont pas connues directement par introspection.

(c) Conclusion : Le contenu qualitatif phénoménologique est une propriété de mes sensations différente des propriétés de mes états cérébraux.

Soit la fonction propositionnelle "$P(x) = x$ est connu par introspection" et soient les termes a = les caractéristiques qualitatives phénoménologiques de mes sensation et b = les propriétés de mes états cérébraux, les propositions précédentes peuvent être reformulées ainsi :

(a) $P(a)$

(b) $\neg P(b)$

(c) Conclusion : $a \neq b$

Mais la conclusion (c) s'ensuit des deux prémisses précédentes uniquement lorsqu'on admet le *principe Leibnizien de l'indiscernabilité des identiques* . Ce principe était pour Leibniz une des bases de son calcul logique bien qu'il ait donné plusieurs versions de cette loi entre 1679 et 1690, mais on peut l'énoncer ainsi :

> ***Deux entités sont identiques si l'une d'elles peut se substituer à l'autre dans toute proposition sans en changer la valeur de vérité.***

Ce qu'on peut traduire de la façon suivante :

$$\forall x \forall y [(x = y) \Rightarrow \forall P[P(x) \equiv P(y)]]$$

La réfutation de la loi de Leibniz est bien connue et appartient à Frege : si les expressions l'*étoile du soir* et l'*étoile de matin* sont considérées comme égales car elles font référence à la même chose (la planète Vénus), alors les propositions suivantes devront avoir la même valeur de vérité.

"Jean croit que l'*étoile du matin* est une planète."
"Jean croit que l'*étoile du soir* est une planète."

Cependant, il est possible que Jean ne sache pas que l'*étoile du matin* et l'*étoile du soir* sont la même planète, ainsi donc il peut croire qu'une des deux n'est pas une planète. Les deux propositions n'ont donc pas forcément la même valeur de vérité, bien que leur objet direct soit le même du point de vue référentiel. Ceci montre l'existence de deux types de propriétés, d'un côté les propriétés *extensionnelles ou de contexte transparent* et de l'autre les *propriétés intensionnelles ou de contexte opaque*. Or le principe Leibnizien de l'indiscernabilité des identiques est uniquement valable pour des propositions qui expriment des propriétés extensionnelles. Pour arriver à démontrer ceci, les physicalistes des types reproduisent le type d'argumentation proposé plus haut avec mais en utilisant l'exemple de Vénus. Soit le fonction propositionnelle $P(x)$ = "Jean croit que x est une planète" et soient les termes a = l'étoile du matin et b = l'étoile du soir :

(a) Jean croit que l'étoile de matin est une planète.

(b) Jean ne croit pas que l'étoile du soir soit une planète

(c) Conclusion : l'étoile de matin $\neq$ l'étoile du soir

On voit bien que la conclusion ainsi obtenue est manifestement fausse. Or, selon les physicalistes des types, on ne peut pas assurer à partir de l'application des lois de Leibniz qu'il existe des propriétés mentales qui ne sont identiques à aucune propriété p physique comme les dualistes de propriétés veulent le faire croire. La proposition

$$P(x) = x \text{ est connue par Jean}$$

n'exprime pas une propriété intensionnelle aux yeux des physicalistes d'où le fait que le principe de Leibniz ne soit pas applicable dans ce contexte.

J'entends maintenant soutenir qu'il existe une autre différence non encore signalée entre les termes suivants: "les propriétés de mes états cérébraux" et "la propriété de qualia d'un état mental". La première n'est pas une propriété intrinsèque tandis que c'est le cas de la deuxième (être un qualia). Le concept d'*intrinsèque* est souvent opposé au concept de *relationnel* . Ainsi, la lune a *intrinsèquement* une masse, mais ce n'est pas intrinsèquement un satellite; c'est un satellite seulement par rapport à la Terre. [Searle, 1994b, cf. page 80]. La propriété du cerveau d' avoir la fibre–C ou la fibre portant le numéro 1895 n'est pas une propriété intrinsèque, mais une propriété relationnelle parce qu'elle existe seulement en relation à un état de choses donné dans le cadre du physicalisme des types. Or les propriétés intrinsèques du qualia ne pourront pas être comparables aux propriétés relationnelles du cerveau, parce qu'elles sont intrinsèques à l'expérience même, ce qui veut dire non analysables et essentielles.

Je pense que ceci justifie une réfutation de l'identité des propriétés puisqu'il y a des propriétés des états mentaux qui sont intrinsèques et qu'on ne peut pas affirmer la même chose des propriétés physiques.

La deuxième objection ressort du sens commun et elle est due à Saul Kripke [Kripke, 1971] . Il utilise un argument modal et prétend qu'il est impossible d'identifier quelque chose de mental avec quelque chose de physique sans laisser de coté le mental. Kripke, dans son argumentation affirme que si les partisans du physicalisme des types veulent défendre l'identité des états mentaux et des états neurologiques par l'analogie avec des identités scientifiques telles que "nuage" et " minuscules particules en suspension" ou "la chaleur" et "des mouvements moléculaires", ils doivent accepter que cette identité est une identité nécessaire et non une propriété empirique. Ainsi, Kripke réfute que l'identité des états mentaux et des états physiques soit une identité contingente.

Il prend l'exemple de la chaleur; il est possible d'imaginer une monde où il se passe l'une des deux choses suivantes: ou bien les mouvements des molécules qui produisent la chaleur dont nous éprouvons la sensation au toucher ne nous donne pas cette sensation ou bien nous pouvons percevoir la même sensation de chaleur à cause d'événements autres que le mouvement des molécules. Il est donc possible d'imaginer des mondes différents de l'actuel où "chaleur" et "mouvements de molécules" ne s'avèrent pas identiques. Ces expériences de la pensée sont telles qu'elles nous donnent l'illusion d'identité contingente. Cependant, cette identité ne peut pas être contingente mais seulement nécessaire étant donné que le but des sciences est de trouver des essences.

La "chaleur" doit signaler le "mouvement de molécules" dans tous les mondes possibles [21] où ces deux objets peuvent exister, ce qui les définit comme étant des *désignateurs rigides*. .[22]

---

[21] Les mondes possibles

Pour parler d'une situation donnée ou d'un état de choses, les logiciens parlent de mondes possibles. L'état actuel des choses n'est qu'un des membres de l'ensemble des tous les mondes possibles. Ainsi les mondes possibles donnent un cadre à ce que nous avions appelé le contexte et permettra en quelque sorte de représenter les concepts d'*intension* et d'*extension* de Frege. Maintenant, dans la théorie des mondes possibles, les expressions contenant des opérateurs modaux ont un sens plus clair. Par exemple la phrase: Φ sera nécessairement vrai dans le monde actuel si et seulement si elle est vraie dans tous les mondes possibles ou la phrase: Il est possible que Φ soit vrai si et seulement si elle est vraie dans un monde possible au moins. Cependant, pour beaucoup de philosophes du langage, le concept de mondes possibles n'éclaire pas le concept de nécessité, voire même le rend plus obscur.

L'intension, l'extension et les mondes possibles

Le premier essai de formalisation de la notion de sens (intension) est dû à Carnap, qui utilise le concept de mondes possibles. De même que la signification d'une expression est déterminée par son extension, Carnap a suggéré que le sens d'une expression (son intension) est une fonction qui, partant de l'ensemble de tous les mondes possibles, donne pour chacun d'eux l'extension de l'expression. Par exemple, l'extension de l'*étoile du matin* est Vénus, l'extension du *nombre des planètes* est neuf puisqu'il y a neuf planètes dans le monde réel. Cependant, dans le monde du Petit Prince de Saint Exupéry, l'extension de la dernière expression est très différente. C'est–à–dire qu'on est obligé de donner comme valeur sémantique à la phrase "le nombre de planètes", une extension pour chacun des mondes possibles. Bref, l'intension d'une expression n'est que la somme de toutes les extensions possibles que cette expression peut avoir, mais assemblées de façon organisée comme une fonction ayant comme arguments tous les mondes possibles et comme valeurs les extensions correspondantes. Kripke a bâti sa sémantique pour logique modale en prenant les mondes possibles comme indices et il a dès lors été possible de donner une définition formelle de l'intension pour les langages formalisés.

[22] [Kripke, 1971, cf. page 42],[Quine, 1981, cf. page 42].

Le même critère est applicable aux propositions telles que: "La douleur est tel et tel état neuronal du cerveau", défendues par le partisans du physicalisme des types. En effet, cette proposition peut paraître contingente pour les raisons suivantes: premièrement on peut s'imaginer être dans cet état neuronal et malgré cela ne sentir aucune douleur. Deuxièmement, on peut même imaginer un agent sans cerveau (par exemple un agent artificiel) qui éprouve de la douleur. On peut donc imaginer, par analogie à l'exemple de l' identité "chaleur/mouvement de molécules", des situations où l'identité "douleur/état du cerveau pertinent" ne se vérifie pas. Cependant, ces situations, bien que plausibles ne sont pas acceptables par un partisan du physicalisme des types parce que s'il les reconnaissait possibles, il devrait alors nier que le but de la quête scientifique soit de trouver des essences et non des situations contingentes. En fait, ce qui nous fait croire qu'elles sont contingentes est que nous appréhendons cet état neuronal à travers une propriété contingente. En effet, ce qui est contingent est *ma* douleur et non la douleur en général. Ainsi, on a déjà prouvé que les états mentaux de douleur sont des désignateurs rigides des états neuronaux.

Néanmoins la démonstration inverse reste à faire, c'est–à–dire, "cet état neuronal (celui pertinent à la douleur) est identique à la douleur".

Mais cette proposition est évidente parce que l'expérience en soi doit être cette expérience–ci, on ne peut pas dire que c'est une propriété contingente de la douleur que je suis en train d'éprouver.

Or, les situations précédentes ont montré que la douleur est un désignateur rigide de cet état neuronal et l'inverse aussi est vrai. Donc, conclut Kripke, il ne s'agit pas d'une hypothèse scientifique mais bel et bien d'une proposition nécessaire.

Cependant si les propositions identitaires postulées sont des propositions nécessaires, les physicalistes des types se trouvent face au problème suivant: d'un côté, si une proposition est nécessaire alors sa négation est impossible et de l'autre les situations où l'on *pourrait* avoir de la douleur sans être pour autant dans l'état neuronal pertinent ne nous semblent pas impossibles. Il faut qu'on nous démontre que ces situations qui nous semblent parfaitement plausibles ne le sont pas. Mais cette démonstration ne peut être faite par analogie avec l'exemple de l'identité de "chaleur"="mouvement de molécules". Ainsi l'argumentation modale développée par Kripke et de son propre aveu n'est pas réfutée. Or les outils analytiques vont à l'encontre de l'identité de type entre états mentaux et états physiques. Pourtant Quine [Quine, 1981, cfr. page 174], qui a critiqué non seulement le concept d'énoncé analytique mais aussi tous les essais de définir le concept de vérité nécessaire, nous fait voir que ce problème ne se pose que si les physicalistes de types croient en la nécessité métaphysique.

La troisième objection a été formulée par Hilary Putnam et Jerry Fodor[23] au début des années soixante et s'adresse aux deux questions suivantes:

La première concerne la difficulté de trouver les mêmes propriétés neuronales (physiques / chimiques) dans des systèmes nerveux appartenant aux espèces qui, tout en ayant la faculté d'éprouver la douleur par exemple, ont des structures physico / chimiques très différentes; par exemple trouver les mêmes propriétés physiques dans les mammifères, les reptiles et les crustacés comme les calamars. Par contre, cet état physique du cerveau ne saurait exister en aucune créature physiquement concevable qui soit incapable d'éprouver cette sensation, mais chaque fois qu'on le détecte, on peut être certain qui l'individu qui le possède éprouve de la douleur même s'il s'agit d'un extraterrestre. Une telle hypothèse, bien qu'audacieuse, ne peut être considérée totalement impossible.

Cependant, il suffirait pour contredire la thèse de l'identité des propriétés qu'un même état psychologique par exemple "avoir faim" soit concevable pour deux espèces très différentes. Putnam prend le cas des mammifères et des pieuvres dont la comparaison des propriétés physico–chimiques respectives exclut l'identité. Dès lors la théorie se trouve infirmée. Comme Putnam le signale, il est tout à fait probable qu' on parvienne à faire une telle démonstration.

La seconde question porte sur le physicalisme des types que Block a appelé le "chauvinisme neurologique". En effet, cette hypothèse restreint l'assignation des états mentaux uniquement aux être biologiques ayant des neurones comme les nôtres, ce qui est incompatible avec l'argument de multiplicité d'implantations de la cognition . La multiplicité d'implantation a été une thèse centrale aux conceptions de sciences cognitives, surtout dans le cadre de la théorie représentationelle de

[23] Voir [Putnam, 1960],[Putnam, 1967a],[Fodor, 1968], [Fodor, 1981b]

la cognition, néanmoins il y des voix comme celle de John Searle récemment, que s'élèvent contre cette conception.

En résumé, nous avons jusqu'ici exposé les deux courants réductionnistes du mental, le behaviorisme et le physicalisme des types. La première théorie veut réduire le mental aux dispositions du comportement; pour la deuxième cette réduction consiste à identifier les propriétés et les états mentaux aux états et propriétés physiques du cerveau. Toutes deux ont échoué pour les raisons que j'ai exposées plus haut. Mais, outre la stratégie réductionniste que toutes deux utilisent, quand bien même leurs programmes respectifs auraient réussi, le statut ontologique n'aurait pas été établi pour autant. Dans le cas du physicalisme des types on ne sait pas si la réduction se rapporte à un type de réduction ontologique faible ni si les termes mentaux ont le même degré d'acceptation que les termes physiques et si l'utilisation d'un ou de l'autre type est une affaire de choix dans un contexte donné. On ignore également si la réduction implique qu'il faut remplacer les termes mentaux par des termes physiques, étant donné que ces derniers sont plus adéquats.

2.5.3 Les positions matérialistes éliminativistes

Le but des positions éliminativistes qui considèrent la psychologie ordinaire erronée est de rejeter toute plausibilité ou explication aux concepts qui en découlent. Elles argumentent que la conception des états ou processus auxquels ces énoncés font référence n'est qu'une erreur ontologique, d'où le besoin de remplacer cette conception erronée par une autre qui soit correcte. Une des théories éliminativistes propose une approche neurologique donc matérialiste. Ses représentants sont R. Rorty, Paul K. Feyerabend , Paul M. Churchland et Patricia Smith Churchland. Le premier à soutenir une position contraire aux termes mentaux du sens commun a été Paul K. Feyerabend [Feyerabend, 1963] qui signale que la tentative de bâtir des lois qui fassent office de ponts entre termes mentaux et termes physiques afin d'opérer une réduction est erronée, étant donné qu'on essaie de perpétuer une terminologie ancienne au sein d'une théorie nouvelle.

Les critiques que Feyerabend fait se basent sur le fait que ces lois ne sont intérieures ni à la théorie réductrice ni à la théorie réduite car ces relations d'identification se font par rapport à des faits externes aux deux théories en question. Ainsi les lois-ponts ne suffisent pas, parce que bien qu'on opére la réduction on ne connait toujours pas les référents de ces termes.

En outre le changement de vocabulaire ne change en rien la valeur empirique de la nouvelle théorie. Cependant, la théorie nouvelle n'est pas, malgré cela, condamnée à échouer. Et il ajoute :

> Finalement, une théorie physiologique de l'épilepsie ne va pas devenir une tautologie vide du fait qu'elle n'utilise pas la phrase – ou la notion – *possédé par le diable, diable* dans le sens 'théologique' du terme. Il y a assez de prédictions disponibles, plus que toutes celles que le mentalisme a jamais pu fournir. [Feyerabend, 1963, page 205]

Selon Feyerabend l'utilisation des théories empiriques comme la psychologie naïve pour mesurer la valeur des théories physiologiques n'est pas soutenable étant donné les conditions sur lesquelles reposent les lois-ponts et je viens de le signaler. En plus, Feyerabend soutient que les théories physiologiques doivent être en mesure de créer leur vocabulaire *ad hoc* sans avoir besoin de se référer aux notions des théories en psychologie.

Les Churchland ont exposé à plusieurs reprises leur conviction que la psychologie ordinaire est une théorie empirique "radicalement fausse dont l'ontologie n'est qu'une illusion" [Churchland, 1981, page 206]. Les raisons qu'ils citent sont les suivantes : Bien que la psychologie ordinaire nous soit utile pour prédire les comportements des autres dans la vie de tous les jours, elle est très incomplète parce qu'il y a des phénomènes pour lesquels elle est incapable d'avancer la moindre explication. L'ignorance de la psychologie ordinaire est flagrante à l'égard des sujets fondamentaux comme la nature et la dynamique des maladies mentales, la capacité de création et d'imagination, les différences entre l'intelligence des êtres, la nature du processus d'apprentissage. En effet, elle n'est pas capable d'expliquer comment l'enfant apprend et emmagasine toutes les croyances et tous les désirs sur lesquels ses propres théories sont fondées. Mais ce n'est pas là ce qui fait de la psychologie ordinaire une théorie radicalement fausse; le problème fondamental tient à ce que les catégories employées dans l'explication sont complètement orthogonales aux catégories de la physique, d'où l'impossibilité absolue de formuler une réduction.

Un autre défaut de la psychologie ordinaire est la stagnation dont cette théorie souffre depuis trois millénaires, car aux yeux des Churchland elle est restée figée depuis l'époque de la Crèce antique. Les interprétations qu'elle donne en termes de croyances et de désirs ne sont, en définitive que les prolongements d'un certain syncrétisme, c'est–à–dire, qu'elles reflètent les attitudes typiques des cultures primitives pour expliquer les phénomènes de la nature. Ces attitudes consistent fondamentalement à attribuer des propriétés intentionnelles aux éléments naturels, par exemple: "Le vent peut connaître la colère, la rivière la générosité, la mer la fureur, et ainsi de suite" [Churchland, 1981, page 211]. La psychologie ordinaire est aussi proche de la psychologie scientifique que l'alchimie de la physique scientifique. En résumé,

> [...] la psychologie ordinaire n'est ni plus, ni moins qu'une théorie ancrée culturellement sur la façon dont nous et les animaux supérieurs fonctionnons. Il n'y a aucune caractéristique spéciale qui la rende empiriquement invulnérable, aucune fonction unique qui la rende irremplaçable, elle ne jouit d'aucun statut spécial.[Churchland, 1981, page 218, ma traduction]

Les Churchlands croient que le matérialisme éliminativiste tel qu'ils le pratiquent n'est pas contraint à donner une explication naturaliste de la cognition, c'est–à–dire d'expliquer scientifiquement l'efficacité causale ou le caractère normatif des états mentaux. Par contre, le but de la démarche sera de donner une explication de l'organisation fonctionnelle neurologique des systèmes cognitifs, celle–ci étant regardée comme l'essence même de la cognition.

Les Churchlands ont aussi à l'appui de cette thèse une vision différente de la démarche scientifique; en particulier ils sont sceptiques quant à l'idée reçue que l'avance de la science est uniforme et sans faille. En effet, Paul Churchland a exposé ses réserves au sujet du réductionnisme nagelien, même à l'égard des réductions considérées homogènes. Pour lui la méthode consistant à proposer des lois ponts (*bridge laws*) chargées d'opérer la corrélation entre les termes descriptifs des deux théories telle qu'on l'a exposée dans le chapitre précédent est erronée parce que, de cette façon nous risquons d'inclure dans la théorie réductrice des termes dénués de toute référence. Nous rappelons qu'une réduction à une théorie nouvelle (TN) d'une ancienne (TO) peut être exprimée comme suit:

TN & (lois ponts) impliquent logiquement TO

Mais pour lui les lois ponts sont en général fausses; il n'est pas vrai par exemple, si nous prenons la réduction de la thermodynamique classique à la thermodynamique statistique, que les gaz réels obéissent aux lois telles que $PV = \mu RT$ vérifiée dans le cas classique. De même, dans la réduction des lois de l'astronomie de Kepler à la dynamique de Newton, il n'est pas vrai que l'orbite des planètes soit elliptique. Il faut borner les lois ponts aux hypothèses qui limitent leur champ d'application. Or on voit que le premier exemple d'équation– pont se vérifie si on suppose que les molécules d'un gaz réel n'ont que de l'énergie mécanique et dans le deuxième cas si l'on admet que la masse des planètes est négligeable vis–à–vis de la masse solaire.

Tout cela justifie pour Churchland la récusation de la réduction nagelienne puisque l'adoption des lois ponts sans plus peut nous amener à valider une théorie dont l'ontologie est inexistante. En effet,

> [...]les caractéristiques conçues comme nouvelles [dans la théorie réductrice] peuvent ne pas être identiques, ou même ne pas être connectées du point de vue nomologique avec les caractéristiques anciennes [relevant de la théorie à réduire] si les anciennes sont illusoires et inapplicables en réalité (*unsubstantiated*).[Churchland, 1992b, page 48]

Toutes ces affirmations témoignent du scepticisme qui anime Paul Churchland quant à la thèse du progrès sans mélange de la science; pour lui la nouvelle théorie ne doit pas réduire l'ancienne sans plus, sinon opérer une espèce de révision. Il propose une nouvelle méthode qui consist à bâtir au sein de la nouvelle théorie une analogue de l'ancienne. Ce schéma de réduction intrathéorique est exprimé par Paul Churchland [Churchland, 1992b] comme suit :

TN & des hypothèses qui limitent & les au bord
implique logiquement
IN (un ensemble des théorèmes dans la théorie TN)

par exemple

$$\forall x(Ax \supset Bx)$$
$$\forall x\,((Bx\&Cx) \supset Dx)$$

qui doivent être isomorphiques avec

$$\forall x(Jx \supset Kx)$$
$$\forall x\,((Kx\&Lx) \supset Mx)$$

Selon Churchland cette réduction est une *déduction* puisqu'on prend comme prémisses non les théorèmes de TO n'appartenant pas à TN mais leurs images à travers une relation isomorphique et ce n'est que plus tard que les paires de correspondances font leur apparition. Ce ne sont pas des implications matérielles comme dans les cas des lois ponts mais seulement des paires i.e.: $< Ax, Jx >$

La projection de TO dans IN et la déduction de IN dans TN apportent la garantie d'exactitude et de prévisibilité de la nouvelle théorie et ces démarches constituent les bases pour la révision de l'ancienne. Il existe deux types de réduction : les réductions douces et les réductions nettes. Pour les premières l'ontologie de la théorie réduite peut se conserver presque totalement dans la réductrice et dans ce cas, les hypothèses limitatives ne sont pas des énoncés contraires aux faits tandis que pour la plupart des principes de TO on peut trouver des analogues en IN. Dans le cas de réductions nettes, l'ontologie de TO est revue et dans certains cas elle en est totalement éliminée. La différence principale entre cette réduction et la réduction nagelienne est que dans celle–ci, ce sont les réductions douces qui soutiennent les énoncés inter–théoriques et non l'inverse comme c'est le cas de la réduction nagelienne qui d'emblée propose les lois–ponts et passe d'une théorie plus simple à une autre plus complexe mais sans mettre en doute la valeur explicative de l'ancienne.

Churchland nie qu'une explication purement neurologique soit incomplète sous prétexte qu'elle n'inclut pas des phénomènes tels que les expériences qualitatives et il récuse aussi l'argument de la connaissance de Frank Jackson cité plus haut. Rappelons que l'objection de cet argument au physicalisme en général est que, même en ayant exhaustivement du point de vue neurologique toutes les données relatives, par exemple la perception d'une couleur, on n'arrive pas à savoir effectivement ce que c'est que de voir une couleur de plus que le commun des mortels .

Cependant pour les Churchlands ceci n'est pas une critique pertinente. Ils disent qu'un agent averti, tel qu'un spécialiste en neurosciences devra être capable de s'imaginer des états internes comme la perception d'une nouvelle couleur à partir d'une description détaillée des états physiques du cerveau. Pour eux une reconceptualisation dans ce sens n'est pas impossible. Ainsi, par exemple ce même agent averti pourrait identifier par la seule introspection la perception du rouge d'une tomate comme correspondant à une fréquence de 90 Mhz d'un faisceau gamma et même sans avoir jamais eu cette sensation, pourrait s'imaginer l'état interne résultant. Ils comparent cette capacité avec la capacité du musicien d'entendre dans son imagination le son d'un accord qu'il n'a jamais entendu auparavant. On peut même jouer à la fois plusieurs accords dont certains contiennent ce son, et le musicien est capable de dire si le son en question forme partie de l'accord ou non. Ainsi le physicalisme neurologique récuse l'objection à l'argument de la connaissance et aussi à la conséquence qui en découle, c'est–à–dire qu'un grand nombre d'événements mentaux, comme par exemple les sensations qualitatives, ne sont pas prises en compte. Pour eux :

> L'avènement d'une véritable kinesthésie et d'une théorie dynamique matérialiste pour expliquer les états psychologiques et les processus cognitifs n'entraînera pas une éclipse de la vie intérieure, mais plutôt permettra d'en dresser le tableau dans lequel sa merveilleuse complexité sera enfin *révélée* – plus directement, si on s'y applique, que par l'introspection consciente. [Churchland, 1992b, page 66]

Je pense que cette position ne parvient pas à donner un vrai compte-rendu du mental; l'explication basée sur des énoncés physiologiques ressort d'un certain niveau d'analyse qui, bien

qu'intéressante voire fondamentale n'est pas pertinente à un autre niveau. La démarche proposée par cette école équivaut à vouloir expliquer le fonctionnement d'un moteur électrique par la description du mouvement des électrons qui parcourent les conducteurs et l'électroaimant, sans dépeindre l'ensemble du système. Si on nous donne la description dynamique complète des molécules composant un moteur électrique et qu'on nous pose la question : A quel objet correspond cette description? Nous aurons du mal à rapporter l'information à son objet, tandis que si l'on nous donne une description fonctionnelle macroscopique des composantes, nous arriverons plus facilement à l'attribuer, parce que ces explications-là sont plus proches du phénomène global. Un exemple célèbre est dû à Putnam [Putnam, 1975b] qui l'applique à la réfutation du réductionnisme matérialiste mais qui peut servir à illustrer mon point de vue sur le problème des niveaux d'explication ou granularité : Comment rendre compte du fait qu'un dé de 0,45 cm de coté ne passe pas à travers un trou circulaire de 0,5 cm de diamètre? On peut l'expliquer en invoquant la mécanique quantique et se fonder sur la composition atomique particulière du dé et sur la composition atomique des bords de chaque orifice. On peut aussi proposer une explication non physique basée sur les propriétés géométriques des surfaces respectives du dé et des deux orifices. Cette dernière explication est plus simple que la précédente et elle se prête mieux à la description du phénomène global qui est celui qu'il nous intéresse expliquer.

Une explication neurologique, aussi intéressante soit-elle, ne suffit pas et ceci n'est qu'un problème de granularité de l'explication. Pierre Jacob [Jacob, 1992a] dit que à la question "Quelle est la composition physico-chimique d'une croyance? " les matérialistes éliminativistes concluent à l'inexistence pure et simple des croyances. Mais on ne peut pas conclure qu'un terme comme croyance est dénué de référence parce qu'il ne permet pas de réponse à une question mal posée car simplement absurde.

Je pense qu'il est très difficile à ce point d'imaginer comment les physicalistes éliminativistes arriveront à décrire les phénomènes qui sont l'objet de la psychologie et dont ils doivent rendre compte en éliminant toute référence au vocabulaire mental. Prenons le cas du dé de Putnam; le physicaliste accepterait de donner une explication de la situation parce qu'un dé est un objet physique et donc pour lui non-vide de signification. Dès lors on peut supposer qu'il acceptera de fournir une explication de ce phénomène en dépit du fait qu'il ne sera probablement pas d'avis d'utiliser le concept de dé comme un objet pour l'explication du phénomène. Voyons maintenant un exemple dans le champ de la psychologie, outre des questions du type général: - quelle origine la schizophrénie a - t - elle? , Qu'est-ce que l'intelligence? , Un physicaliste éliminativiste accepterait-il d'expliquer un symptôme particulier appartenant à la psychose comme par exemple celui d'un psychotique qui croit qu'il est Dieu. Ai-je le droit ne serait-ce que de poser le problème dans ces termes? Mais ceci n'étant pas le cas, comme pourrais-je faire autrement? L'autre méthode acceptable serait de donner la description de l'état neurologique de l'individu en question, mais dès ce moment-là, je n'aurais plus besoin de poser la question puisque la réponse me serait déjà connue.

2.6 Conclusion

Les positions matérialistes que j'ai présentées sont les résultantes du projet positiviste et elle doivent adopter une stratégie pour résoudre le trilemme classique du problème corps - esprit. Les deux premières positions matérialistes exposées ont appliqué la stratégie de l'identité : le behaviorisme philosophique propose qu'il y a identité logique ou analytique entre les propriétés mentales et les propriétés physiques représentées par des dispositions du comportement. Le physicalisme des types soutient la double identité des états et des propriétés mentales et physiques, mais cette identité n'est pas logique ou nécessaire mais bien empirique ou synthétique. Finalement, le physicalisme éliminativiste neurologique nie l'existence des états/propriétés mentaux et propose une explication des événements mentaux en termes purement et exclusivement neurologiques, en récusant toute allusion possible aux termes de croyances, désirs.

J'ai déjà fait état des critiques du matérialisme émises par les dualistes des propriétés portant sur le fait que le compte-rendu physique formulé à la troisième personne de certains événement

mentaux n'arrive pas a décrire totalement ces derniers et que ceci est dû à l'irréductibilité des propriétés mentales aux propriétés physiques. Dernièrement, John Searle [Searle, 1994b] considère ces arguments irréfutables. Néanmoins il ne pense pas qu'ils suffisent à démontrer l'irréductibilité du mental au physique. Pour Searle, le problème est mal posé parce qu'on ne peut redéfinir des termes comme "chaleur" par la phrase "mouvements cinétiques des molécules" tel qu'on le fait dans le cadre de la réduction de la thermodynamique classique à la thermodynamique statistique, ni redéfinir "avoir mal" par "la décharge de la fibre-c dans le cerveau" selon la réduction matérialiste. Dans le premier cas ce qu'on essaie de redéfinir ce sont des phénomènes appartenant à la *réalité physique objective* tandis que dans le deuxième il s'agit de l'*apparence subjective des choses*. Ainsi, dans le premier cas la question de comment on sent la chaleur, c'est-à-dire la question de la connaissance épistémique, ne se pose pas puisque l'élimination de l'aspect subjectif de l'objet est possible en ce contexte. Cependant, lorsqu'il s'agit par exemple de sensations qualitatives, cette élimination ne peut pas être opérée, parce que la réalité des ces phénomènes est justement leur apparence [Searle, 1994b, page 122]. Or, pour Searle toutes les réductions qui laissent de côté les bases épistémiques, les apparences ne seront jamais pertinentes pour donner le compte-rendu des phénomènes qui sont basés justement sur les sensations qualitatives. Mais, que le mental ne puisse pas être réduit ne tient pas à des aspects mystérieux des états mentaux, sinon tout simplement à ce que la réduction qu'on a utilisée jusqu'à présent n'est pas adéquate. Le mental reste un mystère empirique mais guère plus que ne l'était l'électromagnétisme lorsqu'on n'avait que les principes de Newton pour l'expliquer. Si on est capable d'opérer une révolution intellectuelle majeure qui donne naissance à un concept de réduction différent de celui qu'on a jusqu'à maintenant, alors le mental pourra enfin être réduit.

Dans un des chapitres suivants je vais présenter les théories fonctionnelles du problème corps-esprit mais auparavant je les reformulerai du point de vue des thèses de l'intentioonalité de Franz Brentano.

Brentano est celui que a essayé de fonder les bases d'une science psychologique autonome vers la fin du siècle passé et au déhut de celui-ci. Dans le chapitre suivant je vais exposer les traits fondamentaux de cette théorie. Il ne démentait pas la possibilité de réduction des propriétés mentales aux propriétés physiques, et ceci a donné lieu à une thèse qui porte son nom.

Chapitre 3

L'intentionnalité

Ce qui nous importe surtout, c'est moins la quantité et l'universalité des thèses que l'unité de doctrine. Notre but est de réaliser dans cet ordre d'idées ce que la mathématique, la physique, la chimie et la physiologie ont déjà réalisé avec plus ou moins de retard, c'est-à-dire trouver un noyau de vérité généralement admise, autour duquel, grâce au concours de forces nombreuses, ne tarderont pas à venir s'agréger de toutes parts de nouvelles cristallisations. Aux 'psychologies', nous chercherons à substituer une 'psychologie'.
Franz Brentano 1847

3.1 Introduction

Dans le chapitre précédent j'ai exposé quelques caractéristiques de la psychologie ordinaire en montrant qu'elle attribue à ses énoncés – les croyances, les désirs et les intentions – une valeur non seulement explicative mais aussi causale. Ces énoncés qui se réfèrent aux états mentaux representationnels reçoivent, en général, le nom d'attitudes propositionnelles et il sont formalisés comme une relation entre l'agent et l'objet représenté et ce en vertu du fait qu'ils ont un rôle causal. On appelle ces états des états intentionnels.

Le concept d'intentionnalité a fait un parcours considérable pour réunir un consensus –du moins pour l'essentiel– autour de son interprétation comme c'est le cas maintenant. En général des questions tantôt épistémologiques, tantôt ontologiques de caractère récurrent mais qui ont emprunté des formes différentes selon les époques, soulignent les difficultés auxquelles il faut faire face lorsqu'on essaie de formuler un compte rendu de ce concept.

Dans ce chapitre, je vais me référer à ce parcours menant au concept proposé par Franz Brentano qui a positionné l'intentionnalité en tant que marque distinctive du mental. Ensuite, j'exposerai la thèse de Brentano qui lance un défi à l'idéal d'unité positiviste de la science puisqu'elle souligne l'impossibilité de réduire les concepts mentaux aux concepts de la science physique.

3.2 Le concept d'intentionnalité

La première signification donnée au terme *intention* (*intentio* de *in* et *tendere*) a été l'activité (désir, volonté, aspiration) de l'agent en relation à un objet vers lequel cet acte se dirige. Cette acception primitive est encore reconnue pertinente de nos jours. Cependant, cette signification n'a pas été constante tout au long de l'histoire. La relation étroite entre l'intention et la connaissance a fait perdre à celle là ce caractère pour ainsi dire actif qu'elle présuppose pour remplacer le concept d'intention par l'objet même de la connaissance en laissant l'acte de côté.

Le néoplatoniciens arabes ont utilisé ce terme d'intentionnalité pour designer la relation de la connaissance avec son objet en appelant intentions les concepts.[1]

Dans cette même ligne, Avicenne [2] faisait une différence entre la logique d'un côté et les sciences réelles d'un autre. Il définissait comme objet de ces dernières les intentions premières (*intensiones primo intelectae*) parce qu'elles se réfèrent aux choses réelles. A l'opposé, la logique a comme objet des intentions secondes (*intensiones secondo intelectae*) parce que les concepts se réfèrent à d'autres concepts.

Les scolastiques apportent une nouvelle distinction. Ils différencient deux aspects : l'acte d'application de l'esprit vers un objet de connaissance et l'objet lui–même. Le concept scolastique d'intentionnalité distingue donc l'*intentio formalis* qui est l'application de l'esprit à un objet de connaissance, de l'*intentio objetiva* qui est le contenu même de la pensée auquel l'esprit s'applique.

Saint Thomas d'Aquin (1224–1274) utilisa le concept d'intentionnalité, dans le cadre de sa théorie de la connaissance; l'intention est assimilée à l'*espèce intentionnelle* qui est considérée *similaire* à la chose pensée. Pour lui, l'*intentio intellectica* est définie comme objet immédiat de la connaissance qui est une réalité intermédiaire entre la connaissance et l'objet connu. [3]

Cependant, plus tard vers la fin du XIIème et le début du XIIIème siècle l'idée de "similarité" dans l'acte cognitif, soit–elle une copie ou une image, est rejetée parce qu'on considère que c'est l'objet même et non son espèce qui est présenté aux sens et à l'intellect. Ainsi, vers le XIVème siècle Durand de Pourçain et Pierre Auriol ont récusé l'existence de l'espèce en affirmant que l'objet de la connaissance est la chose même et non une image de cette dernière. Avec une nuance pour Pierre Auriol: l'objet de la connaissance est la chose même mais en tant qu'être intentionnel ou objectif (*esse objectivum ou esse intentionale*). Cet *être intentionnel* est la manifestation de la chose à l'intentionnalité cognitive de l'esprit mais il ne doit pas être considéré comme une entité psychologique parce qu'il a un statut ontologique différent: c'est un concept objectif possédant un *esse* intentionnel.

Guillaume d'Occam (1285–1347), amateur des ontologies austères[4], considère l'*être intentionnel* comme un intermédiaire inutile entre l'esprit et la chose. Il estime que l'acte cognitif même

[1] La différence entre la théorie des concepts et la théorie des objets se base sur le fait qu'elles concernent des domaines différents du savoir philosophique. La première appartient au cadre épistémologique et de la théorie de la pensée tandis que la deuxième relève du cadre ontologique ou métaphysique. Ce contraste met en évidence une des principales caractéristiques attribuées aux *concepts* : ils appartiennent au domaine interne –c'est–à–dire à l'esprit– et peuvent être considérés comme des manières différentes qu'a un agent pour se représenter un objet, une propriété ou une relation dans l'esprit. Par exemple, l'héroïne de Superman, Louise Laine se représentait un de ses collègues de travail parfois par le concept de Klark Kent, parfois comme Superman. Cependant, elle ne savait pas que tous les deux étaient la roême personne. Or, les concepts ont une autre caractéristique: l'opacité.

[2] Le terme d'*intentio* traduirait le termes de *ma'qal* et *ma'na* que l'on trouve dans les écrits des philesophes Al-Farabi et Avicenne. Ces termes seraient à leur tour, des traductions du grec *noema*. [Pacherie, 1993, cf. page 8 note pie de page 2]

[3] La position de Saint Thomas d'Aquin est proche de la théorie péripatéticienne des espèces intentionnelles (*species intentionales*) car toutes deux considèrent ces dernières comme intermédiaires dans le processus de la connaissance. Les espèces intentionnelles sont des objets internes; elles appartiennent à la sensibilité ou à l'esprit parce que leur forme est le résultat que la sensibilité abstrait des choses.

Dans la doctrine des espèces, aussi appelée *de la similitude*, l'espèce prend le rôle d'intermédiaire entre l'objet et la potentialité cognitive humaine.

[4] J'utilise le terme *ontologie austère* ou *ontologie peu peuplée* tout le long de ce texte pour désigner des systèmes ontologiques dont les critères déterminant l'existence des objets et des catégories structurelles où les objets se rangent s'avèrent très stricts. En ce faisant, le nombre des catégories pertinentes à un système ontologique se voit réduit. Guillaume d'Occam a récusé le postulat de différents types d'entités pour chacune des dix catégories d'Aristote. Il a voulu restreindre le nombre des substances et des qualités à celles qui se vérifient nécessairement à sa discussion théologique.

est une *intentio* au sens où il se réfère directement à la chose objet de cet acte. Ainsi, "les intentions ne seront que des signes naturels dans l'esprit, signifiant les choses dont elles tiennent lieu." [Pacherie, 1993, page 9].

Je pense que dans l'évolution conceptuelle que je viens d'exposer, il y des problèmes récurrents auxquels je faisais allusion dans l'introduction. Un des problèmes provient de la difficulté d'expliquer la présence dans l'esprit de l'objet que vise l'acte intentionnel. Un autre est posé par le statut ontologique de l'objet intentionnel car déjà les scolastiques avaient reconnu qu'on peut entretenir des croyances vis-à-vis d'un objet inexistant, Pégase par exemple. Enfin, le processus et les moyens dont l'agent dispose pour saisir l'objet dans la pensée constituent aussi un problème. Deux solutions sont proposées, l'une d'elles affirmant l'existence d'une entité intermédiaire et similaire (les *spèces*) qui est présentée à l'agent. Une autre solution consiste à dire, au contraire, que ce qui se présente à l'esprit est l'objet même.

Jusqu'au XIXème la notion d'intentionnalité ne s'utilise pas comme notion autonome, sa signification médiévale est bornée à la relation entre signe et signifiant.

3.3 L'intentionnalité selon Brentano

Franz Brentano (1838-1917) a fait revivre la notion d'intentionnalité médiévale en l'identifiant comme la marque des phénomènes psychiques dans son livre *Psychologie vom empirischen Standpunkt* publié à Vienne en 1874. Dans un passage devenu déjà célèbre de son livre Brentano écrivit:

> Des psychologues de l'Antiquité n'en ont pas moins noté la parenté et l'analogie particulière qui existent entre tous le phénomènes psychiques, tandis qu'on ne la rencontre pas dans les phénomènes physiques. Ce qui caractérise tout phénomène psychique, c'est ce que les Scolastiques du moyen âge ont appelé la présence intentionnelle (ou encore mentale) et ce que nous pourrions appeler nous-mêmes - usant d'expressions qui n'exeluent pas toute équivoque verbale - rapport à un contenu, direction vers un objet (sans qu'il faille entendre par là une réalité) ou objectivité immanente. Tout phénomène psychique contient en soi quelque chose à titre d'objet, mais chacun le contient à sa façon. Dans la représentation, c'est quelque chose qui est représenté, dans le jugement quelque chose qui est admis ou rejeté, dans l'amour quelque chose qui est aimé, dans la haine quelque chose qui est haï, dans le désir quelque chose qui est désiré et, ainsi de suite. Cette présence intentionnelle appartient exclusivement aux phénomènes psychiques. Aucun autre phénomène physique ne présente quoi que se soit de semblable. Nous pouvons donc définir les phénomènes psychiques en disant que ce sont les phénomènes qui contiennent intentionnellement un objet (*Gegenstand*) en eux. [Brentano, 1924, page 102]

Or ce qui caractérise les phénomènes psychiques est:

(a) qu'ils possèdent des références ou des directions "vers quelque chose" (*Gerichtetsein*),

(b) qu'ils font "référence à un contenu" ; ceci en opposition aux phénomènes physiques à qui cette dualité est totalement étrangère.

La caractéristique distinctive et privative des phénomènes psychiques par rapport au phénomènes physiques est donc cette dualité appelée par Brentano existence (ou *inexistence*) ou présence intentionnelle, c'est-à-dire le rapport avec un objet. Mais Jean-Pierre Dupuy attire l'attention sur l'importance de chaque mot car ils s'avèrent "des pièges en puissance" dans la citation précédente.

> Dans l'objet réel, c'—est-à-dire situé en dehors de l'esprit, la forme est unie à la matière; l'objet 'intentionnel', lui, n'est présent que par sa forme. *Inexistenz* ('inexistence') vient du latin *in-esse*, qui signifie : 'être à l'intérieur de'. Le contresens serait ici, évidemment, de comprendre' inexistant' comme voulant dire 'non existant'. L'objet vers lequel tend l'esprit (son intention) se situe à l'intérieur de l'esprit; voilà pourquoi sa présence est immanente. [Dupuy, 1994, page 103]

Toutefois cette présence immanente d'un objet extérieur est un des points centraux de la doctrine les plus difficiles à éclaircir. On a déjà vu qu'il existe deux positions possibles. D'un côté assumer le dédoublement de l'objet intentionnel, de l'autre le considérer comme totalement extérieur. Dans le premier cas, l'objet intentionnel auquel l'acte psychique s'applique se trouve donc dédoublé puisque il est séparé de l'objet réel et il faut alors découvrir une façon de relier

l'objet intentionnel à l'objet réel. La difficulté de cette thèse est qu'on ne peut pas admettre un objet dont l'existence soit liée et bornée à l'esprit de l'agent seulement, car cela introduirait un aspect subjectif qui ne serait d'aucune utilité pour justifier l'existence ultime des objets. [5] Ceci a été la position de Brentano dans la première époque.

Lorsque l'objet est considéré comme extérieur, problèmes qui se posent: quel est le statut de cet objet dans l'esprit, comment peut-il se faire présent intentionnellement dans l'esprit? Comment expliquer par exemple la croyance que peut entretenir un agent qui soutient que les centaures ne sont pas aussi grands que les chevaux pur–sang étant donné que les centaures ne peuvent pas être des objets extérieurs puisqu'ils sont inexistants[6]?

Or, les deux positions comportent à la fois des difficultés et des avantages. Brentano soutient d'abord la première mais plus tard, vers 1909, il changea d'avis en adoptant la seconde, qui consiste à postuler que l'objet d'un acte psychique doit être extérieur ou existant, alors que l'objet intentionnel est bel et bien un objet physique.

Pendant que Brentano opérait ce changement, un de ces disciples Alexius Meinong (1853-1920) esquissa une réponse au problème des objets inexistants qui a été réfuté par la théorie des descriptions de Russell. Les deux postulats fondamentaux de la théorie des objets (*Aussersein*) sont les suivants:

(a) il y a des objets qui n'existent pas

et

(b) les objets qui bien qu'il n'existent pas sont néanmoins constitués d'une manière ou d'une autre et peuvent ainsi être les sujets de prédicats vrais.

Selon Roderick Chisholm[7] (1916–) les caractéristiques de la théorie des objets de Meinong sont les suivantes:

(a) que de tous les objets, il y en a qui existent et d'autres pas.

(b) que d'une partie des objets qui n'existent pas on peut dire qu'ils sont ou qu'ils subsistent tandis que pour les autres ne peut pas du tout dire que s'ils existent.

Mais tous les objets existants ou non peuvent avoir des caractéristiques ou des propriétés, ce qui revient à dire que malgré le fait qu'ils sont dépourvus d'existence, on peut dire qu'il *sont* d'un certaine façon. En définitive, Meinong prône que le fait d'être ou d'exister (*Sein*) pour un objet n'est pas du tout nécessaire pour attribuer à l'objet une certaine caractéristique ou un être quelconque (*Sosein*). Par exemple les montagnes d'or n'existent pas; néanmoins on peut dire qu'elles sont faites d'or de la même façon qu'on peut dire que les centaures ne sont pas plus grands que les chevaux pur–sang.

Toutefois Chisholm fait remarquer qu'en soutenant cette théorie, Meinong ne postule aucunement que les objets inexistants jouissent d'un être différent de l'existence, car bien au contraire, ces objets sont démunis de toute existence; il n'existent même pas dans le paradis platonicien puisque ce sont des objets *sans foyer* (*homeless*). [8]

[5] Si on est de l'avis que la tâche fondamentale de la métaphysique est d'interpréter "l'être" (*Sein*) à partir de "l'étant"(*Seiende*) et que "l'étant" a comme caractéristique l'objectivité, on ne peut pas accepter un objet existant seulement dans l'esprit de quelqu'un, puisque "l'étant" de cet objet sera intrinsèquement subjectif. De cette façon-là, sur ces bases on ne peut bâtir aucun concept ontologique.

[6] Dans le sens courant du terme.

[7] cfr. [Chisholm, 1982]

[8] An objet is homeless, he there said in effect, if it does not fall within the subject/matter of any of the traditional or generally accepted branches of knowledge. But I think we may find it instructive to contrue the term in a slightly different way. We may think of physical things and of person as being concreta, and of attributes, classes, and numbers as being abstracta. Such an objet would be homeless, not only because it is not covered by the usual branches of knowledge, but also because there would seem to be no place for it either in Plato's heaven or on earth. [Chisholm, 1973, page 37]

Ainsi, pour Meinong la phrase "la montagne d'or est d'or" est vraie. Cependant, la transcription faite par Bertrand Russell à partir de la théorie des descriptions donne une valeur de vérité opposée car elle sera "Il existe un x tel que x est à la fois en or et une montagne" qui est un énoncé manifestement faux.

Pour Chisholm [Chisholm, 1972, cfr. page 61] il est clair que la théorie des descriptions russellienne s'applique seulement aux objets existants et il s'étonne donc de voir qu'un énoncé faux puisse être paraphrasé par un énoncé vrai. En fait, la théoire russellienne des descriptions et de la quantification du *Principia Mathematica* ne semble ni a Meinong ni a Chisholm la plus adéquate pour rendre compte des objets qui tout en n'existant pas sont néanmoins les cibles des attitudes intentionnelles ou –dans le langage de Brentano– des actes psychiques.

Cependant Brentano ne voit pas d'un bon oeil cette proposition et dans l'*Appendice* de 1991 Brentano récuse la position que prônait Meinong:

> Je ne peut pas clore cette discussion sur la relation psychique être et exister sans dire un mot d'une opinion selon laquelle il faudrait distinguer de façon radicale être et exister, ce qui suppose qu'on prend ces termes dans un sens tout particulier. En ce cas, en effet, on pourrait peut-être affirmer qu'un objet avec lequel notre pensée entre en relation doit toujours être aussi proprement que nous sommes, mais qu'il n'existe pas toujours nécessairement à la façon dont nous existons. Il se peut que, parmi les partisans de cette opinion, personne ne soit encore allé jusque-là. Mais à propos de rouge, du bleu que nous voyons, des sons que nous entendons, et d'autres objets sensibles auxquels la science refuse l'existence, beaucoup de ces philosophes enseignent bien qu'ils sont sans exister ... Je suis incapable, je l'avoue, de trouver un sens quelconque à cette distinction entre l'être et l'existence. [Brentano, 1924, page 269]

Si on récuse la différence entre *être* et *exister* de Meinong alors la théorie ontologique de l'intentionnalité n'a toujours pas répondu à la question suivante: comment est–il possible, pour une entité qui n'existe pas, d'entretenir des relations avec un sujet qui doit nécessairement exister, comme dans le rapport psychique tel que Brentano l'a interprété? En effet, si l'activité psychique comme par exemple le cas où un sujet pense à quelque chose, apparaît *comme* un rapport entre l'esprit et "une réalité posée comme objet", et si la seule condition nécessaire est l'existence du sujet pensant, comment peut–on continuer à définir l'activité psychique comme rapport, cela veut dire comme relation entre le sujet et l'objet si ce dernier peut ne pas exister? Pour que puisse exister une relation dans le sens aristotélicien du terme, les deux éléments doivent exister. Pour remédier à cette faille logique, Brentano postule que le rapport psychique n'est pas une relation dans le sens strict mais un rapport qu'on peut définir comme quasi–relatif (*relativlich*).

Mais quoique Brentano récuse la différence entre *être* et *exister* maintenue par Meinong, on voit qu'il a subi l'influence des concepts de *Sein* et *Sosein* dans le texte de l'*Appendice* suivant:

> L'analogie avec la relation proprement dite vient de ce que dans les deux cas [la relation et le rapport quasi qualitatif], la pensée considère deux objets, l'un pour ainsi dire de façon directe et l'autre pour ainsi dire de façon latérale. Quand je considère par la pensée un amateur de fleurs, l'amateur de fleurs est l'objet de ma pensée *in recto*; les fleurs en constituent l'objet *in obliquo*. De même, quand je considère un homme plus grand que Caïus, *plus grand* est l'objet direct, Caïus l'objet latéral de ma pensée. [Brentano, 1924, page 268]

Je crois voir dans le dédoublement qui est appliqué aux objets *in recto* et *in obliquo* l'influence de la théorie de Meinong du *Sein* et du *Sosein*. Plus loin dans le texte Brentano nie que l'on puisse appliquer des actes psychiques aux objets inexistants mais soutient que le véritable objet vers lequel se dirige l'esprit est *in recto* ou *in obliquo* un objet réel. Voilà pourquoi, selon lui on peut toujours trouver une traduction dans "une proposition équivalente où le sujet et le prédicat seront remplacés par une chose réelle" [Brentano, 1924, page 286]. Il abandonne ainsi l'aspect référentiel de la doctrine et donne aux objets inexistants un rôle de composante *syncategorématique* [9] employée

[9] Selon Quine les philosophes de moyen age avaient déjà la notion des mots *syncategorématiques*. Ces mots ne sont définissables qu'en relation à un contexte parce qu'ils sont en soi vides de dénotation.

Jeremy Bentham a tenté de développer la définition contextuelle non seulement pour des particules grammaticales – comme "si" ou "mais"– mais aussi pour de vrais termes; c'est–à–dire ceux qui sont considérés comme *catégorématiques*. Lorsque Bentham trouvait des termes pratiques mais d'ontologie douteuse il observait que c'était le contexte qui permettait de découvrir le sens de la phrase, malgré l'absence de dénotation. Il déclarait ces termes *syncatégorématique* et il montrait de façon systématique qu'il pouvait toujours paraphraser l'énoncé auquel ce terme

pour décrire un agent ayant comme objet de sa pensée quelque choses d'inexistant, par exemple un centaure.

Cette composante *syncatégorématique* en tant que telle peut jouer un rôle dans des opérations logiques en les rendant plus simples

> [...] dans la pensée comme dans l'expression, tout comme le mathématicien recourt avec profit aux nombres négatifs ou aux nombres irréels. Une telle méthode permet de traiter, comme s'ils étaient simples, des représentations et des jugements très complexes, et l'on s'épargne ainsi, dans certains cas, la peine de préciser davantage un processus psychique confusément saisi. [Brentano, 1924, page 286]

Elisabeth Pacherie[10] affirme que Brentano s'éloigne ainsi d'une conception relationnelle de l'intentionnalité où la notion de renvoi est centrale, au bénéfice d'une conception attributive puisque le sujet qui pense à une chose inexistante n'est pas en relation avec elle mais est simplement modifié par elle d'une certaine manière.

Je ne puis finir cette description de la thèse ontologique de Brentano sans citer Jean-Pierre Dupuy qui nous fait part d'un mouvement révisionniste de la philosophie brentanienne et dont l'un des principaux animateurs est Stefano Franchi de l'Université de Stanford. Selon ce courant auquel appartient aussi Dupuy, un des commentateurs les plus inconstestés de Brentano aurait commis une erreur digne d'un "(mauvais) élève de philosophie de premier année" en tombant dans le piège des mots-clés de la définition d'intentionnalité de Brentano. Spécifiquement, Roderick Chisholm n'aurait pas compris correctement le terme *inexistant* parce qu'il l'aurait mépris pour le terme *non existence* en concluant que l'objet vers lequel tend la représentation est un objet physique inexistant. Comme je l'ai exposé plus haut, l'objet en soi est considéré comme physique quoiqu' inexistant. Aux yeux de ces révisionnistes cette interprétation est fausse.

> L'intentionnalité n'est donc pas l'activité psychique se dépassant elle-même à l'intérieur d'elle-même en direction d'un objet qui lui reste intérieur, cette 'transcendance dans l'immanence' que tente de cerner Husserl; cela devient un état doté d'un contenu, lequel se rapporte à un objet dont l'existence n'est pas garantie par le fait que l'état mental, lui, existe. Le contenu ne peut pas être qu'intentionnel, donc linguistique.[Dupuy, 1994, page 103]

Il serait très surprenant que la théorie linguistique de l'intentionnalité dérive de la mauvaise lecture des textes originaux de Brentano par Chisholm. Selon Dupuy ce dernier a toujours soutenu cette interprétation en se basant sur le texte original de *Psychologie vom empirischen Standpunkt* de 1874 sans tenir compte de la nouvelle version de 1911. Ceci aurait induit en erreur W. V. O. Quine pour qui la thèse de l'indétermination radicale de la traduction d'une langue dans une autre (qui sera exposée plus loin dans ce texte) ne fait qu'un avec la thèse de Brentano telle qu'elle est *(mé)comprise* par Chisholm.

3.4 La thèse psychologique de l'intentionnalité brentanienne

La motivation de Brentano en écrivant son *Psychologie vom empirischen Standpunkt* est son désir de fonder une psychologie qui soit une science à la fois rigoureuse et descriptive. Vers la fin du XIXème siècle, la méthode scientifique en vogue, responsable de l'épanouissement des sciences de la nature était l'empirisme. Brentano redéfinit la psychologie comme la science des phénomènes psychiques et il la considère comme la science de l'avenir car elle est appelée à exercer une influence considérable dans la vie pratique de tous les jours [Brentano, 1924, cfr. page 45].

A la différence de la psychologie génétique – récusée par Brentano – qui emprunte aux sciences de naturelles sa méthode constituée de l'expérimentation, l'induction et la probabilité, la psychologie descriptive proposée par lui part avec un "désavantage". En effet, dans sa psychologie descriptive l'observation fait défaut car les objets de la psychologie sont les phénomènes psychiques et la source principale des expériences pour les recherches est la perception interne de ces phénomènes qui nous sont propres. Malheureusement la perception interne ne peut se transformer en aucun cas en observation.

---

appartenait de manière que le nouvel énoncé ainsi obtenu ait du sens. Ainsi ce que Bentham appelait paraphrase est devenu ce qu'on appelle aujourd'hui la théorie contextuelle de la définition [Quine, 1981, cfr. page 68]

[10][Pacherie, 1993, cfr. page 11]

> Ainsi donc le fondement expérimental de la psychologie resterait toujours aussi insuffisant et incertain, si cette science se limitait à la seule perception interne de nos propres phénomènes psychiques et à leur observation par l'entremise de la mémoire. Mais tel n'est pas le cas. A la perception directe de nos propres phénomènes psychiques s'ajoute une conscience indirecte des phénomènes psychique d'autrui. Les phénomènes de la vie intérieure se manifestent, comme on dit, au dehors: ils entraînent ainsi des modifications que l'on peut constater extérieurement.[Brentano, 1924, page 56]

Les objets de cette nouvelle psychologie seront les actes psychiques qui diffèrent des actes physiques parce qu'ils sont intentionnels, ce qui revient à dire aussi que l'intentionalité est la marque du mental. Cette opposition entre psychique et physique et la place de l'intentionalité comme caractéristique exclusive du domaine psychique, font qu'on interprète la dernière phrase de la définition d'intentionalité (cf.3.3) donnée par Brentano comme une admission de l'irréductibilité du mental.

Dans la philosophie de l'esprit cette définition est paraphrasée par les deux énoncés suivants[11] :

(a) Tous les phénomènes mentaux manifestent de l'intentionalité.

(b) Aucun phénomène purement physique ne manifeste d'intentionalité.

Or le mental, en raison de son intentionalité n'est pas réductible au physique. Ainsi, le rêve positiviste d'unification de la science se voit gravement menacé; mais si aucun phénomène physique ne manifeste d'intentionalité, comment pourra-t-on mener à bien sa naturalisation? Comment arrivera-t-on à décrire les termes intentioonels dans un vocabulaire pertinent à un domaine dénué de cette faculté? En définitive, comment expliquer qu'un système matériel –i.e. un système de traitement de l'information ou l'homme tel qu'il est conçu dans un cadre physicaliste– puisse avoir des actes psychiques?

Ce qu'on appelle la thèse de Brentano peut être exprimée comme suit: l'intentionalité ne peut pas être réduite à une science naturelle, elle doit être une science autonome qui puisse user licitement dans son vocabulaire de termes intentionnels tels que les croyances, les désirs. Le problème est que cette science ne sera pas acceptée comme une science respectable par les positivistes.

Quine s'identifie à Brentano en ce qui concerne l'irréductibilité du mental mais il prend ses distances dans l'entreprise de construire une science autonome non-réductible puisque le résultat d'une telle tentative ne saurait être appelé une *science* au sens propre du terme [Quine, 1960, cfr. page 307].

Cependant, si on convient que la psychologie ordinaire est capable de faire des prédictions justes et si elle fonctionne si bien dans la vie de tous les jours, bref si elle a une valeur pratique difficile à nier, on pourrait peut-être trancher d'une façon qui nous permette de laisser la porte ouverte à l'unité des sciences tout en conservant l'efficacité des explications de la psychologie. En définitive, il s'agit d'éviter de jeter le bébé avec l'eau du bain en acceptant le *principe de la double norme*. La double norme a été proposée par Quine et il sera suivi en cela par Daniel Dennet, Donald Davidson et Steven Stich.

> Lorsque l'entreprise scientifique est en cause, 'si nous nous aventurons à formuler les lois fondamentales d'une discipline scientifique' la norme qui doit prévaloir est celle de l'austérité ontologique. Par contre, lorsque le but est pratique, 'dissoudre des perplexités verbales ou rendre plus aisées les déductions logiques', c'est la tolérance qui est la norme. Quine se trouve ainsi être l'un des principaux théoriciens non seulement de l'éliminativisme, qui dénie toute valeur scientifique aux concepts de la psychologie ordinaire, mais aussi de l'instrumentalisme qui leur reconnaît une valeur pratique" [Pacherie, 1993, page 18]

Ainsi cette position dite de la "double norme" est en accord avec celle des éliminativistes au sens où elle n'accepte pas la réduction du mental au physique mais elle s'en sépare en reconnaissant à la psychologie ordinaire une valeur utilitaire. On peut appeler cette position un *éliminativisme utilitariste*.

Le problème de l'irréductibilité du mental n'est pas l'unique controverse que Brentano ait engendré. Il en existe un autre et c'est celui de la relation entre "intentionalité" et "langage".

[11] cfr.[Pacherie, 1993] aussi [Cayla, 1991]

3.5 L'intentionalité et le langage

Les états intentionels sont les produits des actes psychiques représentant un contenu qui *renvoie* à un état de choses dans le monde. Mais ce ne sont pas les seules entités qui soient doués de cette faculté de *renvoyer*. Les images, les symboles comme par exemple ceux du code routier et le langage ont aussi cette capacité; ils représentent quelque chose, ils *signifient* les choses auxquels ils font référence.

La grande question est savoir où se situe la différence entre la capacité de renvoi des phénomènes psychiques et celle de renvoyer les phénomènes physiques par des objets matériels tels que les images? Si on accepte avec Brentano que la marque du mental est l'intentionalité, il faut qu'on puisse trouver au moins une différence entre ces deux types de phénomènes. En général, les objets matériels jouissent d'un intentionalité *dérivée* à la différence des actes psychiques pour lesquels cette intentionalité est *intrinsèque* au système intentionnel.

Pour montrer la différence, prenons un exemple qui l'illustre; dans un des épisodes de la série américaine "Colombo", le détective doit éclaircir le meurtre commis par un grand maître des échecs contre son rival qui détenait le titre mondial. La clé de l'enquête s'avère être une rencontre à laquelle Colombo a assisté par hasard la veille du meurtre dans un restaurant. Cette nuit-là, après avoir dîné les deux adversaires jouent une partie en utilisant comme pièces les ustensiles de la table tels que les verres, le poivrier et autres objets de ce type. Si Colombo n'était pas arrivé au moment où ils étaient en train de jouer, il aurait difficilement pu déduire que les deux maîtres avaient joué après manger parce que les choses dont ils se servaient ne représentent pas intrinsèquement les pièces du jeu; le pouvoir de représentation est dérivé et *relatif* aux personnes qui ont passé cette accord.

Même un échiquier ne représente pas intrinsèquement le jeu d'échecs car ce n'est qu'un produit relatif à notre culture ou civilisation. S'il y avait un grand cataclysme sur Terre auquel aucun parmi nous ne survive et que plus tard des êtres d'autres mondes arrivent ici et trouvent un échiquier, cela ne représenterait pas le jeu des échecs à leurs yeux. Ils devront faire des recherches anthropologiques pour reconstituer notre cadre socio-culturel et seulement par ce moyen pourront-ils déduire à quoi servaient ces objets.

Cette distinction entre intentionalité intrinsèque propre aux actes mentaux et intentionalité dérivée relative aux systèmes de représentation matérielle est en général acceptée.

Cependant il y a un sujet d'étude qui produit des controverses: le langage.

En effet, on peut considérer le langage comme un simple système matériel mais alors le problème est d'expliquer par quel processus la série de symboles que je suis en train d'écrire ou la suite de sons que j'émets renvoie les auditeurs à un état de choses? On peut supposer que le langage est le précurseur de l'intentionnalité, ou au contraire qu'il en est un dérivé.

Pour certaines auteurs dont John Searle le langage a une intentionnalité *dérivée* et en contrepartie, les états mentaux ont une intentionnalité intrinsèque.

> **Mental states have *intrinsic* intentionality, material objets in the world that are used to represent something have *derived* intentionality. The most important form of derived intentionality is in language and there is a special name in English for this form of intentionality. It is called 'meaning' in one of the sense of the word.[Searle, 1994b, page 386]**

L'enjeu est de savoir si l'intentionnalité est dérivée du langage ou bien au contraire si c'est le langage qui est dérivé de l'intentionnalité ou, dit d'une autre façon il s'agit d'éclaircir le rapport entre les propriétés de la pensée qui sont considérées comme des propriétés des entités mentales et les propriétés sémantiques du langage.

Pourquoi le langage est-il un sujet de controverse plus que les autres systèmes? Je pense qu'il y a pour cela des causes historiques car le noyau dûr de la philosophie du langage a essayé de donner un compte-rendu des propriétés sémantiques en excluant toute intention de locution, en vidant le langage de son but principal: la communication. L'idée initiale fut de bâtir une théorie du langage incluant une explication des propriétés sémantiques comme la référence et le sens sans intervention aucune de la philosophie de l'esprit.

Un exemple a été la théorie chomskienne du langage qui en prônant la division entre *compétence* et *performance* vise à faire abstraction de toute contexte intentionnel d'un énoncé ainsi que

de tout interaction possible entre le locuteur et son milieu. En effet, l'école de la *sémantique générative* cherchait à analyser la structure" profonde" des phrases en termes de leur "forme logique", assimilant celle-ci à une structure sémantique assignée directement à ces phrases. [12]

Deuxièmement le double niveau d'analyse propre du langage: la signification des signes et l'adscription des attitudes propositionnelles que l'on peut attribuer au locuteur à partir de ses énoncés.

Cependant tout le monde n'est pas de cet avis. Donald Davidson, entre autres, voit le concept de signification comme étroitement liée à ceux de croyance et d'intention, "et qu'il ne peut pas être réduit à des notions primitives ni éliminé" [Engel, 1994a, page 5]. [13] Le terme *théorie de la signification* est un terme ambigu. Cependant on peut néanmoins distinguer deux sens: le terme *théorie sémantique* qui est utilisé pour les langages naturels et le terme *théorie de la signification* pour designer l'analyse philosophique du concept de signification. Néanmoins, ceci n'implique pas pour autant que les deux théories sont indépendantes l'une de l'autre.

Les énoncés émis par un locuteur, même lorsqu'ils signalent des données objectives comme par exemple: "le chat est sur le paillasson" expriment, en plus de ce fait, l'adscription du locuteur à ces énoncés. Ils seront donc des comptes rendus de l'intentionnalité du locuteur. Ainsi, le langage aura un double niveau, d'un côté la relation entre les signes utilisés et leur signification et de l'autre l'interprétation du langage et l'interprétation des contenus mentaux et des actions. Ces deux niveaux du langage, comme renvoi aux choses et son potentiel interprétatif des attitudes d'autrui fait, à notre avis que les opinions sur les relations entre intentionnalité et langage ne sont pas unanimes.

La question qui vise à mettre au clair les relations entre les propriétés sémantiques du langage et la propriétés intentionnelles de la pensée a été attisée tout au long de ce siècle. La question de base est si ces deux types de propriétés sont indépendants ou bien s'il est possible de réduire les propriétés intentionnelles de la pensée au concept de renvoi dans la signification du langage. Une de ces controverses est celle de Chisholm–Sellars.

La controverse Chisholm–Sellars Pour Cayla, l'intérêt de cette controverse est qu'elle abrite le noyau central de toutes les discussions sur l'intentionnalité de ce siècle, avec l'atout additionnel qu'elle n'est pas coupée de la tradition historique non analytique ou *continentale* puisque les deux philosophes sont des grands connaisseurs de la philosophie scolastique médiévale et Sellars un spécialiste reconnu de Kant. [14]

Selon Sellars les pensées sont dérivées du langage, selon Chisholm bien au contraire, "les pensées sont les sources d'intentionnalité, et elles le seraient *même* s'il n'y avait pas d'entité linguistique. Sellars admet le premier point mais conteste la validité de la conditionnelle contra-factuelle: sans entités linguistiques, il n'y aurait *pas* d'intentionnalité des pensées. Selon Sellars, donc, ce sont les catégories sémantiques du discours qui doivent expliquer les catégories intentionnelles de la pensée, tandis que Chisholm affirme exactement la proposition inverse".[Cayla, 1991, page 50]

Le but de cet exercice est de mettre en concurrence les deux modes de renvoi, le renvoi sémantique et le renvoi des états intentionnels. Le concept de renvoi sémantique dans un langage peut être représenté par un énoncé du métalangage de la forme "'...'signifie p'.

Sellars prône la théorie *analogique* de la peosée. Il propose l'existence d'un langage mental ou *Mentalais* et d'un isomorphisme entre le langage et la pensée. Cbaque épisode mental a ainsi un épisode verbal qui lui correspond. Ainsi, les propriétés intentionnelles de la pensée, son mode de

[12]"Chomsky oppose, comme on sait, la théorie de la compétence linguistique, c'est-à-dire de la connaissance que le locuteur a de sa langue, à l'étude de la performance, concernée par l'emploi effectif de la langue dans des situations d'énonciation concrètes. La première est l'objet de la linguistique; la seconde relève de disciplines multiples, comme la psycho-physiologie, la sociologie des faits de la parole. L'hypothèse fondamentale qui doit rendre compte de l'*usage créatif de la langue dont un locuteur natif es capable est le mentalisme: la théorie chomskienne postule* en effet que l'on peut développer un modèle de la compétence du locuteur, c'est-à-dire reconstruire le système de règles sous-jacentes qu'il a intériorisées, sans avoir à prendre en compte l'interaction du locuteur avec son milieu, ni d'ailleurs le fait que la langue puisse lui servir à communiquer."[Proust, 1982, page 23] Aussi voir [Engel, 1994a, principalement chapitre I]

[13]*Aussi voir* [Engel, 1994a, *principalement chapitre I*]

[14]Pour excellent compte rendu aussi pour parcourir la totalité de la correspondance voir [Cayla, 1991]

renvoi ne seront qu'une analogie *interne* des propriétés sémantiques des langages natureles et elles acceptent aussi une réécriture du type "'..."signifie p' saos faire aucun appel au concept de pensée.

Pour Chisholm, cette transcription est vide de substance car le renvoi intentionnel doit être autre chose qu'une suite de signes non interprétés; elle doit "contenir *analytiquement* le concept de pensée; en conséquence, ils doivent s'analyser en "'..."exprime t et t porte sur p' où t est une pensée." Les énoncés sémantiques sont, dans cette perspective, des *abréviations* d'énoncés intentionnels, de sorte que, loin d'être aptes à rendre compte de l'intentionnalité des pensées, ils dérivent de celle-ci."[Cayla, 1991, page 51]

Chisholm soutient contre Sellars une thèse *forte* sur l'irréductibilité intentionnelle aux propriétés du langage: les énoncés métalinguistiques qui les expriment sont dérivés des propriétés de la pensée et ces dernières sont le support, de l'intentionnalité. Dans cette démarche Chisholm a proposé des critères sémantico–logiques de l'intentionnalité et a été un des premiers à montrer la différence entre intensionnalité–avec–un–s et intentionnalité–avec–un–t.

Les deux termes seront souvent amalgamés malgré le fait que Chisholm a montré que les contextes intensionnels débordent largement le cadre des phénomènes intentionnelles. John Searle a été celui qui a le plus souvent signalé cette confusion dans les vingt dernières années. Elle aurait son origine dans le fait que toutes les phrases qui sont intensionnelles–avec–t sont à la fois intentionnelles–avec–s, mais les premières ne sont pas les seules. Les énonces modaux sont aussi des cas d'intensionnalité–avec–s mais il ne sont pas forcement intentionnels–avec–un–t.

> **Croire que quelque chose d'intentionnel-avec-un-t appartienne en propre à l'Intentionnalité–avec–s procède d'une erreur dont la philosophie linguistique semble chroniquement affectée: La confusion entre les caractéristiques du rapport et celles des choses rapportés. C'est une caractéristique des rapports d'états Intentionnels-avec-un-t qu'ils soient des rapports intensionnels-avec un-s.Mais il ne s'ensuit pas et n'est généralement pas vrai que les états Intentionnels-avec-un-t soient comme tels intensionnels-avec-un-s. [Searle, 1983, page 41]**

John Searle attire l'attention que le compte rendu des attitudes propositionnelles sont en fait des représentations de représentations. Quand j'ai dit, par exemple: "Je crois que les centaures ne sont pas plus grands que les chevaux purs sang" la vérité de la représentation ne dépend pas du contenu; cela veut dire que "les centaures ne sont pas plus grands que les chevaux purs sang" dans la réalité sinon du fait que dans mon esprit j'ai cette représentation et que je la tienne comme vraie, parce que, ce que je suis en train de rapporter, est le contenu d'une représentation mentale, et c'est par ce fait que les intentions–avec–t sont des contextes opaques ou des intensions–avec–s.

Pour finir la section et à titre de corollaire, j'aimerais signaler que l'irréductibilité de l'intentionnalité au langage telle que Chisholm la conçoit n'est que la version linguistique de la thèse de Brentano.

3.6 Conclusion

Nous avons vu que la mise au jour du concept d'intentionnalité réalisé par Brentano pose des questions qui ont nourrit la philosophie dans ce siècle.

D'un côté nous avons la thèse d'irréductibilité du mental qui est une menace pour le programme positiviste d'unification des sciences. En effet, une fois acceptée elle établit une tension entre deux faits contradictoires: le principe de l'existence d'une science avec un vocabulaire physique et des lois strictes et l'existence "du mental" qui tout en n'entrant pas dans ce cadre puet être considéré comme tel.

Nous avons vu qu'une solution est la solution utilitariste de Quine.

En suite, nous avons montré les controverses entre la relation entre langage et intentionnalité. Dans le chapitre suivant nous allons continuer dans la même ligne. Je ne vais pas me référer spécifiquement au langage mais à un problème voisin, le problème de la présence intentionnelle de l'objet dans l'esprit, plus précisément les concepts de "contenu" et de "représentation".

Chapitre 4

Representation et contenu

Yo soy yo y mi circunstancia.
José Ortega y Gasset

4.1 Introduction

Au cœur de l'intentionnalité brentanienne et des sciences cognitives on trouve les concepts de *contenu* et de *représentation*. Dans le cadre des sciences cognitives, ces concepts sont nécessaires pour expliquer non seulement la capacité de renvoi que démontrent les phénomènes intentionnels mais aussi le rôle causal que l'on octroie aux choses auxquelles ils renvoient. Or, les différentes façons de concevoir et de définir ces concepts s'avèrent centrales à la compréhension des différentes entreprises de naturalisation de l'intentionalité; c'est–à–dire toutes les démarches visant à mettre en évidence les propriétés *non-sémantiques* qui confèrent à un système physique des propriétés sémantiques, voire intentionnelles.

Le plan de ce chapitre est tout d'abord, de montrer le résultat d'un exercice d'exegèse des termes *représentation* et *contenu* en partant de Brentano.

Ensuite, je vais exposer le problème que pose l'attribution des propriétés sémantiques à l'esprit dans le cadre des *sciences cognitives*. Je vais aussi exposer les difficultés que l'on éprouve à individualiser des contenus selon chacune des stratégies de la dichotomie traditionnelle *externalisme-internalisme*.

Finalement je vais présenter la position phénoménologique de Husserl comme une position intermédiaire au sein de la première de ces dichotomies.

Je discuterai deux interprétations différentes de l'œuvre de Husserl: l'interprétation analytique qui soutient un parallélisme entre les concepts phénoménologiques et les concepts fregéens en contraposition à l'interprétation *gestaltica* qui, entre autres caractéristiques, assimile le concept de noème à celui de mode de présentation.

4.2 Représentation et Contenu chez Brentano

4.2.1 La représentation

Dans le chapitre I (§. 2) du Livre II de la *Psychologie vom Empirischem Standpunkt* Brentano utilise le terme *représentation* comme l'exemple par excellence du phénomène psychique mais il signale:

> [...] et par représentation j'entends ici non pas ce qui est représenté, mais l'acte de représenter. [Brentano, 1924, page 93]

Et après il ajoute:

> [...] qu'on considère quoi que ce soit avec haine, amour ou indifférence; qu'on l'accepte ou qu'on rejette ou qu'on réserve entièrement son jugement, on ne peut mieux s'exprimer qu'en disant qu'on se représente cet objet. Dans le sens que nous donnons au mot *représenter*, *être représenté* est synonyme d'*apparaître*. [Brentano, 1924, page 95]

De la lecture du point §.3 du même chapitre on déduit qu'il fait un amalgame entre les phénomènes psychiques et les représentations, puisqu'il considère celles–ci comme le fondement de tout acte psychique. Tout acte psychique sera une représentation.[1]

À la différence du Livre *I* qui est dédié à montrer la différence entre les actes physiques et les actes psychiques, un des objectifs du Livre II est de donner des bases homogènes à la nouvelle psychologie scientifique brentannienne. Les représentations sont citées dans les premiers chapitres du Livre *II* comme étant la dénomination, caractéristique commune à tous les actes psychiques.

> Nous pouvons donc considérer comme indubitablement correcte cette définition des phénomènes psychiques: ce sont ou bien des représentations, ou bien ... elles ont pour fondement des représentations. [Brentano, 1924, page 99]

Ensuite il présente d'autres auteurs dont la tendance est au psychologisme, comme par exemple Wilhelm Wundt (1832–1920) auteur entre autres de *Physiologische Psychologie* et que Brentano nomme explicitement dans une note en bas de page[2]. Selon Brentano cette conception des représentations est plus bornée que la sienne propre. Brentano dit se recommander d'une conception plus large du concept de représentation, plus proche de celle de Johann Friedrich Berbart (1776–1841)[3] et Rudolf Herman Lotze (1817–1881)[4].Ces derniers auteurs sont des opposants au psychologisme

[1] J'établirai plus tard dans cette section que le concept de *représentation* chez Brentano a un double volet, à la fois actif et passif. Il est actif car il se réfère à un acte de l'esprit et il est passif parce que c'est la représentation qui permet la présence d'un objet dans l'esprit.

[2] De prime abord, il semble contradictoire que Wunt, considéré comme le père de la psychologie expérimentale se trouve inscrit dans la lignée du psychologisme. Ceci peut être dû au fait que l'influence de Wundt sur son époque ait été davantage attachée à l'importance de son laboratoire qu'à son œuvre.

Néanmoins pour Wundt l'expérimentation n'a de sens que si elle s'appuie sur l'introspection. C'est cette dernière qui lui fournit les objets pertinents et c'est seulement à partir de l'introspection que les composantes du processus que l'on veut étudier peuvent être isolées. L'expérimentateur a pour but de mesurer la durée des différentes phases de ces processus et d'étudier les conditions externes qui influent sur leur fonctionnement. La méthode qu' il utilise est la mesure du temps de réaction. Le temps de réaction est considéré, à son tour, comme une mesure des la complexité du processus. [de Souabe Zyriane, 1985c]

[3] Bien que l'on désigne le plus souvent Wundt comme étant à l'origine de la psychologie expérimentale vers 1879, ce dernier a eu deux prédécesseurs importants dont l'un est Herbart. En effet, la publication de l'œuvre principale de Herbart intitulée *Wissenschaft neugegründet auf Erfahrung, Metaphysik und Mathematik* (La Psychologie comme science fondée sur l'expérience, la métaphysique et les mathématiques) date de 1824–1825. Dans ce livre Herbart expose une problématique dans le but de fonder une nouvelle discipline du type rationalisme métaphysique. Il s'agit d'opérer une fusion entre les enseignements mécanistes des sciences naturelles, l'associationnisme anglais et l'idéalisme kantienne.

De la physique newtonienne Herbart conserve dans son système l'idée de *force*, qu'il transpose au domaine des idées et des représentations, en dehors de toute critique épistemique. Les phénomènes d'association constituent sous ce rapport un matériel extrêmement démonstratif, vu qu'ils se prêtent à une analyse en termes d'attraction et de répulsion réciproques. L'influence de Kant apparait dans le fait que pour Herbart comme pour Kant les faits observables renvoient à l'essences des choses. Mais à la différence de Kant, Herbart donne à ce renvoi un certain rôle causal. Ainsi, l'analyse empirique des phénomènes introduit automatiquement aux choses en soi. Le dualisme du sujet peut être contourné puisqu'il peut être abordée empiriquement par le biais de l'action manifeste. Cette démarche ne pouvait pas aboutir à la découverte du psychisme, étant donné le point de départ et le contexte de l'époque. Elle postule, en revanche une dualité de la connaissance psychologique elle–même et prescrit de passer du métaphysique à l'empirique, et réciproquement selon les cas. Herbart se disait partisan d'une psychologie qu'il voulait quantitative mais non–expérimentale avec une forte tendance formaliste. [de Souabe Zyriane, 1985a]

[4] Comme son prédécesseur Herbart, Lotze veut donner à la pédagogie des bases scientifiques. Tout en affirmant, de même que Herbart, la nature foncièrement métaphysique de ses recherches, Lotze aborde divers domaines de la psychologie étroitement apparentés à la physiologie, science qui l'intéressait vivement. Il est considéré comme un précurseur de la psychophysique et il développe entre autres la théorie de la perception qui aura une grande influence sur les théories postérieures de Wundt et de Helmholtz. Cette théorie est fondée sur la notion des *signes locaux*. Ces derniers sont définis comme des variations d'intensité consécutives à l'application d'un stimulus et déterminant une structure intensive différente selon le lieu du sensorium atteint. Ainsi, dans le domaines tactile, la sensation sera différente pour toutes les régions de la peau en raison des variations de résistance mécanique des tissus; dans le domaine visuel, ce sont le mouvements des yeux, différents pour chaque point du champ, qui fournissent les signes locaux sous le rapport de l'attention. L'espace perçu est donc référé à la topographie du corps

et il maintiennent la différence kantienne entre psychologie et critique de la connaissance. Selon cette différence, les contenus de pensée sont irréductibles aux conditions de leur genèse.

En particulier il vise ceux pour qui les actes que nous appelons *qualia* ne contiennent pas de représentations. Mais il nous explique que dans ce cas, les sensations par exemple de brûlure ou de douleur reposent aussi sur une représentation.

> [...] Si vous vous coupez, vous n'éprouvez d'ordinaire aucune perception de contact; si vous vous brûlez, aucune perception de chaleur; dans les deux cas, il ne semble y avoir que de la douleur.
>
> Il est hors de doute, néanmoins, que dans ces cas également le sentiment repose sur une représentation. Nous avons toujours la représentation d'une localisation précise que nous lions d'ordinaire à telle ou telle partie visible et palpable de notre corps. Nous disons que nous avons mal au pied, à la main, que nous souffrons de telle ou telle partie du corps. Ceux qui estiment qu'une représentation de ce genre découle naturellement et primitivement d'une excitation des nerfs seront les derniers à pouvoir nier l'existence d'une représentation à la base de ces sentiments. [Brentano, 1924, page 96]

Le fait de désigner la représentation comme le lien ou la caractéristique commune de tous les actes psychiques contraste avec la classification ultérieure qu'il donne des modes fondamentaux de renvoi. Selon cette première classification, la *représentation* n'apparaît que comme un mode parmi d'autres.

> [...] nous croyons nous aussi qu'on doit distinguer d'après leur mode de relation avec l'objet, trois classes principales d'activités psychiques. Mais ce trois classes ne sont pas celles que l'on établit communément. En l'absence d'expressions plus appropriées, nous donnons à la première le non de *représentation* (*Vorstellung*), à la seconde le nom de *jugement* (*Urteil*), et à la troisième le nom de *mouvement affectif* (*Gemütsbewegung*), d'*Interesse* ou d'*amour*(*Liebe*). [Brentano, 1924, page 203]

Les trois modes sont considérés comme différents et non-réductibles entre eux par Brentano dans la première époque. Cependant, Brentano change d'avis plus tard et dans l'*Appendice* du *Psychologie vom Empirischem Standpunkt* il abandonne la différence entre *jugement* et *représentation*. [Brentano, 1924, cfr. pages 278-279 et 282]

A mon avis, il ne faut pas interpréter le terme représentation comme étant ambigu, moins encore contradictoire. Je pense que les difficultés que nous trouvons à l'instar de Brentano pour saisir le concept sont dues d'une part à la complexité du concept de représentation et de l'autre à la tentation toujours présente de l'assimiler simplement au concept de renvoi. Brentano même le suggère

> Nous parlons de représentation chaque fois qu'un objet nous apparaît. Quand nous voyons quelque chose, nous nous représentons une couleur; quand nous entendons quelque chose, nous nous représentons un son; quand nous imaginons quelque chose, nous nous représentons cette image. Employant le mot avec cette signification générale, nous avons pu dire que l'activité psychique ne pouvait jamais se rapporter à quelque chose que ne fût pas objet de représentation. [Brentano, 1924, page 203]

Or, la complexité du concept de *représentation* est en partie due au fait qu'il comporte deux rôles différents. D'un côté un rôle actif car il est considéré comme un *acte* psychique et dans ce sens il est une des trois formes de relation psychique possible avec un objet intentionnel. De l'autre, un rôle passif car il révèle la présence de l'objet dans l'esprit et je reprends les mots de Brentano cités plus haut, *être représenté* ou *d'apparaître*.

La représentation selon Brentano est une activité psychique intrinsèquement consciente. Comme le signale Jean-Pierre Dupuy:

> L'activité psychique est intrinsèquement consciente d'elle-même. Quand nous pensons, nous avons une perception immédiate du fait que nous pensons, et la perception de l'activité pensante est simultanément perception de l'objet de la pensée. Cette perception interne ne peut pas être une observation, note Brentano, car il y aurait alors régression infinie d'activités psychiques pointant les unes vers les autres. C'est en une appréhension globale et unique que la pensée comme activité se rapporte à la fois à elle même et à son objet intentionnel... [Dupuy, 1994, page 103]

---

et du champ; il résulte finalement, dans son organisation subjective, de la tendance constitutive de la conscience à disposer spatialement les contenus sensoriels des signes locaux. Lotze explique l'espace à partir de la non-spatialité en dotant la conscience d'une potentialité organisatrice étrangère en soi aux mécanismes perceptifs proprement dits. Cette doctrine aura une influence non négligeable sur les théories de la *Gestaltheorie*.[de Souabe Zyriane, 1985b]

4.2.2 Le contenu

Brentano introduit tardivement le concept de *contenu* dans l'*Appendice* de la *Psychologie vom Empirischem Standpunkt* pour fournir une solution à l'éternel problème des actes psychiques qui portent sur des objets intentionnels inexistants. Le fait que l'on accepte des relations psychiques avec des choses (*Dinge*) qui n'existent pas ne veut pas dire que ces choses puissent être considérées comme des objets.

La démarche de Brentano consiste en établir une différence entre l'objet et l'aspect de la chose sur lequel l'acte porte. L'agent peut entretenir des relations de diverses façons avec une même *chose*. Ces différentes manières de relation deviendront

> [...]en quelque sorte plus que l'objet lui-même [et] qui contiendrait en soi cet objet et que se trouverait en même temps dans l'agent psychique. C'est ce qu'on a appelé le « contenu » du rapport psychique. [Brentano, 1924, page 284]

Or, *le contenu n'est jamais représenté comme l'objet d'une représentation* [Brentano, 1924, cfr. page 285]. Un exemple servira à éclaircir cette notion de contenu et d'objet. Brentano prend comme exemple le jugement "Il n'y a pas de centaure". Dans ce cas l'*objet* du jugement sera le centaure mais le *contenu* sera l'inexistence du centaure, donc le véritable sens de ce jugement serait "un agent psychique est en train actuellement de nier un centaure" [5]

La différence pour Brentano entre *objets* et *contenus* est que les premiers peuvent être sujets ou objets des phrases, ces phrases étant le contenu des actes intentionnels qui expriment des propriétés ou des prédicats. Cela veut dire que le contenu peut être assimilé à la notion soit de concept, soit de mode de présentation. Ceci est en accord avec la tradition de Russell et de Frege où le contenu est assimilé au *mode de présentation* ou au sens (*Sinn*) fregéen tandis que l'objet même constitue le référent.

L'on est amené à se demander si cette solution syntaxique de Brentano aux problèmes des objets inexistants aurait amené à Chisholm à proposer l'interprétation linguistique de l'intentionalité que j'ai déjà exposée dans le chapitre 3 (§3.4) et à faire la mauvaise traduction du concept d'intentionalité que certains courants révisionnistes lui attribuent? Je ne suis pas en mesure de répondre à cette question, mais en lisant Brentano on ne peut mettre l'accent que sur un de ses aspects. Il ne s'agit pas d'une lecture que l'on fait sans *a priori*. C'est un type d'exégèse porté sur la coïncidence des concepts que l'on utilise maintenant. Cette lecture orientée, je l'avoue est celle que j'ai faite moi-même. Bien que cette lecture soit intéressante du point de vue historique, elle n'est nullement naïve et elle ignore des aspects de l'œuvre de Brentano qui lui furent essentiels, du moins dans le deuxième livre de *Psychologie vom Empirischem Standpunkt*. Ainsi, la gerbe riche en concepts et les caractéristiques de la notion même d'*intentionalité* est effeuillée comme une marguerite dont on ne prend que les pétales propres à justifier une démarche représentationnelle, tout en laissant tomber les autres. Surtout, on laisse tomber les aspects qui pourraient nuire au projet de naturalisation de l'intentionalité.

L'intentionalité est donc l'*activité* psychique qui se dépasse à l'intérieur d'elle-même en direction d'un objet lui demeurant extérieur; mais elle devient, dans l'interprétation cognitiviste, un état mental doté d'un contenu et celui-ci se rapporte à un objet dont l'existence n'est pas garantie hélas par le fait que l'état mental lui-même existe. Ceci amène à une conception intensionnelle du contenu, c'est à dire simplement linguistique et dans laquelle le concept de *contenu* est assimilé au concept de *sens*. En définitive, il s'agit de rapporter le problème du domaine de la philosophie de l'esprit à la philosophie du langage.

4.3 Représentation et contenu en sciences cognitives

Aujourd'hui, les différents courants des sciences cognitives s'accordent à classer les états mentaux en deux grands groupes. D'un côté les *qualia* dont j'ai déjà parlé en dans le chapitre 2 (§2.2, i.e. une douleur, la perception d'une couleur) et qui ne sont pas considérés comme représentationnels.

[5] Puis dans le texte Brentano poursuit avec les différences d'être en *recto* et en *en obliquo* que nous avons exposées plus haut.

De l'autre, les états mentaux représentationnels qui sont groupés sous l'appellation que leur donne Bertrand Russell d'*attitudes propositionnelles*.

Russell y a fait référence en maintes occasions, par exemple dans l'introduction au *Tractatus logico-philosophicus* de Wittgenstein :

> ... le problème [...] de la forme logique de la croyance, c'est-à-dire la question de savoir quel est le schéma qui représente ce qui se produit quand un homme croit. Évidemment, le problème s'applique non seulement à la croyance mais aussi à la foule d'autres phénomènes mentaux que l'on peut nommer des attitudes propositionnelles : doute, réflexion, désir, etc. Dans tout ces cas, il paraît normal d'exprimer le phénomène sous la forme "A doute de p", "A désire p", etc. , *ce qui montre que c'est comme si nous traitions d'une relation entre une personne et une proposition.* [Russell, 1953, page 17]

La postulation des attitudes propositionnelles par Russell est motivée en partie par sa démarche épistémologique et par sa quête de solutions aux problèmes liés à la connaissances empirique.

> Both for logic and for theory of knowledge, the analysis of such occurrences [*attitudes propositionelles*] is important, especially in the case of belief. We find that believing a given proposition does not necessarily involve words, but requires only that the believer should be in one of a number of possible states defined, mainly if not wholly, by causal properties. When words occur, they 'express' the belief, and if true 'indicate' a fact other than the belief. [Russell, 1940, page 18]

La conception standard des attitudes propositionnelles fait d'elles des états mentaux possédant un double aspect; d'un coté ils ont un contenu propositionnel et de l'autre des attitudes envers ce contenu. On trouve dans cette caractérisation le double aspect brentanien des représentations : la proposition en tant que présence dans l'esprit de l'objet sous un certain aspect et l'attitude qui est l'acte même de représentation

Le fait que le contenu soit conçu du point de vue ontologique comme une proposition, comme les propriétés d'un objet ou un état de choses ne sera pas pris en compte dans la présente discussion.

Du point de vue formel les attitudes propositionnelles peuvent être analysées de façon relationnelle ou de façon monadique. La première est celle qui considère qu'une attitude propositionnelle comme dans l'exemple "Jean croit que l'étoile de matin est Vénus" exprime une relation (croire) entre Jean et l'objet de sa croyance tandis qu'une analyse monadique conclut que le prédicat "croit que l'étoile de matin est Vénus" est attribué à Jean. Je vais montrer que cette différence entre les conceptions des attitudes propositionnelles devient centrale à la naturalisation des théories fonctionnalistes.

4.3.1 La représentation intentionnelle

Mais par quel moyen l'esprit peut-il accomplir le rôle de renvoi caractéristique de l'intentionnalité? Comment rendre présent dans l'esprit un état de choses? Comment est-t-il possible de le représenter? Qu'est-ce qu'une représentation?

Le concept de représentation a un double volet; d'un côté il se réfère à la reproduction d'un objet, par exemple une photo d'identité qui reproduit en quelque sorte l'individu photographié. D'autre part, il a aussi une acception liée au pouvoir d'agir au nom de ce qu'il représente. Tandis que la première acception a un rôle passif de "rappel", la deuxième par contre est active parce qu'elle implique la possibilité de modifier une situation ou un état. Ainsi, par exemple, dans un mariage le maire représente l'état ou la loi en vertu des pouvoirs qui lui sont conférés de changer les états civils des membres du couple.

Pour qu'une représentation soit effective il faut un observateur qui ait la connaissance explicite ou implicite du code d'interprétation, c'est-à-dire du sens de la représentation. Le sens de la représentation reçoit le nom de *contenu*. Dans des environnements informatiques lorsqu'on parle de représentation on pense tout suite à un système formel, ce qui signifie que nous assimilons l'idée de représentation à celle d'un langage. Cependant, il y a d'autres moyens de représentation tels que les images. Il y a également toute une norme du langage gestuel puisque chaque société possède un ensemble de gestes spécifiques chargés de signification. [6]

[6] A ce sujet John Searle vient d'achever une théorie visant à expliquer l'intentionnalité dans une société. Il a exposé une grande partie de cette théorie à l'Universite de Génève pendant le semestre d'hiver 1994-95 dans le cadre du Séminaire Philosophique du Prof. Kevin Mulligan.

Le concept de représentation a été systématisé par Charles Sanders Pierce qui vers 1935 développe un théorie des signes en les classifiant en

a.– *Icônes* : signes qui ont une ressemblance structurelle avec la chose représentée. (i.e. photo, portrait)

b.– *Indices* : où le signal est connecté avec l'objet par une relation causale (i.e. la fumée indique le feu)

c.– *Symboles* : les signaux sont en relation avec les objets en vertu de l'usage ou de l'association.

Le différentes théories du mental ont différentes façons de considérer la représentation. L'école représentationnelle née au MIT et dont un des plus importants partisans est Jerry Fodor soutient que l'unique type de représentation possible pour le mental est symbolique. Fodor propose l'existence du *langage de la pensée* ou *Mentalais* comme celui proposé par Sellars mais dans le cadre de la théorie fonctionnelle qui a comme métaphore de base l'ordinateur. Je revendrai sur ce point lors de chapitres à venir.

D'autres comme par exemple Stephen Michael Kosslyn acceptent l'existence d'images mentales dans l'esprit. L'existence ou non d'image mentales dans le cadre intentionnel est un sujet de controverse qui anime depuis plus de vingt ans les sciences cognitives. Les principaux et les plus actifs détracteurs des images mentales sont Jerry Fodor et Zenon Pylyshyn, et un de leurs arguments contre les images est qu'elles ne peuvent pas faire partie d'un langage génératif, c'est à dire, qu'on ne peut pas engendrer l'infinité des images qu'on pourrait se représenter en ayant seulement un alphabet fini.

Finalement, Fred Dretske soutient que le lien entre la représentation et l'objet représenté est de type indicatif ou en d'autres termes qu'il existe une relation causale entre le signe et l'objet. Je vais présenter cette théorie en détail dans des sections ultérieures.

4.3.2 Le contenu mental

Dans le cadre des sciences cognitives, les termes *contenu* et *sens* sont souvent amalgamés et utilisés comme synonymes. Néanmoins, il est juste de faire la différence suivante: le terme *sens* s'applique pour exprimer des propriétés sémantiques appartenant aux phrases ou aux expressions de langages naturels tandis que *contenu* est applicable aux représentations mentales ou en général à des attitudes propositionnelles.

Mise à part cette différence, les notions sémantiques de *signification* et de *référence* présentent le même type de difficultés que celles de *contenu* et de *sens*.

Si la définition d'*intentionnalité* donnée par Brentano correspond au lien sémantique par excellence et si les représentations sont le moyen d'expression des états d'affaires auxquelles les attitudes propositionnelles font référence, le contenu est ce à quoi elles renvoient, ce sur quoi elles portent, alors un des buts poursuivis par la démarche de naturalisation des sciences cognitives est d'expliquer comment on peut appliquer aux représentations la propriété d'"avoir du sens"[Proust, 1990, cfr. page 13].

Une théorie psychologique doit fournir un moyen d'individualisation des contenus, donc des critères qui permettent de décider si deux contenus sont différents ou bien au contraire si ce sont des instances d'un même et unique type. Ceci est un problème fondamental aux tentatives de naturalisation de l'intentionnalité. Les difficultés seront exprimées par la suite.

Premièrement l'individualisation des contenus doit être solidaire avec le type de relation (soit de dépendance soit de corrélation) entre les contenus et les états du cerveaux. L'individualisation des contenus devient une condition préalable à la cobérence de toute solution que l'on peut proposer au problème corps–esprit. Autrement dit, les propriétés sémantiques des états internes devront se montrer survenantes ou du moins en corrélation entre elles sur les propriétés intrinsèques physiques et ceci de façon synchrone. Pour que tel soit le cas, il faut que les propriétés sémantiques des contenus des états internes ne soient déterminées que de façon intrinsèque ou non–relationnelle. Elles devront être déterminées sans faire aucun appel à des faits dépendants de l'environnement

et sans relation avec l'historique du sujet. Or, ceci n'est pas le cas. L'exercice de la pensée de la Terre jumelle de Putnam (parmi d'autres expériences) montre que deux organismes qui sont totalement identiques du point de vue interne peuvent, néanmoins s'avérer différents par rapport aux propriétés sémantiques qu'ils possèdent. Le contenu de leurs croyances et de leurs désirs par exemple n'est pas identique du point de vue sémantique. Cette constatations a motivé le dualisme des contenus que j' exposerai dans la section suivante.

Deuxièmement, les contenus des états mentaux jouent un rôle fondamental dans les explications que l'on donne aux comportements dans la vie de tous les jours. Souvenons-nous que la psychologie ordinaire explique les comportements en fonction des attitudes propositionnelles qu'elle attribue aux agents. Si "j'ai soif et je crois qu'il y de l'eau fraîche dans le frigo" ces deux attitudes (mon désir d'appaiser ma soif et ma croyance qu'il y a de l'eau dans le frigo) vont entretenir une relation causale avec mon action de me mettre debout, de marcher jusqu'au frigo, d'ouvrir la porte et de boire de l'eau. Aussi c'est le contenu de la croyance qu'il y a de l'eau fraîche dans le frigo qui produit mon action et non pas, par exemple ma conviction que le franc français va chuter dans les marches internationaux dès que Chirac s'installera à l'Elysée. Or, les attitudes propositionnelles non seulement semblent participer aux processus causaux de l'action mais aussi elles le font en fonction de leur conteou. L'individualisation des contenus selon leur rôle causal devient centrale.

Résumons: l'explication du rôle des contenus mais aussi leur individualisation se heurtent à deux difficultés dans la démonstration de deux parallélismes ou corrélations. Premièrement le parallélisme ou la corrélation, entre les propriétés sémantiques et les propriétés physiques des états qu'elles dénotent. Deuxièmement la corrélation sémantico-causale: la difficulté de démontrer que les raisons (contenus sémantiques) peuvent être aussi des causes sans violer le principe physicaliste.

Avant de décrire les solutions aux problèmes que je viens d'exposer quant à l'individualisation des contenus dans le cadre de la philosophie de l'esprit, j'aimerais faire état des problèmes que cette même individualisation présente pour la philosophie du langage. Dans ce dernier cas, l'individualisation des contenus se heurte aux mêmes difficultés que celles que nous avons éprouvées pour la détermination des concepts.

Premièrement, les attitudes propositionnelles sont des concepts opaques; elles n'admettent pas d'analyse extensionnelle. Cette difficulté se retrouve aussi pour d'autres entités telles que les énoncés modaux dans la linguistique. Nous rappelons que dans un contexte opaque il n'est pas possible de remplacer *salva veritate* le contenu par un autre co-référentiel et ce problème est connu sous le nom de "problème de Frege".

Deuxièmement, il existe le problème *des modes de présentation*. Nous allons illustrer ce point par un exemple. Soient les expressions suivantes: "être un triangle équiangle" et "être un triangle équilatéral", synonymes non seulement dans le monde réel mais dans tous les mondes possibles, elles ont donc les même extensions et la même intension[7] footnote et il est possible de les accorder aussi comme la même propriété (en faisant abstraction de toute coosidération causale) bien qu'elles soient différentes du point de vue de la présentation dans l'esprit. C'est une chose de penser à un triangle équiangle et une autre de penser à un triangle équilatéral. [Rey, 1994, cfr.] D'où l'intérêt du théorème qui montre que les deux ensembles sont, en fait, co-référentiels.

On trouve un équivalent à ce problème dans la linguistique, pour les énoncés indexicaux. Dans ce cas les indexicaux sont résolus en utilisant des règles d' inférences. Ainsi, dans l'énoncé "Je suis français" l'indexical "Je" est interprété comme "le locuteur est français".

La solution du problème des modes de présentation est équivalente à la solution du problème

[7]Le premier essai de formalisation de la notion de sens *(intension)* est dû à Carnap, qui utilise le concept de mondes possibles. Etant donné que la signification d'une expression est déterminée par son extension, Carnap a suggéré que le sens d'une expression (son intension) est une fonction qui, en partant de l'ensemble de tous les mondes possibles, donne pour chacun d'eux l'extension de l'expression. Par exemple, l'extension de l'étoile du matin est Vénus, l'extension du nombre des planètes est neuf puisqu'il y a neuf planètes dans le monde réel. Cependant, dans le monde du Petit Prince de Saint Exupéry, l'extension de la dernière expression est très différente. C'est-à-dire qu'on est obligé de donner comme valeur sémantique à la phrase "le nombre de planètes" pour chacun des mondes possibles. Bref, l'intension d'une expression sera la somme de toutes les extensions possibles que cette expression peut avoir, mais assemblées de façon organisée comme une fonction ayant pour ensemble de départ tous les mondes possibles et pour valeurs les extensions correspondantes. Or dans le cas qui nous occupe en référence au triangle, les deux intensions coïncident parce qu'il s'agit d'une identité analytique.

du lien sémantique car une des difficultés majeures dans les deux cas tient au rôle du contexte. Si Jean croit que l'"étoile du matin" est Vénus tout en ne croyant pas que l'"étoile du soir" est Vénus, si l'énoncé "Je suis française" est faux lorsque c'est moi qui l'énonce (puisque je ne suis pas française) et vrai quand c'est Mme Arlette Laguiller[8] footnote qui le dit, on voit bien que l'on ne peut pas déterminer la référence à partir du seul énoncé. Il faut tenir compte des autres données. Pour ce qui est de Jean, il faudrait tenir compte de son réseau épistémique.[9] Dans le cas de l'affirmation de nationalité française il faut savoir si la locutrice est française ou non.[10]

Le problème du contexte ou du cadre, typique en intelligence artificielle s'avère être un des problèmes centraux aussi bien dans la philosophie de l'esprit que dans celle du langage. Les difficultés rencontrées pour parvenir à l'individualisation des contenus représentent de véritables défis dans la tâche de faire de la psychologie une science naturelle.

De tout ce qu'on vient d'exposer, il ressort que l'individualisation du contenu dépend non seulement du *mode de présentation* mais aussi de l'agent cognitif qui est placé dans un environnement ou un milieu. La question pertinente devra donc tenter d'éclaircir le rôle de chacune de ces parties dans la détermination du contenu. Quelle est la place de l'agent dans la détermination du contenu et quelle est celle de l'environnement et par quels moyens ce dernier agit–il sur l'esprit de l'agent?

La tension existante entre le rôle de l'esprit du sujet et le rôle de l'objet qui appartient à l'environnement n'est que le produit de la superposition de la rationalité de l'agent faisant jouer l'ensemble de son réseau epistémique avec les objets qui déterminent le contenu de ses croyances.

Ceci place au centre de la discussion deux types de stratégies foodamentales en philosophie : d'un coté nous avons le débat entre l'externalisme et l'internalisme qui est enraciné dans la pure tradition analytique; de l'autre la solution phénoménologique de Husserl. Cette dernière proposition ne peut être considérée comme foncièrement analytique mais elle appartient à la tradition phénoméoologique qui l'a longtemps délaissée, peut être à tort.

Je vais traiter de ces deux stratégies dans le paragraphe suivant étant donné que leur compression s'avère être le point d'appui fondamental des différentes approches fonctionnalistes.

4.4 L'internalisme et l'externalisme

Une position externaliste extrême permettrait de soutenir que les états mentaux de deux individus identiques du point de vue des leurs état internes –donc ayant les mêmes croyances et les mêmes désirs– nés de leurs perceptions respectives de deux pommes mûres identiques placées dans des environnements identiques seront différentes.[11] La croyance qui a comme contenu "Cette pomme est mûre" sera différente ches les deux individus. En revanche un intemaliste pourrait affirmer que les deux états sont les mêmes. Voilà pour une illustration un peu caricaturale mais le problème est un peu plus complexe.

Aujourd'hui il n'existe pas beaucoup d'auteurs qui soutiennent l'externalisme extrême, pas plus que l'ioternalisme *tout court*.[12]

Une des pierres d'angle dans le bâtiment de l'externalisme a été posée par Hilary Putnam qui en 1975 publie "The meaning of 'meaning'".[13] Putman y prône une position contraire à l'internalisme; les contenus mentaux selon lui, ne sont pas déterminés par des états internes de l'agent. Pour le démontrer, la méthode qu'il utilise consiste à prouver qu'il existe une faille dans la survenance

[8] Mme. Laguiller a été candidate à la Présidence de la République lors des trois dernières échéance électorales.

[9] C'est–à–dire, l'ensemble de croyances et désirs tant explicites que tacites de Jean.

[10] Ces situations constituent la contrepartie des autres énoncés comme le déjà célèbre "la neige est blanche" et leur *vérité ne dépend pas du sujet qui les énonce, quelle que soit sa nationalité ou son réseau épistemique*. Je laisse de côté ici le cas selon lequel on peut supposer que la neige n'est pas blanche dans un monde possible. De tout façon le problème indexical existe déjà dans la monde réel donné alors que pour relativiser la vérité de l'énoncé de la "neige est blanche" il faudrait "émigrer" du monde actuel.

[11] J'emprunte cet exemple au Prof. Richard Glauser .

[12] Pascal Engel a exprimé cette situation par un jolie métaphore lors de la synthèse du Colloque *Esprit, représentation, contexte: externalisme et internalisme* du Séminaire de philosophie de l'Université de Neuchâtel. Il a dit que quand nous étions à l'école il y avait les étudiants externes que nous admirions, il y avait aussi des étudiants internes que nous méprisions mais en réalité nous étions pour la plupart des demi–pensionnaires.

[13] Voir aussi [Kripke, 1980] et [Burge, 1986]

du mental sur le physique lorsqu'on individualise intérieurement les contenus. Le but est donc de montrer que l'on peut avoir deux sujets identiques molécule-à-molécule c'est à dire, possédant les mêmes états physiques internes mais avec des états mentaux différents. Si tel est le cas, on réfute la survenance (et/ou l'identité) du "mental" sur (et/avec) le physique parce que les contenus des croyances ne surviennent pas sur les états internes de l'individu, et s'il n'y a pas de survenance les contenus ne peuvent pas être déterminés par les états internes du sujet. Putnam a proposé un exemple qui est devenu célèbre, celui des Terres jumelles. L'argument est le suivant:

> Supposons qu'il existe une autre planète, une Terre jumelle, identique en tout à la Terre, sauf que dans la première, il existe une substance que les gens appellent aussi eau, mais dont la composition chimique n'est pas H_2O. Il n'y a pourtant aucune différence perceptible car elle a les même caractéristiques que l'eau chez nous, mais sa formule abrégée est XYZ. Donc, le terme "eau" dans les deux mondes a la même intension, parce qu'elle a les mêmes propriétés physiques perceptibles. Les gens des deux planètes, lorsqu'ils parlent d'eau, sont dans le même état psychologique. Cependant les extensions de leurs contenus sont différents: pour notre planète, l'extension du mot eau est l'ensemble de molécules H_2O, et pour la terre jumelle, les molécules du type XYZ.

Putnam a suggéré cet exercice dans le cadre de la théorie de la signification pour mettre en évidence la primauté de la référence sur la signification (sens). Par la suite, cet exemple a été repris par la philosophie de l'esprit et a donné naissance à une formidable discussion.

Mais avant d'aller plus loin dans le cadre de la philosophie de l'esprit j'aimerais parler de la conception de la signification dans la philosophie du langage.

La signification chez Putnam Les théories de la signification de type fregéen reposent sur deux principes:

1. Connaître la signification d'un mot est équivalent à être dans un état psychologique déterminé.

2. La signification d'un terme (où l'on comprend signification dans le sens d'intension) détermine son extension.

Hilary Putnam nie que l'on puisse soutenir la conjonction des deux axiomes, et par là il essaie de prouver qu'il est possible d'avoir des états psychologiques identiques, qui déterminent donc une même intension quoique les extensions résultantes soient différentes. Je vais exposer les arguments principaux de [Putnam, 1975a] à la suite.

Putnam ne veut pas faire reposer la signification sur la théorie des ensembles dans le cas des extensions, mais sur quelque chose d'objectif et appartenant au monde: les structures internes des objets. L'intention de Putnam est ainsi de normaliser la signification à partir d'un vecteur de quatre dimensions. Chaque composant du vecteur représente une hypothèse dans la compétence de l'orateur, sauf l'extension qui fait aussi partie de ce vecteur. Dans le tableau suivant, nous allons montrer les éléments composants de cette normalisation, leur définition et le cas spécifique de l'eau comme exemple.

-	marqueurs syntaxique	marqueurs sémantiques	stéréotypes	extension
définitions	sont appliqués aux mots (groupe syntactique auquel le mot appartient)	sont appliqués aux mots (p. ex. animal, période de temps)	description de types additionnels le cas échéant	description de l'extension
eau	nom	concret, type naturel, liquide	sans couleur, transparent, sans savoir, désaltérant	H_2O (avec ou sans impuretés)

Cela ne veut pas dire que toutes les personnes qui parlent de l'eau savent qu'elles parlent de H_2O, mais plutôt que l'extension des mots repose sur un critère qui va au-delà des locuteurs. Voilà comment Putnam trouve, en quelque sorte, un critère opérationnel pour la détermination de l'extension. L'extension est basée sur les structures du monde, qui ne sont pas intelligibles par tous. Putnam propose une division du travail de langage. Mais admettons que je considère le bracelet que j'ai acheté comme étant en or et que je l'appelle ainsi parce qu'il a une apparence métallique de couleur jaune et d'autres caractéristiques suggestives; néanmoins, un connaisseur dira que ce n'est qu'un alliage du cuivre, d'or et probablement d'aluminium. Nier cette distinction entre apparence et essence revient, pour Putnam, à nier la dimension sociale du langage. Enfin, quelle sera la différence dans la signification du mot "eau" sur la Terre et sur la Terre jumelle? Putnam nous dit:

> In particular the représentation of the words 'water' in the Earth dialect and 'water' in the Twin Earth dialect would be the same except that in the last column the normal form description of the Twin Earth word 'water' would have XYZ and not H_2O. This means, in view of what has just been said, that we are ascribing the *same* linguistic competence to typical Earthian / Twin Earthian speaker, but different extension to the world, nonetheless. This proposal means that we keep assumption (II) of our early discuesion. Meaning determines extension - by construction, so to speak. But (I) is given up; the psychological state of the individual speaker does not determine 'what he means'.[Putnam, 1975a, page 270]

Voilà comment, chez Putnam, l'extension est déterminée par les structures du mondes, et non par l'individu, à un niveau ontologique. Mais du point de vue cognitif, l'individu peut ne pas avoir connaissance de la véritable extension du mot qu'il utilise. Ainsi, je n'arriverai probablement jamais à savoir quelle est la vraie extension du mot "or" dans la phrase "J'ai acheté ce bracelet en or".

La théorie de Putnam essaie de concilier deux aspects qui apparaissent plutôt dissociés: la connaissance des individus et les faits objectifs du monde. Il dit: "Traditional philosophy of language, like much traditional philosophy, leaves out other people and the world; a better philosophy and a better science of language must encompass both." [Putnam, 1975a, page 271]

Conséquences dans la philosophie de l'esprit Cet exercice de la peosée montre que la théorie de dépendance systématique (survenance) des états mentaux et des états physiques ne tient pas, puisqu'une réplique molécule-à-molécule de vous sur la Terre jumelle aura les mêmes états physiques internes que vous et que malgré cela vous garderez la croyance que l'eau est H_2O tandis que pour votre jumeau elle sera XYZ. [14]

4.4.1 Le contenu étroit et le contenu large

La différence faite entre les contenus qui vous sont propres et ceux de votre jumeau malgré l'identité de leurs rôles causaux nous mystifie. Comment est - il possible que l'individualisation basé sur le sens commun nous joue un si mauvais tour? Lorsque l'on crie 'Au feu! ' ici comme sur la Terre jumelle, les gens se mobilisent dans tous les cas pour amener de l'eau , ou quand il fait chaud et que vous et votre "double" êtes au bord de la mer, vous plongez tous deux dans l'eau pour vous rafraîchir.

Par conséquent, les contenus en relation avec l'eau aussi bien ici que dans l'autre monde ont les mêmes propriétés causales. Cela veut dire que les contenus ne sont pas identiques du point de vue de l'extension puisque le contexte de l'objet-eau n'est pas le même. Les conditions de vérité sont différentes et certains énoncés qui se réfèrent à l'eau ne sont pas les mêmes dans les deux cas.

[14] Comme le dit Putnam en se référant à la sémantique du langage ,

> We claim that it is possible for two speakers to be in exactly the same psychological state ..., even though the extension of the term A in the idiolect of the one is different from the extension of the term A in the idiolect of the other. Extension is not determined by psychological state. [Putnam, 1975a, page 222]

Par exemple, la proposition "la composition chimique de l'eau est H_2O" est vraie seulement dans ce monde et non sur la Terre Jumelle.

Une façon de résoudre ce problème est d'accepter un dualisme des contenus. Les deux contenus auxquels nous faisons référence dans l'exemple en question, pourraient être considérés comme étant les mêmes puisqu'ils ils sont tous deux "eau(s)" de même aspect et ont exactement le mêmes rôles causaux. Cette façon différente de considérer ou d'individualiser les contenus non selon leurs référents *mais* selon leurs rôles causaux permet de les classer en deux types différents : les contenus larges (*wide ou broad content*) et les contenus étroits (*narrow content*).

L'individualisation du contenu *étroit* se fait en vertu des états internes de l'individu; c'est ce que Fodor appelle la manière non *relationnelle*.

En revanche, le contenu *large* fait référence à des entités ou à des propriétés sémantiques specifiables seulement en mentionnant les conditions de vérité des contenus et l'environnement du sujet, c'est à dire de façon relationelle.[Fodor, 1987, pages 29–30]

Pierre Jacob résume très bien l'intérêt que présente le dualisme des contenus aux yeux de ses partisans.

> Dans l'esprit de ses partisans, le dualisme sémantique n'est pas apparu seulement comme une solution permettant de concilier des théories du contenu mental de type *informationnel*[15] avec la reconnaissance du rôle que jouent les relations *internes* (entre attitudes propositionnelles). Le dualisme sémantique est aussi apparu comme permettant de satisfaire une contrainte sur le rôle causal (ou explicatif) des propriétés sémantiques des attitudes propositionnelles d'un individu dans la production du comportement intentionnel de l'individu. Selon cette contrainte, une propriété sémantique d'une attitude propositionnelle d'un individu ne peut posséder d'efficacité causale dans le processus de production du comportement intentionnel de l'individu qu'à condition qu'elle *dépende systématiquement* des propriétés physiques du cerveau de l'individu. Si on admet ... que la fameuse expérience de pensée de Putnam (1974) démontre la propriété sémantique *large* d'une attitude propositionnelle d'un *individu*, alors le *dualisme sémantique* permet d'*espérer* qu'au *moins* les *propriétés physiques* du cerveau de l'individu peuvent servir de substrat à la propriété sémantique *étroite* de l'attitude propositionnelle de l'individu. Du moins était-ce l'espoir de Fodor (1980; 1987) lorsqu'il supposait que les lois psychologiques intentionnelles font référence aux propriétés sémantiques étroites (non aux propriétés sémantiques larges) des attitudes propositionnelles d'un individu. [Jacob, 1995, page 5–6]

Le dualisme de contenus selon Fodor en 1987 Dans le deuxième chapitre de son ouvrage intitulé *Psychosemantics* [Fodor, 1987] Fodor récuse l'importance de l'argumentation de Terre jumelles. Si on l'utilise pour l'iodividualisation non relationnelle, c'est à dire si on considère les contenus étroits à la place des contenus larges, la survenance n'est pas mise en échec, elle tient toujours.

Pour montrer la vacuité des contenus étroits dans l'explication psychologique Fodor propose un autre exercice de la pensée, moins cité en général dans la bibliographie, mais intéressant (surtout par les temps qui courent où le dollar semble être en chute libre dans les marché internationaux).

Fodor dit posséder une pièce de 100 cents de dollars capable de contrôler toutes et chacune des particules de l'univers. Si sa pièce est côté pile dans l'instant t alors toutes les particules de l'univers sont définies comme étant des H particules à l'instant t. En revanche, si sa précieuse pièce est côté face à l'instant t alors toutes les particules de l'univers vérifient le prédicat "être une particule T à l'instant t". [Fodor, 1987, page 33] En dehors de cette définition des particules comme étant H ou T en fonction de la pièce de Fodor, il n'y point de changement dans l'univers, toutes les lois de la physique se vérifient comme auparavant.

Pourquoi serait-il non pertinent (ou une folie, comme le dit Fodor) de tenir compte dans une théorie physique du fait que les particules soient définies comme T ou comme H? Tout simplement parce que *ce fait, le fait d'être définie comme T ou d'être comme H, n'a aucune pouvoir causal.* Dès lors, il en va de même pour la propriété d'être un état mental d'un personne que vit dans un contexte où l'eau est H_2O ou XYZ. Le fait de tenir compte d'une telle propriété n'a aucun rôle explicatif/causal dans une théorie psychologique.

[15] J'adopte à titre provisoire la définition suivante du terme *informationnel* : terme qui caractérise un type de théorie externaliste dont le but est de mettre au clair la genèse du lien sémantique. À la base de cette explication on trouve le concept d'*information* qui a comme fonction d'indiquer les données echangées par des objets auxquels nous actes intentionnels renvoient. Je reviendrai ultérieurement sur ce terme pour apporter des précisions.

Selon Fodor la situation pourrait se résumer comme suit:

> In short, what you need in order to do science is a taxonomic apparatus that distinguishes between things insofar as they have *different* causal properties, and that groups things together insofar as they have the *same* causal properties. So now we can see why it would be mad to embrace a taxonomy that takes seriously the difference between H-particles and T particles. All else being equal, H particles and T particles have identical causal properties; whether something is an H-(T) particle is irrelevant to its causal powers. To put it a little more tensely, if an event e is caused by H-particle p, then that *same event e is also caused by p in the nearest nomologically possible world in which p is T rather* than H... So the properties of being $H(/T)$ are taxonomically irrelevant for purposes of scientific causal explanation. ...
>
> And similarly, *mutantis mutandis, for the property of being a mental state of a person who lives in a world where there is property of XYZ rather than H_2O in the puddles.* These sorts of differences in the relational properties of psychological (/brain/particle) states are irrelevant to their causal powers; hence, irrelevant to scientific taxonomy. [Fodor, 1987, page 34, les morceaux en italique ont été ainsi soulignés par Fodor dans le texte original]

De la même manière que la propriété d'"être définie comme étant une particule T/H" n'a aucune pertinence pour une théorie physique car cette propriété n'a pas de pouvoir causal, le *contenu large* n'a pas de pertinence pour une théorie psychologie.[16]

Fodor relativise ainsi l'importance du problème de la Terre jumelle et des autres expériences de la pensée du même type. Il soutient qu'ils ne coupent pas la connexion entre les extensions et les contenus mais que tout simplement ils la font se rapporter au contexte. Or Fodor, à cette époque, récuse l'externalisme au bénéfice du dualisme de contenus .

Cependant, la notion de contenu étroit telle que Fodor la présente n'est pas non plus exempte de problèmes. Quels sont les problèmes qui s'avèrent pertinents dans la notion de *contenu étroit*? Les difficultés d'un telle approche sont les suivantes. Premièrement, le contenu étroit mérite à peine le nom de contenu. Comme Joëlle Proust l'a bien signalé

> [...] le contenu étroit paraît à peine mériter le nom de contenu dans la mesure où il n'est *pas encore sémantiquement évaluable*, puisqu'un contexte n'est pas encore donné qui rende possible cette évaluation. Par opposition à un tel contenu étroit, ce qu'on appelle contenu dans la réflexion sémantique traditionnelle comme la pensée frégéenne, la proposition en soi de Bolzano ou la phrase éternelle de Quine ont une valeur de vérité déterminée en ce sens qu'elles *incluent* les déterminants contextuels du sens. [Proust, 1990, page 24, italiques dans le texte original]

En second lieu, on ne voit pas comment on pourrait déterminer l'extension à partir du contenu étroit même une fois le contexte fixé.

La discussion de ces critiques requiert qu'on fasse appel à un nouveau concept, le langage de la pensée (*The langage of thought*). Je reviendrai sur le problème de la détermination de l'extension et de la relation entre contenu étroit et contenu large lorsque je développerai la théorie de Fodor dans le chapitre 7 (§7.5.2 ss).

4.4.2 La théorie externaliste de Drestke

Fred Drestke a proposé une théorie externaliste dont l'objectif est d'expliquer la genèse du lien sémantique et par cela même comment les propriétés relationnelles extrinsèques peuvent être considérées comme des causes.

En s'inspirant de la théorie de l'information proposée par Shannon il a pour but de mettre en évidence le type de lien entre les contenus et les état des choses auxquels il se réfèrent. Ce lien est pour lui un lien informationnel.

La notion de contenu informationnel est la suivante :

DEFINITION 1
Un signal qui porte l'information que quelque chose est F sans porter l'information qu'elle est G malgré le fait que tous les Fs sont des Gs.

[16] Je reviens sur ce problème dans le chapitre 7.

Par exemple (emprunté à Drestke), Herman entend la sonnette de sa porte et marche vers la porte. Nous pouvons en déduire qu'Herman croit que quelqu'un est à la porte. Mais le fait que quelqu'un sonne à la porte implique aussi que le bouton de la sonnerie a été pressé. En effet, chaque fois que la sonnette retentit, le houton a été pressé; cependant l'information dont ce signal est porteur pour Herman est plutôt que quelqu'un est à la porte .

Or la notion de contenu informationnel est une notion intentionnelle parce que seules quelques unes des caractéristiques des objets de l'environnement font partie dudit contenu. Elle est néanmoins différente d'une autre notion intentionnelle: le contenu sémantique.

Dans [Drestke, 1981] on signale la différence entre contenu informationnel et contenu sémantique. L'information contenue dans une structure definit un contenu propositionnel qui est intentionnel parce qu'une structure qui porte l'information que t est F peut ne pas révéler un F qui est aussi un G. Si F est G analytiquement alors le contenu ayant comme seule proposition que t est F est considéré un *contenu sémantique*.

Le système a choisi une seule de toutes les informations qui sont apportées par la source; il a donc fait une *interprétation*. Cette interprétation consiste en la transformation de l'information contenue par le signal que produit la source de manière analogue au système digital.

Pour modéliser ce processus informationnel nous serons tentés de postuler l'existence d'un lien causal entre la source A et le récepteur de l'information B. Cependant, les relations de causalité entre A et B ne suffisent pas pour conclure qu'un flux d'information se produit entre A et B. Par exemple une mouche dans le champ de vision d'une grenouille provoque une excitation neuronale particulière qui déclenche à son tour la réponse "happer au vol". [Proust, 1990, cfr.page 26] mais ce n'est pas pour autant que l'on peut affirmer que la grenouille a *interprété* ni même perçcu le contenu de la proposition suivante: *il y a une mouche dans mon environnent.*

Cette thèse s'avère trop stricte pour rendre compte d'une notion intentionnelle telle que la transmission et la réception de l'information, et c'est pourquoi Drestke fait la différence entre causalité et *régularité nomique*[17][Drestke, 1981, cfr. page 33],[Proust, 1990, cfr. page 26] tout en adoptant ces régularités comme pertinentes au processus intentionnel.

Processus sensoriel et processus cognitif: de la structure au contenu Fred Drestke a commencé son travail philosophique dans le domaine de l'épistémologie. Il a récusé les théories de la connaissance basées sur une justification épistémique.[18] Il met en avant, en revanche le *caractère*

[17]Par *régularités nomiques* il faut comprendre, pour le moment, des relations non–déterministes mais contrefactuellement stables. Je reviendrai sur ce point dans la prochaine section.

[18]Ce type de théories de la connaissance doit répondre à la question suivante: si la connaissance nécessite que nos croyances s'avèrent justifiées, alors ces justifications doivent constituer un réseau *épistémique*. Par exemple supposons que quelqu'un croie que l'on peut réaliser tout ce qu'on veut si on travaille suffisamment. La justification de cette croyance est qu'elle l'incite à entreprendre de choses qu'il ne serait pas disposé à tenter autrement. Cette justifieation est psychologique mais non–épistémique. Une telle croyance peut être bénéfique à la personne mais on ne peut pas déduire d'elle la vérité de la proposition "un personne peut réaliser tout ce dont elle a envie".

La justification épistémique se doit de démontrer que ce qu'on pense de juste est aussi vrai. Cela veut dire qu'elle met l'accent sur la connexion entre les croyances que l'on considère justifiées et leur probabilité d'être vraies. Ceci amène tout de suite au problème de la régression infinie. En effet, supposons que vous ayez une croyance justifiée. Même dans le meilleur des cas, lorsqu'il s'avère que cette croyance est vraie, elle n'est pas justifiée pour autant car il reste encore à démontrer les raisons sur lesquelles notre croyance se fonde. Mais les raisons que vous pouvez donner ne sont pas elles–mêmes justifiées à moins que vous puissies donner des raisons justifient les raisons précédentes, et ainsi de suite. Le problème de la régression infinie est une des armes dont disposent les sceptiques pour nier l'existenes de la connaissance. Une des stratégies contre cette argumentation est le fondationisme.

L'idée du fondationisme est qu'il existe une fondation de la connaissance qui se base sur certaines croyances, mais que ces dernières ne sont justifiées que par leurs relation à d'autres croyances. Ce sont des croyances de base: immédiates et justifiées de manière non–inférentielle. Elles servent à la justification des autres croyances qui ne sont pas des croyances de base et qui ont donc besoin d'une médiation inférentielle.

Le débat épistémologique de ce siècle porte sur la nature ou sur les caractéristiques que ses croyances de base doivent avoir.

Roderick Chisholm défend une version forte du fondationisme selon laquelle les croyances de bases sont directement évidentes. Par exemple, il ne vous semble pas que vous ayez un rage de dents et toute la justificstion dont vous avez besoin pour justifier que vous avez un rage de dents est d'avoir effectivement une rage de dents. Dans son texte [Chisholm, 1977] ils discute un certain nombre des principes épistemiques qui font le lien entre les faits directement évidents et la plupart des nos croyances que ne le sont point. Il affirme qu'une partie de la justification

modal de la connaissance. Selon Drestke, on fait la justification d'une croyance en analysant les conditions nécessaires présentes pour qu'elle soit vraie, mais aussi les autres conditions nécessaires dans des cas similaires; non seulement la situatioo telle qu'elle est mais aussi telle qu'elle aurait pu être. L'utilisation de la théorie de l'information lui donne la possibilité de mettre en évidence le lien existant entre une croyance justifiée et la connaissance qui en découle et aussi entre une expérience et la perception dont elle est l'origine.

Ainsi, lui même a résumé son parcours :

> Even more basic to my view of the mind is the externalism I brought from my work in epistemology. Knowledge [...] is not a matter of justification, not a matter of getting your belief secured by an evidential chain to a foundational rock. It is, rather, a matter of such belief being connected to the facts in the right way, a relationship whose existence, because external or extrinsic to the total system of beliefs, might be quite unknown (perhaps even unknowable) to the knowing mind. Sense perception is one way, the most direct and reliable way, of getting oneself so connected.[Drestke, 1994, page 260-261]

Le but de Drestke est d'expliquer la genèse du lien intentioonel en utilisant des bases non intentionnelles dans un perspective externaliste, c'est–à– dire en évoquant des propriétés relationnelles extrinsèques du contenu. Ainsi, il propose une théorie de l'information qui permettra une description du rôle causal du contenu en termes des caractéristiques nomiques de la nature.

Rappelons–nous quelques concepts de base de la théorie de l'information. Soit s une source, N la totalité des événements possibles et n sous–ensembles désigné par l'information. La *quantité d'information* sera :

$$I(s) = log_2(N/\mathrm{n})$$

Considérons maintenant non un seul sous–ensemble mais une partie séparée de l'ensemble original. Appelons E l'ensemble original et $E_1, \ldots, E_n$ ses différents éléments. La quantité d'information liée à E_i sera par définition

$$I(E_i) = log_2(N/n_i)$$

la quantité d'information que l'on appelle *entropie* sera

$$I(s) = \sum_i p(E_i).I(E_i) = \sum_i \frac{n_i}{N} log_2 \frac{N}{\mathrm{n_i}}$$

Cette notion est applicable à un ensemble d'éléments régi par une loi de probabilité, car toute loi de probabilité a la propriété de définir une division de l'ensemble des événements (entre événements favorables et événements défavorables par exemple). La tbéorie de l'information mesure la quantité d'information de trois entités : la quantité d'information engendrée par la source, la quantité d'information transmise et la quantité d'information perdue.

En utilisant le concept de quantité d'information de la tbéorie mathématique des signaux Drestke peut définir les conditions pour qu'un signal r transmette une information, par exemple que s est F

CONDITION 1

Si un signal est porteur de l'information que s est F, alors la quantité d'information véhiculée par le signal est au moins aussi grande que la quantité d'information contenue dans le fait que s est F (p. ex. s est rouge).

Cette condition est nécessaire mais non suffisante. Ainsi, il se peut qu'un signal ait besoin de 2 bits pour porter l'information sur la couleur. Aucun signal ne peut porter l'information sur la

des propositions que ne sont pas directement évidentes est obtenue par l'intermédiaire des celles qui le sont.

En revanche, Alvin Goldman met l'accent sur le processus même de formation des croyances; il a une approche pour ainsi dire plus fonctionnaliste du problème. Pour lui les caractéristiques que l'on donne aux croyances dignes de fois sont inspirées du processus de formation des croyances justifiés. L'explication qu'une croyance soit sûre, digne de fois (de l'anglais *reliable*) doit être cherchée dans les précurseurs. Ainsi Goodman met l'accent sur la connexion causale entre la croyance justifiée et ce qui fonde ou justifie cette croyance. [Goldman, 1967]

couleur (s est rouge) s'il ne consiste pas d'au moins deux bits. Cependant un signal peut porter l'information sur une couleur de s (par exemple que s est bleu) sans contenir l'information que s est rouge. Voilà pourquoi on a besoin d'une deuxième condition.

CONDITION 2
s est effectivement F

Les deux conditions sont individuellement nécessaires mais l'union des deux n'est pas suffisante. Soit un signal qui a besoin de trois bits pour transmettre la forme (p. ex. un carré) et de trois autres bits pour transmettre la couleur (p. ex. rouge). Si la signal transmettant que s est un carré vérifie les deux conditions et néanmoins ne transmet pas l'information que s est rouge, la quantité d'information sera suffisante mais l'information transmise n'est pas la bonne (que s est rouge). Il faut satisfaire une dernière condition :

CONDITION 3
La quantité d'information dont le signal est porteur comprend l'information engendrée par le fait que s est F.

Lorsqu'on observe la définition mathématique de l'entropie d'une partie on se rend compte que l'utilisation des probabilités conditionnelles nous oblige à supposer l'existence de l'espace universel d'événements dans lequel la fonction probabilité est définie.[19]

Selon l'approche de Drestke ceci se traduit par le fait que le récepteur doit connaître toutes les possibilités pour un signal de la source. Par exemple, dans le cas de la couleur, le récepteur doit savoir qu'il s'agit d'un ensemble de deux couleurs (rouge et bleue).

La valeur k représente ce que le récepteur connaît déjà de la source du signal.

A partir des trois conditions citées plus haut, Drestke peut définir la notion de contenu informationnel comme suit :

DEFINITION 2 (CONTENU INFORMATIONNEL)
Un signal r est porteur de l'information que s est F si est seulement si la probabilité conditionnelle que s soit F, étant donné r (et k), est égale à 1.

La définition de contenu informationnel satisfait les trois conditions citées plus haut. La première condition est satisfaite parce que si la probabilité conditionnelle de l'événement s est F étant donné r, alors toutes les probabilités du complément seront nulles, et si l'on remplace cette valeur dans la formule d'entropie, on obtient que la quantité d'information de s est la même que celle de s est F.

La deuxième condition est satisfaite trivialement étant donné que la probabilité que s soit F est égale à 1.

La troisième condition est satisfaite parce que la définition affirme explicitement que le signal porte la quantité pertinente d'information en vertu du fait que l'on a exclu toute situation qui contraindrait à la réduire.

Une autre caractéristique du contenu informationnel est qu'il ne garantit en rien l'unicité de l'information. Toute signal qui porte l'information que s est F porte aussi toutes les information implicites, analytiquement ou nomologiquement dans ce fait. Par exemple, si s est F porte l'information que s est un carré alors le signal porte aussi l'information que s est un rectangle. Cette non–unicité des contenus fait la spécificité de l'information et cause les différences de signification. En effet, l'énoncé "ceci est un carré" ne signifie pas "ceci est un rectangle".

Or, le contenu informationnel d'une structure s'avère être plus large que son contenu sémantique.

La démarche de Drestke a les mérites suivants; d'un côté l'information portée par une signal dépend de l'existence des corrélatioos nomiques entre les propriétés du signal et celles de la source. Par exemple, si un signal porte l'information que le mercure se dilate, il porte aussi l'information que la température monte; ainsi le contenu concernant la montée de la température est

[19] Pour une discussion en detaille [Kistler, 1995a, page 612-620]

induit à partir des corrélations nomologiques. Il y a d'autres exemples établis sur des corrélations analytiques.

D'autre part, le fait que le contenu informationnel d'une structure soit plus large que celui du contenu sémantique permettra de justifier l'opacité des contenus mentaux. En effet, c'est la sélection faite dans l'information par l'agent à partir de la structure du contenu qui permettra de justifier, par exemple le fait que "Jean croie que l'étoile de matin est la planète Vénus" sans qu'il croie le même prédicat pour "étoile de soir".

La situation est différente lorsque les informations impliquées ne le sont ni nomologiquement ni analytiquement mais qu'elles sont coréferentielles de manière contingente. Or le signal "s est F" ne porte pas toutes les informations enchâssées de manière contingente par ce contenu.

Soit l'exemple suivant [Drestke, 1981, cf. page 74]:

EXEMPLE 1
Supposons que tous les enfants d'Herman aient contracté la rougeole. Le fait d'être un enfant d'Herman et d'avoir la rougeole est une corrélation contingente. Soit Alice une enfant d'Herman. Un signal peut être porteur de l'information qu'Alice est un des enfants d'Herman sans pour autant être porteur de l'information qu'Alice a la rougeole.

L'exemple montre au premier abord que la probabilité conditionnelle qu'Alice ait la rougeole étant donné qu'elle est la fille d'Herman n'est pas l'unité. Drestke l'attribue au fait que les enfants ont deux parents puisque il y n'a pas une loi de la nature ou une nécessité métaphysique qu'il en soit ainsi. Mais comme l'indique Élisabeth Pacherie[20] cette argumentation n'est pas convaincante. Il n'y a rien d'incohérent à donner à la théorie des probabilités une interprétation extensionnelle à la place de l'interprétation intensionnelle qui découle de l'exigence d'une loi de la nature ou d'une nécessité logique. Le problème est ailleurs.

Rappelons que Drestke avait pris la notion d'information de la théorie des signaux pour tisser un lien entre les croyances et la connaissance sans faire appel à d'autres concepts que l'états des affaires et les notions en rapport avec l'information. Cependant, l'exemple montre que ceci ne suffit pas et qu'il faut postuler une corrélation nomique entre les propriétés du signal et les propriétés de la sources. Ainsi Pacherie signale:

> Il y a donc quelque chose de trompeur dans la définition qui est donnée par Drestke du contenu informationnel. En définissant le contenu informationnel en termes de probabilités, il occulte en fait l'essentiel. Le fait qu'un signal ne puisse être porteur de l'information que s est F que si la probabilité que s est F, étant donné le signal, est égale à un n'est qu'une conséquence d'un fait plus important, à savoir qu'un signal ne peut être porteur de l'information que s est F que s'il existe une corrélation nomique entre les propriétés du signal et les propriétés de la source. C'est donc cette dernière condition qui est en fait la condition essentielle d'une définition de contenu informationnel. [Pacherie, 1993, page 219]

En conclusion, il semble que Drestke n'a pas totalement atteint son but. En effet, quand il s'agit des corrélations contingentes des informations enchassées dans un signal, il appert qu'il faut plus que les notions relatives au concept d'information. L'explication des la genèse du lien sémantique externaliste semble avoir ainsi échoué.

Le contenu sémantique Le contenu informationnel, nous l'avons vu est une structure plus large que le contenu sémantique. En effet, prenons une exemple de Drestke: la phrase "Cette étendue d'eau est en train de geler" ne signifie point que "cette étendue d'eau est en train de s'étendre" malgré le fait que la deuxième information soit enchâssée dans la première.

DEFINITION 3 (CONTENU SÉMANTIQUE)
Une structure S a pour contenu sémantique que t est F si et seulement si:

1. *S est porteur du contenu informationnel que t est F*

[20] [Pacherie, 1993]

2. *le contenu informationnel que t est F n'est pas enchâssé dans un autre contenu informationnel dont S serait également porteur.*

Les contenus sémantiques ont une intentionnalité plus grande que le contenu informationnel. Drestke distingue trois ordres différents d'intentionnalité. Soit le contenu informationnel s est F et le contenu s est G, alors la structure informationnelle qui a l'information que s est F sans avoir l'information que s est G aura un ordre d'intentionnalité selon que tous les F se trouvent être des G en vertu de:

1er **ordre** Une relation coextensionnelle.

2ème **ordre** Une loi naturelle.

3ème **ordre** Nécessité analytique.

Dans le contenu sémantique s'est opéré une perte d'information en relation au contenu informationnel qui en est à l'origine.

Drestke dit que cette différence est semblable à la différence entre représentation analogique et représentation digitale.

> I will say that a signal (structure, event, state) carries the information that s is F in *digital* form if and only if the signal carries no additional information about s, no information that is not already nestled in s's being F. If the signal *does* carry additional information about s, information that is *not* nestled in s's being F, then I shall say that the signal carries this information in analog form. When the signal carries the information that s is F in analogue form, the signal always carries more specific, more determinate information about s than that it is F. Every signal carries information in both analog and digital form. The most specific piece of information the signal carries (about s) is the only piece of information it carries (about s) in digital form. All other information (about s) is coded in analog form [Drestke, 1981, page 137]

La différence entre analogique et digital montre le clivage entre contenu informationnel et contenu sémantique. La conversion d'une structure de l'analogique au digital comporte une perte d'information. Dans ce processus les parties de information qui sont irrelevantes sont mise de côté. La forme de codage digital correspond à des représentations cognitives tandis que la forme digitale correspond à un format sensoriel. Avant la conversion en digital, l'information ne peut pas être soumise à des processus cognitifs supérieurs tels que des catégorisations, ou des généralisations.

C'est le format digital qui permet de reconnaître une entrée comme étant une instance d'un type plus général. Le passage d'analogue à digital est aussi motivé par le fait que l'information provenant des sources de l'environnement est trop vaste pour les capacités des systèmes perceptifs.

> What *is* digested are bits and pieces of information the sensory structure carries in analog form. [Drestke, 1981, page 148]

Par ailleurs, la différence que Drestke fait entre les deux types de contenus sert à éclairer le dualisme de contenus (étroit – large) prôné par Fodor. Comme le remarque Joëlle Proust

> Dans la mesure où elles éclairent la complémentarité du contenu étroit et du contenu large, les analyses de Drestke apportent de l'eau au moulin de Fodor: le "*narrow content*" renvoie à la structure cognitive d'un concept (c'est-à-dire à ses effets et conséquences fonctionnels), tandis que le "*wide content*" renvoie aux origines informationnelles des concepts. [Proust, 1990, cfr. page 26]

Mais revenons aux concept des régularités nomiques. Ce ne sont pas des lois de la nature mais seulement des contrefactuelles stables, donc comment l'agent a-t-il accès à elles? Par quels moyens peut-il appréhender ces régularités?

Selon Drestke, l'acquisition des régularités nomiques se fait à partir de l'apprentissage permis par l'exposition des signaux porteurs de cette information. Ce sont ces régularités nomiques qui déterminent le contenu informationnel de concept "eau" dans le cas des Terres Jumelles. Ainsi, le concept "eau" a des valeurs informationnelles différentes; cette différence a son origine dans les différences entre les types d'information auxquels vous et les habitants de votre terre jumelle étiez exposés pendant l'apprentissage et ceci malgré le fait qu'à oeil nu deux échantillons appartenant

chacun à une des deux terres soient impossibles à differentier. Bien que la notion de contenu informationnel selon Drestke comme le note Proust aille dans le sens de la dualité étroit/large, il faut souligner que les explications de Fodor et de Drestke sont foncièrement différentes. Pour Fodor les contenu large n'ont aucun rôle à jouer dans les explications en psychologie, mais le mécanisme d'appropriation du contenu étroit pertinent pour une situation donnée reste un mystère. Une preuve de cela est qu'il n'existe aucune explication de la relation perception–intentionnalité chez Fodor; pis encore il ne voit pas pourquoi il serait pertinent d'en donner une (communication personelle). La métaphore de l'existence d'un oracle est pour lui plus que suffisante à cet égard.

Drestke, en revanche fournit une explication de la manière d'appropriation du contenu; le sujet et l'environnement entretiennent des relations qui déclenchent des processus d'apprentissage de leurs régularités nomique chez le sujet. Ces processus renforcent la détection d'une partie de l'information totale portée par la source en détriment des autres éléments.

Jusqu'à maintenant j'ai exposé le concept de contenu informatif et j'ai essayé de circonscrire l'externalisme dretskien, mais il reste à répondre aux deux questions que l'on s'est posé auparavant: comment se fait–il que les propriétés sémantiques aient un corrélation avec les propriétés physiques? et d'autre part, comment est–il possible de concilier les raisons avec les causes sans abandonner la perspective physicaliste?

Il s'agit d'expliquer, par exemple pourquoi Herman va–t–il à ouvrir la porte (action physique) quand il entend sa sonnerie (état mental résultant d'uo perception auditive) et "Comment la sonnerie cause–t–elle que Herman aille vers la porte pour l'ouvrir? "

Les questions ainsi posées sont très complexes mais on pourrait prendre une analogie avec une situation similaire [Kim, 1991, cf.].

On pourrait dire, "Pourquoi le thermostat a–t–il déclenché la chaudière? " question que l'on pourra à son tour diviser en deux: (1)"Pourquoi le thermostat a–t–il déclenché la chaudière maintenant? " et (2) "Pourquoi le thermostat a–t–il déclenché la chaudière plutôt qu'autre chose, par exemple l'éclairage des couloirs? "

Une réponse à (1) pourrait être "Parce qu'il a détecté une température inférieure à 20 degrés maintenant". Cette réponse donne une information sur la *cause* du déclenchement.

Une réponse à (2) pourrait être: "Parce que les circuits électriques sont faits de façon telle qu'ils relient le thermostat à la chaudière plutôt qu'à autre chose comme par exemple l'éclairage des couloirs? " et ceci est une *raison structurelle*.

Pour Drestke le rôle des raisons dans l'explication causale est de structurer les causes des résultats moteurs ou des actions.

Comment peut–on appliquer ce type d'explication au cas d'Herman? Appelons la propriété F la résonance de la sonnerie de la porte d'Herman. Comment est–il possible qu'un état interne du système en l'occurrence Herman, en vertu de son contenu (la résonance de la sonnerie de la porte de chez Herman) explique pourquoi l'état interne (entendre la sonnerie de la porte) est relié avec la réponse motrice (aller vers la porte). Une explication pourrait être que pour le système (Herman) il est avantageux de produire la réponse motrice (aller vers la porte) lorsque la propriété F se présente dans l'environnement. Comment est–il possible de défendre l'existence d'une liaison confiable entre la présence de F et la production de l'action (aller vers la porte)?

Tout d'abord Herman est muni d'un récepteur, disons d'un F–détecteur et dans le cas qui nous occupe le système auditif d'Herman sera une des composantes de ce récepteur. Ce détecteur est activé toutes les fois que la présence de F se produit dans l'environnement, disons que le détecteur de F se met d'un certaine façon. Le résultat de détecter F dans le voisinage d'Herman, produit chez lui un certain état interne qui est une instantiation de la propriété F: l'état interne qui peut être exprimée par "percevoir que la sonnerie de la porte a résonné".

On peut considérer que l'état interne ("percevoir que la sonnerie de la porte a résonné") est un "token" référent à la propriété F.[21]

Mais ceci a lieu parce que l'état interne d'Herman (celui de "percevoir que la sonnerie de la porte a résonné") a une propriété neurologique que nous nommerons N, et ceci est *en général* le

[21] Il n'est pas clair que Drestke soit partisan de considérer ces états internes comme des *tokens* (identités occasionnelles)

cas lorsque la présence de F apparait dans l'entourage d'Herman. Or, la détection dans le temps t de la propriété F consiste en la présence de l'état interne du token doté de la propriété N.

Au point où nous sommes arrivés, assurer l'existence d'une relation fiable entre la présence de F dans le voisinage d'Herman et le résultat moteur ("aller vers la porte") revient à affirmer la seule connexion fiable entre l'occurrence d'un état neurologique N et le résultat moteur.

Pour Drestke cette relation est assurée par le fait que, pour Herman, il est avantageux de produire le résultat moteur ("aller vers la porte") chaque fois que se vérifie la présence de F dans son entourage ("la sonnerie de sa porte résonne") et ceci est le résultat de l'apprentissage qu'Herman a fait auparavant. Herman a développé un réseau épistemique des croyances (p. ex. "que le facteur ne sonne qu'une fois", "que les voisin peuvent avoir besoin de lui", "que sa grand-mère vient lui rendre visite", "qu'il ne s'agit pas d'un créancier parce qu'il est trop tard pour que les gens travaillent encore") dans le processus d'apprentissage de sa vie et que ces croyances ont renforcé la relation entre N et le résultat moteur ("aller vers la porte").

Cependant il me faut encore montrer comment les propriétés sémantiques deviennent des causes et aussi quelle est leur corrélation avec des états physiques tels que, dans ce cas N? C'est ici qu'intervient le concept de contenu informationnel. L'existence d'une corrélation nomique entre F et la réponse motrice ("aller vers la porte") présuppose l'occurrence occasionnelle d'un état interne portant le contenu informationnel F ("la sonnerie de la porte a résonné") et implique que cet état interne ("percevoir que la sonnerie de la porte a résonné") cause l'action par le fait d'être une instance de l'état physique N. Ainsi on dit que l'état interne de Herman ("percevoir que la sonnerie de la porte a résonné") a pour propriété la représentation du F comme contenu *intentionnel*. Le fait que la réponse motrice "aller vers porte" soit considérée comme une conséquence de l'état N est dû essentiellement au caractère représentationnel de N. En effet ladite réponse est simplement le résultat de l'acquisition d'un *fonction d'indication* de F. En conclusion, Dretske sauve le physicalisme en soutenant que la cause de l' action motrice est l'état physique N, mais que l'état N est la cause parce que le système en question a acquis la fonction d'indication de F, ce qui veut dire qu'en fonction du contenu et de cette façon nous avons répondu aux deux questions posées au début.

4.4.3 Le problème de la méprise représentationnelle

Toutefois il existe une ombre dans ce beau tableau comme le note Dretske:

> [...] since the information a structure carries cannot be false, a structure's semantic content cannot be false. But we can certainly have false beliefs.[Drestke, 1981, page 1909]

Ce que Drestke vient d'évoquer dans cette citation est le problème de la *méprise représentationnelle* commun à toutes les théories causales ou informationnelles du contenu. Pour Drestke, c'est au cours l'acquisition de cette fonction d'indication que les méreprésentations deviennent possibles. Il est plausible qu'Herman ait par exemple une hallucination auditive sur la sonnerie de sa porte ou qu'il confonde la sonnerie de son voisin avec la sienne. Alors Herman éprouvera l'état N et bien que F ne soit pas présent, Herman aura le contenu de représentation F. Pour Drestke, si un état représentationnel est susceptible de méreprésentation, alors le système qui le possède exhibe une intentionnalité appropriée à la génération des croyances. En définitive, un système intentionnel doit être capable de méprise représentationnelle.

Le problème de la disjonction

La théorie de Drestke relève d'un groupe d'homologues qui partagent une stratégie commune: elles expliquent la relation de référence entre une représentation et son objet par l'existence d'une relation convergente de causalité entre l'objet ou ses propriétés dénotées et la proposition que sa représentation exprime. Toutes ces théories groupées sous le nom de théories causales/informationnelles comportent le même problème: elles ne peuvent expliquer de façon satisfaisante les cas de méprise représentationnelle que Fodor a décrit comme le *problème de la disjonction*.

Supposons que je suis en train de me promener dans le Jura un matin d'automne et qu'il y ait du brouillard. Lors de ma promenade dans la campagne je vois un animal et je me dis "tiens,

une vache". Pourtant il ne s'agit pas d'une vache mais d'un cheval. Selon la théorie causale, il faut soutenir que le contenu de la représentation "vache" est la propriété disjonctive "être soit une vache soit un cheval" parce qu'il doit exister une relation causale convergente à la référence entre l'objet ou la propriété et sa représentation. La propriété disjonctive permettra donc de surmonter ce problème. Cependant, cette solution n'est pas satisfaisante parce que malgré que l'on l'on conserve une propriété convergente à la référence, les théories causales/informationnelles deviennent, par le même fait, impuissantes à rendre compte des erreurs de représentation. En effet, elles ne parviennent pas à distinguer un concept disjonctif ("être une vache ou un cheval") de la méprise représentationnelle qui est de faire tomber quelque chose qui n'est pas une vache, en l'occurrence un cheval, dans l'extension de ce concept. Le défi est de pouvoir expliquer le problème disjonctif en termes autres qu'intentionnels et d'éviter ainsi une solution circulaire.

Il serait préférable, de plus, de ne pas faire appel aux concepts étrangers à la théorie jusqu'à présente, comme par exemple d'invoquer "les circonstances normales", "la sélection naturelle faite par l'évolution", entre autres.

La solution proposée par Fodor en *Psychosemantics* est de postuler une relation d'esymétrie entre l'occurrence de la propriété causale de la représentation "vache" et sa représentation.

C'est parce que les vaches donnent lieu à la représentation de "vache" que les chevaux conduisent parfois à se représenter une vache, tandis qu'il n'est pas vrai que symétriquement ce soit parce que les chevaux produisent le symbole "vache" que les vaches conduisent à se représenter "une vache".

> Cette théorie permet d'expliquer l'erreur en termes non-intentionnels, empiriques, de dépendance asymétrique entre des relations causales. Selon la formule de Fodor (1988) "les occurrences fausses dépendent métaphysiquement des vraies"(page 19) [22] [Proust, 1990, page 28]

J'ai exposé le problème de la disjonction et aussi du contenu large et étroit en parlant de représentation dans son sens large bien que dans les écrits cités ici la représentation à laquelle on se réfère soit celle de symbole dans le langage de la pensée.

Dans les théories de type externaliste comme celle qui vient d'être exposée le concept pour ainsi dire "s'impose" au sujet et la relation est *objective*. Il faut rappeler que la relation de référence entre une représentation et son objet est la résultante d'une relation convergente de causalité entre l'objet ou les propriétés qu'il dénote et la proposition que cette représentation exprime. Le rôle du sujet est plutôt passif et le sujet diffère radicalement de l'environnement.

La phénoménologie présente un tableau different du précédent. Le rôle du sujet est un rôle actif ainsi que les objets qui se présentent à lui. La relation intentionnelle est *réflexive* parce que l'acte est en quelque sorte "plié sur lui même".

Nous allons expliquer qu'il existe deux manières d'approcher la théorie de Husserl. La première conçoit les relations entre sujet et objet comme totalement semblables à la relation fregéenne. La deuxième, au contraire, tient compte de la richesse de la relation intentionnelle ainsi qu'Husserl mais interprète mal certains concepts en tentant de les insérer dans le cadre de la *Gestaltheorie*.

Toutefois j'aimerais exprimer deux réserves sur mon exposé de la théorie de Husserl. La première est que je ne prétends pas exposer ici la théorie de Husserl en elle-même, mais expliquer l'usage que l'on en fait. L'autre est que j'ai volontairement laissé de côté des aspects importants et fondamentaux de l'approche phénoménologique. Je songe ici aux développements de Martin Heidegger (1889–1976). Les raisons qui m'ont amenée à faire ce choix tiennent d'une part au but instrumental que cette partie poursuit dans le cadre de mon travail. En effet, il s'agit de montrer les usages qui s'avèrent pertinents aux explications et critiques postérieures que je dédie ensuite au modèle du Professeur Jean Petitot. D'autre part, l'exposé de la problématique à partir de Husserl et de Heidegger exigerait un ouvrage entier et il dépasse donc les limites de ma propre recherche dans le but initial que je me suis fixé.

4.5 L'approche phénoménologique

Edmund Husserl (1859–1938) est le créateur de l'important courant de la *phénoménologie*.

[22] Le texte de Fodor cité par Proust est un manuscrit intitulé *Information et représentation*

Mathématicien de formation, il travailla comme assistant de Karl Weierstrass et soutiot sa tbèse de doctorat en 1882 sous le titre de *Contributions à la théorie du calcul de variations*. Il se tourne vers la philosophie à cause de Franz Brentano dont il devient l'élève à Vienne vers 1883. D'origine juive il se convertit au lutbérianisme en 1886. Il a été professeur extraordinaire à l'Université de Göttingen en 1901 mais cette même université lui refuse en 1905 le titre de professeur ordinaire pour "manque de qualifications scientifiques".

En 1916 il devient professeur ordinaire à l'Université de Fribourg–en–Brisgau et sa leçcon inaugurale a pour sujet *La phénoménologie pure, son domaine de recherche et sa méthode*. En 1917 il est nommé conseiller secret du grand–duc de Bade. En 1919 l'Université de Bonn lui décerne le titre "doctor juris honoris causa". En mars 1933 Busserl est rayé de la liste des professeurs à cause de son ascendance juive. Cette exclusion est reportée parce qu'il a donné un de ses trois fils à l'Allemagne pendant la grande guerre, mais elle devient effective en 1936, année où il meurt à Fribourg–en–Brisgau. Selon certaines notes biographiques, sa déportation en camp de concentration était, à ce moment–là, immineote. Sa famille, sa bibliotbèque et tous ses manuscrits sont sauvés et sortis d'Allemagne par le franciscain Van Breda qui établit les archives Busserl à Louvain où le matériel est maintenant accessible aux chercheurs du monde entier.

4.5.1 Brentano et Husserl

Brentano a influencé non seulement la vie personnelle de Busserl mais aussi son oeuvre. Néanmoins, sous de nombreux aspects, l'élève a dépassé le maître.

Dans des chapitres précédents j'ai déjà exposé les difficultés que Brentano avait rencontrées pour expliquer de manière satisfaisante la relation entre l'objet intentionnel et le sujet de l'acte intentionnel. Brentano essaie de les résoudre en se bornant à deux notions : sujet et objet. La différence de base avec Husserl est que ce dernier introduit une nouvelle notion, la notion de *noème*.

L'influence de Brentano sur Husserl est susceptible de deux lectures différentes selon qu'on part d'une conception à la Chisholm de Brentano ou d'une conception plus proche de la phénoménologie du Husserl de la dernière époque.

Selon la conception apparentée au courant analytique, Busserl a bâti un tbéorie générale des contenus des états intentionnels qui consiste eo une analyse détaillée de la notion de "renvoi intentionnel" par opposition à Brentano dont le compte–rendu de l'intentionnalité était plutôt centré sur l'objet intentionnel. L'originalité de l'approche busserlienne est qu'elle soutient que, dans tout acte intentionnel, l'attention non seulement s'écarte de l'objet auquel cet acte se réfère mais s'éloigne également de l'expérience psychologique qui a trait à ce dernier pour se concentrer sur l'acte lui même et plus spécifiquement sur son contenu intentionnel. Il y a donc un changement d'attitude dans l'analyse, car elle n'est plus *objective* comme le croit Brentano et comme on l'assume intuitivement, mais devient *réflexive* et repliée sur l'acte même. Ce changement de cible est appelé par Husserl la *réduction transcendentale ou epochè*.

Selon Dagfinn (1932–) Føllesdal [Føllesdal, 1982a, cfr. page 36] grâce à l'introduction de trichromie entre l'acte du aujet, le noème et l'objet, Busserl est en mesure de retenir la notion de renvoi des pbénomènes mentaux tout en surmontant les difficultés posées par l'inexistence de l'objet.

Comment Husserl est–il arrivé à résoudre le problème de renvoi aux objets inexistents? Il serait erroné de penser que l'acte psychique n'est pas dirigé vers un objet à part entière puisque pour Busserl ainsi que pour Brentano, l'objet de l'acte intentionnel est un objet à part entière mais parfois, nous observons (rappelons le cas de centaure) que cet objet n'est pas réel. Néanmoins, l'acte psychique a toujours la notion de renvoi et notre conscience fait *comme s'il existait*. Le noème explique justement ce que signifie dans ce contexte l'expression *comme s'il existait*. L'étude du noème doit viser à mettre en clair la relation entre les différentes caractéristiques de la conscience et expliquer comment elles peuvent s'accomoder des caractéristiques d'un objet. On voit bien que la solution de Husserl paraît plus satisfaisante que celle de Brentano, et qu'elle fait la part des choses entre les structures qui sont dans la conacience et celles du monde extérieur.

L'intentionnalité occupe une place centrale dans la phénoménologie husserlienne mais qu'est-ce qu'un acte intentionnel dans ce cadre? Tous les états mentaux ne sont pas pour Husserl des actes; par exemple: les états qualitatifs comme "avoir mal" ne seront pas considérés comme tels. En revanche, si un état mental est bien un acte alors il est considéré comme étant toujours intentionnel.

La courant plus fidèle à l'approche phénoménologique explique plus aisément les chemins qui ont amené Husserl à cette position philosophique. À l'origine on trouve une préoccupation épistémologique : celle de lutter contre les théories de la connaissance en vogue à l'époque qui menacaient l'objectivité cognitive. Une des théories est le psychologisme qui fait une identification entre le sujet de la connaissance (ou de la science) et le sujet psychologique. Alain Renaut résume très bien cette approche dans la phrase: "nul jugement ne serait indépendant du moi qui l'énonce." [Renaut, 1993, cf. page 80]. L'autre théorie est le logicisme, en fonction de quoi la validité des principes logiques ne se fonderait que sur notre organisation psychique, ce qui revient à dire que la logique ne serait qu'une branche de la psychologie. Ces deux théories amenaient de l'eau au moulin du scepticisme épistémologique.

Le chemin frayé par Husserl passe donc par deux récusations: celle du psychologisme et celle du logicisme. Selon Renaut la première récusation lui permet de faire la part des choses entre le phénomène psychologique et le phénomène pur par la mise en ouvre de la *réduction phénoménologique*. Comment se fait-il que puisse exister une connaissance quelconque à partir d'un vécu intellectuel par lequel quelque chose m'est donné? . Comment peut-on passer de l'"évidence immédiate" à la "une sphère de présence absolue"?

> Le moyen en est, selon Husserl, "la réduction phénoménologique" qui permet de dissocier le phénomène pur du phénomène psychologique: "a tout vécu correspond, sur la voie de la réduction phénoménologique, un phénomène pur, qui révèle son essence immanente (prise individuellement) comme une donnée absolue"[Husserl, 1985, page 69]. C'est ici qu'intervient le concept d'intentionnalité – pour justifier le caractère propre à ce phénomè ne pur qui a la vertu de se rapporter à un être objectif, sans qu'il faille en rien préjuger de l'existence ou la non – existence de cet être.[Besnier, 1993, page 480]

La récusation du logicisme d'autre part permet d'éviter toute acceptation dogmatique des essences transcendantales.[23]

Comme le signale Jean Michel Besnier

> C'est avec la théorie de la réduction phénoménologique que s'opère la rupture de Husserl avec le psychologisme aussi bien qu'avec le logicisme qui en est la tentation inverse: comprendre le monde, ses manifestations, et avec lui, le Moi de l'homme incarné, suppose en effet la mise en suspens de toute relation spontanée de la conscience avec ce que n'est pas elle. Par là même, le sujet transcendantal se démarque clairement du sujet empirique, impliqué dans l'espace et le temps; il apparaît comme le foyer de toute signification, le principe de toute intentionnalité constitutive d'objet. [Besnier, 1993, 484-485]

Lorsqu'on se place dans cette perspective, l'inclusion du noème dans la relation intentionnelle brentanienne n'est pas le première caractéristique à relever.[24]

Selon Alain Renaut les emprunts que fait Husserl à son maître portent d'abord sur la distinction entre les *phénomènes psychiques* et les *phénomènes physiques*, ce que définit l'essence du concept de conscience en termes du *vécu intentionnel*; le terme *vécu* sera adopté par Husserl pour désigner les phénomènes psychiques. Ensuite on retrouve sous sa plume la conviction brentanienne que l'esprit se dirige toujours vers quelque chose, ce qui se traduira par la formule « Toute conscience est conscience de quelque chose » reprise ultérieurement par Jean Paul Sartre.

La relation intentionnelle implique non seulement que l'objet se rapporte au sujet, mais aussi que ce mouvement se vérifie dans le sens inverse.

---

[23] Le psychologisme est la théorie qui considère le contenu théorique des normes logiques comme dérivé de la psychologie empirique. Le logicisme est en revanche la tendance opposée.

La première formulation du logicisme dû à Gottob Frege et à Bertrand Russell soutient que la théorie des nombres est totalement réductible à la logique déductive. Or, les concepts arithmétiques ne sont que des concepts logiques. Une position logiciste extrême affirmerait que la connaissance arithmétique a les mêmes origines, contenus et justification que notre connaissance des vérités logiques.[Bell, 1994, cf. page 265] [Besnier, 1993, cf. page 477]

[24] Pour un étude détaillée des conséquences de la récusation du logicisme et du psychologisme voir [Philipse, 1995]

La première direction du rapport montre néanmoins que les actes intentionnels ne sont pas uniquement une affaire de psychologie. La deuxième direction du rapport signale l'activité que déploie le sujet dans l'appréhension intentionnelle. Ainsi, les deux solutions classiques: le réalisme et l'idéalisme ne peuvent pas être considérées comme caractéristiques des vécus intentionnels. Le réalisme soutient que la chose en soi produit dans le sujet sa représentation, laquelle est alors conçue comme l'effet d'une affectation de nos facultés de connaître (tenue pour essentiellement passives) par le choc de la réalité extérieur. L'idéalisme, en revaoche déclare illusoire l'impression d'extériorité présente dans la conscience d'objet pour réduire la réalité à une simple représentation produite par l'activité du sujet. [Renaut, 1993, cf. page 95]

Le mérite de Brentano a été de dégager les caractéristiques de tout acte intentionnel, et à partir de cette démarche une question fondamentale pour la théorie de la connaissance peut être posée: il s'agit d'éclaircir non comment tel objet est en relation avec tel sujet mais plutôt de définir la relation

> [...] entre l' *objet comme tel* et le *sujet comme tel*, ce sujet de la connaissance en général que Husserl, après Kant, nomme *sujet transcendantal*. [Renaut, 1993, page 98]

Husserl place ainsi sa quête philosophique dans le cadre la corrélation subjective–objective qui est au centre des vécus de la conscience, étant donnée que selon la tradition brentanienne la conscience est toujours conscience de quelque choses.

Résumons: j'ai présenté deux filères de l'héritage brentanien chez Husserl. La première consiste en une une lecture qui vient tout droit de l'interprétation analytique de Husserl et à mon avis, elle ne prend point en considération la direction du rapport intentionnel dans la direction sujet–objet. Le noème est présenté ainsi comme une entité qui parvient à surmonter l'inexistence de l'objet de cet acte, mais le processus d'acquisition n'est pas tenu en compte. L'acte psychique et la représentation conçus comme étant une entité et non comme une activité sont analysés *a posteriori*. De cette façon uniquement les partisans de cette interprétation arrivent à amalgamer (à réduire) le vécu intentionnel avec une relation plus proche des relations mathématiques. Cette démarche permet à Føllesdal d'établir une correspondance presque bijective entre les coocepts fregéens et ceux qui sont issus du la phénoménologie transceodantale d'Husserl. Je reviendrai sur ce point plus tard.

Pour l'autre courant, plus fidèle à l'oeuvre d'Husserl, l'béritage fondamental que Brentano lui a légué est la quête des caractéristiques communes à tous les actes psychiques. Cette quête qui permet à Husserl de se poser la question sur les rapports subjectifs–objectifs et ensuite de proposer la réduction transcendantale pour faire la part de choses dans les vécus intentionnels puis finalement de proposer les structures, comme par exemple le noème, qui y interviennent.

4.5.2 L'interprétation fregéenne de Husserl

Les notions phénoménologiques de base

Selon ce courant idéologique, l'*analyse phénoménologique* se déroule autour des trois éléments ou concepts: le *noème*, la *noèse* et l'*hyle*. Ces trois éléments ne sont pas couverts par la réduction transcendantale et ils sont donc l'objet de l'analyse phénoménologique.

Le *noême* est une structure abstraite qui aurait lieu à nouveau seulement dans le cas très improbable où on aurait le même type d'expérience du même objet placé sous le même point de vue, avec les mêmes attitudes anticipatoires, etc... Selon Hubert Dreyfus, Hussserl a attribué au noème un tâche presque impossible.

> The noema, as conceived by Husserl, is a complex entity that has a difficult – perhaps impossibly difficult – job to perform. It must account for the mind's directedness towards objects. Therefore it must contain three components. One component must pick out a particular object outside the mind, another component must provide a "description" of that object under some aspect, and a third component must add a "description" of the other aspects which the object picked out could exhibit and still be the same object. In short, the noema must "refer", "describe", and "synthesize".[Dreyfus, 1982b, page 7]

La *noèse* est l'acte concret que reflète ce noème. L'*hyle* est un type d'expérience que l'on a typiquement quand nos organes sont affectés. Dans le cas de la perception ou de la mémoire l'hyle a la fonction de restreindre les noèses possibles pour ce type de noème appartenant à l'acte. La noèse et l'hyle sont des processus temporaux ou des expériences alors que ces caractéristiques ne sont pas applicables au noème. Le noème, la noèse et l'hyle ne sont pas l'objet des actes auxquels ils appartiennent. L'objet est toujours un objet dans le sens propre de ce concept.

Pour illustrer tous ces concepts prenons le cas d'une expérience perceptive [Føllesdal, 1982a, cf. page 40]. Dans cette situation, selon Husserl un objet cause l'affectation de nos organes sensoriels. À ce moment la couche qui octroie la signification active une hyle; cette couche est la noèse. La noèse informe l'hyle de façon telle qu'elle éveille en nous un acte qui est renvoyé à l'objet approprié.

Selon l'imagination on peut avoir toutes les noèses que l'on veut tandis que dans le cas de la perception, l'hyle joue le rôle de condition de borne et ainsi il élimine la plupart des noèses auparavant possibles. Néanmoins l'hyle ne réduit pas tous les possibilités à une seule, il reste encore différentes possibilités. Pour un instant donné nous allons avoir une noèse particulière mais lors de la poursuite de notre expérience nous obtenons d'autres byles et il se peut que ces nouveaux hyles ne correspondent pas à la noèse originale. Si tel est le cas, nous avons souffert une méprise perceptive et nous devons alors prendre une autre noèse que soit compatible non seulement avec l'hyle original mais aussi avec le nouveau.

Mais où est le noème dans tous cela? Le noème est la structure abstraite, le noème peut être le méme d'acte en acte mais il est instantié par la noèse qui est, en revanche l'acte mental concret. Dans le cas précédent, on peut dire que toutes les noèses possibles sont des instantiations possibles du même noème. Or la noèse n'est que la contrepartie temporelle du noème abstrait.

Le concept de noème

Le concept de noème est souvent comparé à deux concepts différents: d'un côté le concept de *forme* dans le sens aristotélicien, de l'autre le concept fregéen de *Sinn*.

L'inspiration aristotélicienne Les concepts de *matière*, de *forme* et de *hyle* sont utilisés par Aristote, entre autres choses pour résoudre la paradoxe de Parménide et aussi pour expliquer le changement, un des phénomènes le plus frappants chez les êtres vivants. En effet, il s'agit d'expliquer la dynamique de la vie, celle qui permet les changements sans qu'il y ait perte de notre identité. La question pertinente est la suivante: quelle est la différence entre un changement et une substitution?

L'hylomorphisme fut une des premières positions soutenues par Aristote pour donner une solution au problème du changement. Selon cette approche l'âme des créatures vivantes est mieux comprise *comme* étant la *forme* (du grec *morphē* de leurs corps ou de leur matière *hulē*). En prônant l'hylomorphisme, il soutient une thèse plus générale sur la priorité de l' organisation ou de la structure vis-à-vis des composantes matérielles pour l'explication de la nature et de l'action de ces substances. En effet, Aristote professe que la matière est la base de la continuité, elle est la garantie qu'il s'agit bien d'un changement et non d'une substitution. En effet, en dépit du fait que les choses changent, la matière préserve son identité et il semble que sa *forme* ou son organisation soit la candidate idéale pour permettre d'expliquer cette permanence en dépit des composantes de la matière même. Cependant, il ne faut pas comprendre le terme *forme* dans le sens d'aspect ou en tant qu'équivalent du mot anglais *shape* parce que les choses vivantes peuvent changer leur aspect sans pour autant cesser d'être elles-mêmes. La signification de la forme sera pour Aristote

> [...] the organisation in virtue of which they function in the way characteristic of their kind. [Nussbaum, 1995, page 222]

Une des solutions qu'Aristote propose au paradoxe de Parménide est une solution sémantique [McMullin, 1995, cf.]. Trois termes sont nécessaires et suffisants pour expliquer les changements: le sujet qui change (la matière qui change), un prédicat particulier qui fait défaut (privation) et la possession de ce prédicat (forme). L'analyse sémantique est la suivante: la matière ne doit pas être comprise comme le matériel qui compose le sujet sinon comme ce qui sert de sujet dans la

proposition qui exprime ou décrit le changement. La matière est donc ce qui assure la continuité et cela montre qu'il s'agit bien d'un changement et non d'un remplacement. Or la *matière* et la *forme* sont des aspects corrélatifs de tout changement.

Selon Dagfinn Føllesdal [Føllesdal, 1982a] la trouvaille d'Husserl a été l'aggiornamiento du concept d'hylomorphisme.

> Husserl therefore propounds a kind of hylomorphism. What corresponds to Aristotle's "form" or *morphe*, could in a first approximation, be called the noesis, which informs the hyle [qui selon Føllesdal peut être assimilé au concept de matière]. However, the noema [...] is an even better conterpart to the form; the noesis is simply the temporal counterpart to the abstract noema. [Føllesdal, 1982a, page 41]

Résumons à nouveau: le garantie de l'identité est l'hyle (la matière) qui restreint toutes les noèses possibles à celles seules qui sont pertinentes; cependant elle n'est qu'une contrepartie temporelle de la forme aristotélicienne, la contrepartie plus pertinente est le noème. Cette application aristotélienne du concept de changement à l'acte perceptif a pour but de rendre compte du fait qu'un même objet peut apparaître à notre esprit sous des modes de présentation différents tout en restant le même.

Le noème est la structure qui constitue le cadre des possibles noèses, celles-ci n'étant que des instantiations possibles du noème donné. Le rôle de l'hyle est de réduire l'ensemble de tous le noèmes possibles en devenant, de cette façon, le garant de l'identité de l'objet.

Husserl et Frege

Husserl et Frege ont échangé de la correspondance entre 1891 et 1894. Les commentateurs sont d'accord sur le fait que la distinction entre représentation (*Vorstellung*) et signification ou référence (*Sinn*) faite aussi bien par Frege que par Husserl a été conçue par tous les deux de façon independante.[25]

Cependant, il n'y a pas unanimité lorsqu'il s'agit de déterminer si Frege a ramené Husserl au champ de l'anti-psychologisme; à la différence de Mohandy, Føllesdal soutient que l'influence de Bolzano est responsable de ce retournement mais quoi qu'il en soit, notre objectif est de voir dans quelle mesure les concepts de *Sinn* et de *Bedeitung* de Frege sont assimilables aux concepts de base de la phénoménologie husserlienne.

Dans son livre *Ideen* Husserl affirme

> Le noème n'est rien d'autre qu'une généralisation de l'idée de sens (*Sinn*) à l'ensemble du domaine des actes. [Husserl, 1950, page 239]

Pour les partisans d'une interprétation fregéenne de Husserl comme Dreyfus, Husserl a simplement adapté le vocabulaire de Frege à celui de la phénoménologie. Husserl aurait fait cela dans le but de projeter la notion frégéenne du cadre étroit de la signification linguistique à tous les actes intentionnels. [26]

Selon Dreyfus la traduction que fait Husserl des termes de Frege est la suivante: Il utilise "objet" (*Gegenstand*) à la place de référence (*Bedeutung*), les termes "signification" et "sens" (*Bedeutung* et *Sinn*) deviennent interchangeables avec celui de "sense". L'étude de la fonction de la "signification" (*Bedeutung* ou *Sinn*) vis-à-vis de la connaissance devient centrale dans le quête de généralisation de ce concept à tous les actes psychiques.

---

[25] [Mohanty, 1952, cf. page 51], [Føllesdal, 1982a, cf. page 53].

[26] Selon Føllesdal

> In many of his other works, Husserl expresses similar views. Thus in *Ideen*, Volume I, he says, "Originally, these words ['Beduten' and 'Bedeutung'] related only to the linguistic sphere, that of 'expressing'. It is, however, almost unavoidable and at the same time an important advance to widen the meaning of these words and modify them appropriately, so that they in a certain way are applicable to the whole noetic-noematic sphere: that is to all acts, whether they are intertwined with expressing acts or not" (page 304) And in *Ideen*, I, p. 233, Husserl characterizes the full noema as a "'Sinn'(in the widest sense)." [Føllesdal, 1982b, page 74-75]

Mais y-a-t'il des différences entre les positions de Frege et de Husserl ou s'agit-il seulement de variantes de vocabulaires? Dans un intéressant article Ronald McIntyre [McIntyre, 1982] cite quelques différences.

Premièrement, pour Frege la signification (*Sinn*) est la relation entre une entité lioguistique et son référent tandis que pour Husserl la signification, voire l'"acte de signification", voire le *Sinn* noèmatique est la relation entre l'acte et l'objet vers le quel cet acte est renvoyé. Or la relation intentionnelle husserlienne joue le même rôle que la relation de référence fregéenne car elles sont toutes deux déterminées par la signification.

Deuxièmement, Husserl et Frege défendent la théorie de l'*identification - description* comme source de référence. Cette théorie affirme que le sens (*Sinn*) donne une description, une série des propriétés qui servent pour déterminer le référent de la signification. Selon les termes de Husserl, le *Sinn* noèmatique qui est dirigé vers un objet particulier a un "contenu" consistant en un "prédicat de sens" qui prescrit les propriétés de l'objet vers lequel l'acte se dirige. [McIntyre, 1982, page 222]

Cependant, la théorie de l'*identification - description* n'est pas exempte de critiques.

Premièrement, Putnam dans son texte *The meaning of 'meaning'*[27] a déjà prouvé que cette théorie n'est pas valable lorsqu'il s'agit des noms propres. Les différentes critiques se basent sur les faits suivants: parfois les noms propres utilisés ne sont pas connus par le récepteur du message; comment peut-on caractériser Socrate, comme "un des plus fameux philosophes grecs"? On voit bien que cette description n'est pas exclusive à Socrate, il pourrait aussi bien s'agir de Platon ou d'Aristote.

Le deuxième problème est que très souvent la description d'un nom propre donne une mauvais référent pour l'acte intentionnel. L'exemple de Kripke est "Colomb" et une description le disant "celui que découvert l'Amérique" alors qu'on sait qu'il ne l'était probablement pas.

Le troisième problème est que les descriptions dépendent de facteurs contingents comme l'actualité. Par exemple, "le président de la France" maintenant est François Mitterand mais quand vous lirez cette page la description fera référence à quelqu'un d'autre. Aussi bien Putnam que Kripke maintiennent que les noms propres sont des "dessignateurs rigides", ce qui veut dire qu'ils doivent se référer au même objet dans tous les mondes possibles.

Nous avons déjà vu que l'on rencontre un problème similaire dans le cas des démonstratifs. Selon une approche fregéenne, ainsi que nous l'avons vu dans des chapitres précédents on propose des énoncés de traduction pour résoudre le référent, du type "Je" est la personne qui formule l'éoncé.

Husserl avait relevé le problème de la référence des noms propres et des démonstratifs mais il apporte une solution qui se place hors de la théorie de l'*identification - description*. Dans son livre *Ideen* il fait la distinction entre le "contenu" descriptif d'un *Sinn* noématique et un autre composant du *Sinn* qu'il nomme le "X déterminable" dans le *Sinn*.

> Husserl describes the structure of a noematic *Sinn* (paradigmatically, the *Sinn* of the perception) in *Ideen* §129–131. There he distinguishes two components in the *Sinn*: its "content" and its "determinable X". According to Husserl, the content of a noematic *Sinn* consists of senses of predicates that attribute *properties* the object is intended as having (§130). But a Sinn's X-component is a sense of a different sort: it correlates with the *object* itself, the object that is intended as "bearing" these properties (§131) [...] Whereas the whole noematic *Sinn* (content+X) relates the object under a particular description (viz., the description prescribed by the *Sinn*'s content), Husserl says that the X relates to the object "simpliciter," "in abstraction from all predicates"... The X thus seems to be a "non-descriptive" component of sense in the noematic *Sinn*, a sense that presents an act's object directly, without prescribing properties of the object, and in a way that is largely independent of the particular descriptive content of the *Sinn*. [McIntyre, 1982, page 227]

Husserl soutient dans les *Recherches Logiques* que la relation entre les démonstratifs ou les noms propres et leur référent est "directe"

> ... le nom propre, lui aussi, nomme l'objet «directement». Il ne le vise pas attributivement, comme porteur de tels ou tels caractères, mais sans cette médiation «conceptuelle», comme celui qu'il est «lui-même» et tel que la perception nous le mettrait sous les yeux. (R. L. *VI*, §5, 20)

[27] [Putnam, 1975a]

La relation intentionnelle, en particulier celle qui traduit la signification des démonstratifs et des noms propres sera ainsi directe sans aucun type de médiation conceptuelle. En fait, pour Husserl le lien entre les deux éléments de la relation est fait par l'*intuition.* (*Anschauung*). Husserl affirme dans son ouvrage *Recherches Logiques* :

> En un certain sens, il faut bien dire que l'intuition contribue à la signification de l'énoncé d'une perception: en ce sens précisément que, sans le secours de l'intuition, la signification ne pourrait s'expliciter dans sa relation *déterminée* à l'objectivité visée. (R. L. *VI*, §5, 18)

La critique que McIntyre fait de cet appel à l'intuition est que Husserl trahit en quelque sorte la méthode phénoménologique par l'introduction d'un élément nouveau. La détermination de l'objet de l'acte intentionnel (ou du référent) ne ressort plus seulement du contenu phénoménologique de l'acte, c'est à dire de la signification, ou du *Sinn* noémique qui sont intrinsèques a l'acte même, mais aussi de facteurs "contingents ou extérieurs" qu'un objet plutôt qu'un autre arrive à présenter dans l'environnement immédiat du sujet.

Selon McIntyre, dans ses derniers travaux Husserl a entendu le concept de "contenu phénoménal" de façon telle qu'il y a inclus non seulement le *Sinn* noématique mais aussi les *Sinn* des multiples expériences et croyances appartenant à l'arrière-plan (*background*) [McIntyre, 1982, cf. page 231]. Ceci indiquerait que sa tendance dernière fut d'aller vers un holisme de la signification et l'intentionnalité. Cependant, McIntyre relève et soutient la critique faite par l'existentialisme: le contenu phénoménologique tout seul (aussi étendu soit-il et quelles que soient les choses qui y sont comprises) n'arrive pas et n'arrivera jamais à expliquer la relation réelle de la conscience avec l'environnement où vit le sujet en question.

Selon cette interprétation, l'acte intentionnel se réfère aux objets à part entière et le concept de noéme est traduit en termes de signification et de références. Par exemple, Føllesdal affirme:

> [...] where as Frege held that in contexts like "believes that ..." terms refer not to their ordinary reference but to their ordinary *Sinn*, Husserl held, [...] that acts normally are directed toward *ordinary objects* and not toward Sinne or noemata of such objects. This leads to major differences in their analyses of act contexts.[Føllesdal, 1982b, page 79]

Jusqu'ici je n'ai fait qu'exposer l'interprétation de Husserl selon Frege; j' énoncerai les critiques que l'on peut lui adresser vers la fin de ce chapitre. Notamment, il faudra d'analyser la relation intentionnelle concept – sujet de Frege et sujet – objet de Husserl et de montrer que la validation d'une cette méthode dépend de la simplification de la seconde relation au hénéfice de la première.

4.5.3 L'interprétation *gestaltique* de Husserl

Outre le courant analytique autour de Husserl, il existe une véritable lignée de commentateurs enracinés dans la *Gestaltheorie.* Un des plus importants tenants de ce mouvement est Aron Gurwitsch dont l'influence selon Dreyfus "emerges transformed into a criticism of Husserl in the writings of Merleau-Ponty" [Dreyfus, 1982b, page 98].

A la différence de la première approche que l'on peut caractériser de réaliste, la seconde est d'inspiration *dualiste.* Sans nier la relation entre les concepts husserliens et ceux de Frege, elle en fait une traduction en termes de *Gestaltheorie*, assimilant ainsi le concept de *noème* à celui de présentation perceptive *percept.*

Dans son livre *Théorie du champ de la conscience*, Aron Gurwitsch affirme que le concept de phénoménologie a eu Descartes pour précurseur et il attire notre attention sur le fait que Husserl définit son approche comme une sorte de *néo-cartésianisme* mais en s'empressant de citer Husserl qui dans ses *Méditations cartésiennes* exprime une nuance:

> [...] bien qu'elle se soit vue obligée de rejeter à peu près tout le contenu doctrinal connu du cartésianisme, pour cette raison même qu'elle a donné à certains thèmes cartésiens un développment radical.(Husserl, *Méditations cartésiennes*)[Gurwitsch, 1957, page 133]

Cette affirmation donne en un certain sens une validité à la démarche, étant donné que l'approche Gestaltique a des bases dualistes et se démarque ainsi des interprétations "naturalistes" au

vu desquelles la perception n'est qu'un acte relevant du "domaine mondain". [28]

Gurwitsch fait ainsi la différence entre *Psychologie* et *Phénoménologie*. La première est une "science positive" qui dans "son exploration et son explication de la conscience ... prolonge les sciences physiques et biologiques, et repose en partie sur elles" [Gurwitsch, 1957, page 131] La phénoménologie par contre aborde la conscience d'un point de vue tout à fait different car

> [elle] vise à la clarification et à la justification ultimes aussi bien du savoir théorique et scientifique au sens propre, que de ce savoir préthéorique et pré–scientifique par lequel nous sommes guidés dans notre vie quotidienne, et d'où procède le savoir scientifique et théorique. Le savoir, à chaque niveau, se réalise par des actes de conscience.[...] Puisque le dessein de la phénoménologie est de rendre compte des objets quels qu'ils soient, et de leur statut ontologique, c'est–à–dire de leur existence et de la signification de leur existence, nous nous trouvons renvoyés aux actes par lesquels les objets en question se présentent à nous pour ce qu'ils figurent dans notre vie consciente, dans notre activité pratique, théorique, artistique, etc. Il est clair que la phénoménologie ne peut procéder à la manière des sciences positives, puisque la clarification et la justification des méthodes des science positives et des notions impliquées dans ces méthodes est une de ses tâches. [Gurwitsch, 1957, page 131]

L'inspiration cartésienne de la phénoménologie n'implique pas uniquement le "neo–dualisme" corps–esprit mais aussi le "doute méthodologique".

La phénoménologie a abandonné l'approche de la psychologie positiviste que Gurwitsch appelle l'"hypothèse de la constance". En effet, la psychologie positiviste ne prend pas seulement l'agent comme un système physique auquel adviennent uniquement des événements physiques mais elle prend pour point de départ de la perception l'univers physique. En revanche une approche phénoménologique prend comme point de départ l'univers tel qu'il nous est familier dans la vie quotidienne et non l'univers construit et élaboré par la physique.

La différence entre une approche phénoménologique et celle qu'on emploie dans la réalité de tous les jours est la suivante: dans la vie nous avons de rapports perceptifs, nous agissons, nous raisonnons en relation à un objet de la réalité. A partir de cette relation, nous formons la croyance consistant à affirmer l'existence de cet objet, ainsi "le caractère existentiel des choses et des êtres que nous rencontrons n'est pas chaque fois dégagé, rendu explicite, posé." Cette supposition d'existence implicite est ce que Gurwitsch appelle l'"hypothèse de la constance" et je l'ai présentée sous ses deux formes. La première est l'hypothèse inhérente au domaine de la physique, la deuxième est celle de la vie quotidienne. La stratégie de la phénoménologie consiste à mettre la croyance de l'existence entre parenthèses, à la suspendre. Il s'agit donc, d'appliquer la réduction phénoménologique.

La réduction phénoménologique selon Gurwitsch

La réduction phénoménologique est pour Gurwitsch "indispensable à une clarification philosophique radicale" [Gurwitsch, 1957, page 136]. Il est faux de penser que la croyance à l'existence de l'objet soit

> [...] mise en doute, ou, au lieu d'être considérée comme certaine, (était) estimée comme seulement probable, etc. Tout cela serait modifier la croyance, et non la suspendre. A proprement parler, la réduction phénoménologique ne concerne pas la croyance en l'existence elle–même ni le caractère existentiel offert par le monde de la perception et par les objets appartenant à ce monde. Elle concerne plutôt le rôle que le phénoménologue fait jouer à cette croyance. En ce sens, la réduction phénoménologique peut être considérée comme un artifice de méthode pour fonder une connaissance philosophique radicale et justifiée.... En vérité, le caractère existentiel des choses réelles, bien loin d'être laissé de côté, est explicitement dégagé et soumis, comme les autres caractères qu'offrent les choses réelles, à une réflexion et à une analyse radicale. Bien plus, la clarification ultime de l'existence des choses perçues et du monde de la perception en général, est un des thèmes centraux de la phénoménologie.[Gurwitsch, 1957, page 135–136]

Les interprétations idéalistes extrêmes de Husserl ne proviennent donc d'une tradition foncièrement gestaltique ni d'une interprétation fondamentalement fregéenne.

[28] En effet, une position naturaliste est aussi une position physicaliste. Elle voudrait expliquer les actes intentionnels en utilisant seulement les termes et les notions appartenant au domaine des sciences naturelles. En ceci elle diffère de la Gestaltheorie qui est dualiste et permet des références à des termes autres que purement physiques.

Le noème perceptif de Gurwitsch

De la même façoo que dans un énoncé linguistique oous avons différentes manières de nous référer à un objet comme par exemple: on peut se référer à Napoléon comme étant "le vainqueur d'Austerlitz", ou "l'initiateur du code civil français" ou " le vaincu de Waterloo", les choses dans l'acte perceptif se présentent sous un certain aspect. Aron Gurwitsch établit donc un parallélisme entre le concept de *mode de présentation* des contenus et l'aspect des choses dans les actes perceptifs.

Supposons qu'un sujet regarde une maison, celle-ci est perçue à partir d'un point de vue où le sujet est placé. Elle peut être de face ou sur le côté, de pleio-pied ou en contre-bas, en pleio soleil ou voilée de brouillard, et ainsi de suite. À travers chaque perception individuelle, la chose perçue s'offre sous une face ou un aspect particulier. Chaque perception présente soo objet par voie d'*esquisse*. Mais pour parvenir à une expérience complète de la chose réelle, il faut dépasser la perception isolée et chercher à avoir d'autres perceptions de cette chose. Chacune des ses perceptions, aussi variées soient-elles, est pourtant vécue comme une perception de la même chose.

> Donc la chose perçue, à titre d'existant, ne s'épuise dans aucune de ses apparences ou présentations perceptives individuelles. On trouve ici la différence ... entre la chose perçue elle-même et un noème perceptif individuel qui s'y rapporte.[Gurwitsch, 1957, page167]

Le noème perceptif est ce que, dans un certain sens, ces perceptions multiples ont en commun, ce à l'égard de quoi les perceptioos multiples s'accordent toutes les unes avec les autres. A ce titre, le noème perceptif ne peut évidemment passer pour une composante réelle d'aucun membre de cette multiplicité.[Gurwitsch, 1957, cf. page145]

Le noème perceptif n'appartient ni au domaine des choses réelles objectives ni au domaine psychologique des actes de conscience, le noème perceptif doit être compris comme le sens de la perception au même titre que la sigoification d'un symbole. [Gurwitsch, 1957, cf. page 146-147]

Selon Gurwitsch, Husserl récuse le théorie "réaliste" de la perception et dans son exégèse husserlienne les termes "apparence" (*Erscheinung*) et parfois le terme "image" (*Bild*) devroot être pris comme des synonymes de "noème perceptif" et il continue

> Quand Husserl emploie le terme "apparence", la différence entre cette apparence et la chose même n'est pas une différence entre ce qui est en fait donné dans la perception et une réalité qui se cache derrière, mais plutôt celle qui existe entre une présentation particulière de cette chose, et la totalité de ses aspects possibles.[Gurwitsch, 1957, cf. page152]

Il continue en affirmant que toutes ces apparences formeot un système. Ce système doit son unité à la cohérence de la *Forme* (*Gestalt*).

> *L'organisation interne du perçu* se révèle ainsi être une *unité par cohérence de Forme* : un système de *significations fonctionnelles solidaires et interdépendantes qui, dans leur coexistence équilibrée même*, constituent le noème perceptif en tant qu'un tout. Il n'y a pas de principe unificateur en addition aux matériaux unifiés. L'unité du noème perceptif consiste en ce que ses composantes ne sont ce qu'elles sont que les unes par rapport aux autres ou bien, dans un certain sens, par la 'présence' des unes dans les autres. [Gurwitsch, 1957, cf. page 244]

C'est dans la citatioo précédente qu'on voit que Gurwitsch fait l'amalgame entre le concept de noème de Husserl et sa notion de présentation de la Forme ("percept" ou "perceptual Gestalt"). Cette ambiguïté est présente entre les notions de *perceptual Gestalt* et de *noème* dans l'oeuvre de Gurwitsch. Je crois qu'on peut l'interpréter de la façon suivante: la puissance structurelle et la cohérence de la forme appartiennent à la *Gestalt* tandis que le noème perceptif sera l'eosemble des éléments commun à toutes les structures apparentes de l'objet perçu. En définitive, l'interface entre la conscience et la réalité ressemble à une pièce doot une des faces est la Gestalt. Cette dernière garde le pouvoir de *donner du sens* puisque c'est grâce à elle et à sa cohérence que *l'organisation interne du perçu* est dévoilée. L'autre est le noème qui détient les caractéristiques communes de toutes les apparences. Néanmoins ces apparences n'appartiennent ni au monde réel ni à la conscience du sujet et ce sont des structures d'anticipation de l'organisation de la *Gestalt*.

4.5.4 Husserl réaliste ou idéaliste?

Avant de répondre à cette question il faut d'abord tirer au clair l'aspect de la phénoménologie husserlienne que nous voulons étudier. Je pense qu'il y en a fondamentalement deux: l'aspect épistémologique ou méthodologique et l'aspect ontologique ou métaphysique.

La réponse quant au premier aspect devra être donnée en fonction de la dynamique de la réduction transcendantale tandis que le seconde sera en fonction du produit final que l'on obtient par la réduction phénoménologique.

Jaakko Hiotikka a récemment publié un texte où il analyse, entre autres, l'aspect épistémologique de la réduction transcendantale et affirme qu'il existe un chevauchement de la conscience et de la réalité. Hintikka signale que la réduction transcendantale ne consiste pas seulement à mettre entre parenthèses l'objet d'un acte et en faire du noème l'unique centre de l'attention du phénoménologue.

> Elle met aussi entre parenthèses tout ce qui, dans le noème, ne nous est pas donné dans une expérience immédiate. Elle sépare ce qui est *intentionné* de ce qui est *donné*, et cherche à réduire le premier au second. [Hintikka, 1995, page 44]

Mais si un des éléments donnés est ce chevauchement de la conscience et la réalité, l'autre est ce que l'on fait subir à cette expérience pour la remener au monde pleinement articulé des objets. Hintikka soutient une position foncièrement différente à celle déjà exposée par McIntyre [McIntyre, 1982]. Pour le premier il est inexact de borner la phénoménologie à l'étude des actes et leurs noèmes. La phénoménologie inclut aussi l'étude de la relation entre les noèmes et leurs objets. Cette relation n'est pas une relation inmediate sinon qu'elle est médiatisée par l'intuition. Pour MacIntyre l'inclusion de l'intuition est une trahison à la théorie mais je pense que ce n'est pas à théorie de Husserl sinon à la théorie à la Frege. Lorsqu'on accepte qu'il existe un chevauchement entre la conscience et la réalité la relation concept–sujet fregiènne s'avère moins nuisible d'être assimilée sans plus à la relation sujet–objet de Husserl. Je pense que c'est ici le problème central; les deux actes intentionnels tels qu'ils sont conçus par Husserl ne sont pas assimilables à des actes de la signification de Frege sans trahir Husserl. Je reviendrai d'ailleurs sur ce point.

Je pense que l'interprétation de Hintikka est celle qui reste la plus fidèle au modèle phénoménologique. Or, il découle de cette interprétation que dans la réduction transcendantale ce ne sont pas les concepts qui sont imposés au sujet, mais que ce "chevauchement" assure le concours du sujet et de l'objet. Dans ce sens on peut dire qu'il s'agit d'un idéalisme méthodologique.

Du point de vue ontologique, à mon avis il ne serait point acceptable de dire qu'Husserl est idéaliste pas plus qu'il ne serait juste de le décreter réaliste naïf. Justement la démarche de la réduction transcendantale repose, entre autres choses, sur le fait qu'Husserl n'accepte pas l'hypothèse de la constance. A la différence du réalisme qui ne met en doute l'existence des objets et qui de surcroît considère ces objets comme les causes à eux seuls des représentations, le rôle de la réduction transcendantale est justement d'arriver aux essences en faisant la différeoce entre ce qui est intentionné et ce qui est donné.

Mais, comme Hall le signale on ne peut pas conclure de ceci à la réduction du nombre des objets réels dans l'univers. La réduction phénoménologique n'a rien à voir avec une réduction métaphysique.

> He calls his method a reduction because one of the elements of ordinary experience drops out of the "reduced" experience of the phenomenologist. But this "dropping out" is solely a function of the direction of attention or point of view which the philosopher adopts. The transformation which occurs is on the side of the subjet. Nothing happens to the real world of which we are ordinarily aware, except that we purposely shift our attention away from it and, as a result, ceases to have it as the object of aur experience. The réduction is a reduction in constituents of our experience, not in the contents of reality. Where natural experience involves both a mediating interpretive structure or meaning and a transcendent objet in the real world, the "transcendentally reduced" experience of the phenomenologist involves only interpretative structures or meanings. [Hall, 1982, page 177]

Enfin, nous pensons comme Hall que la question est tout simplement mal posée car toutes deux sont des positions philosophiques et

> His realism is prephilosophical; his 'idealism' nonmetaphysical. [Hall, 1982, page 186]

4.5.5 Comparaison et critique des deux interprétation de Husserl

Hubert Dreyfus [Dreyfus, 1982a] fait une comparaison des deux interprétations de Husserl que je viens d'exposer. Pour lui, l'interprétation de concept de noème de Føllesdal est plus cohérente avec le développement de la théorie de Husserl. En revanche, l'interprétation de Gurwitsch du noème perceptif ne lui semble pas consistante avec le projet de Husserl car ce dernier vise à bâtir une phénoménologie transcendantale.

La critique fondamentale de Dreyfus concerne le fait que Gurwitsch considère le noème perceptif comme un synonyme du terme "apparence" et rassemble les concepts d' *objet tel qu'on s'y réfère (as referred to)* avec celui de *référence à l'objet (reference to)*. Pour Dreyfus, Gurwitsch a mal compris Husserl qui n'a jamais voulu identifier l'apparence avec le noème, étant donné que son but était la généralisation de la signification (*Sinn*) de Frege aux actes de la conscience. Dreyfus dit que probablement ce que Gurwitsch avait dans l'esprit était d'assimiler le concept de noème avec celui de point de vue. Je suis d'accord avec cette interprétation de Dreyfus car Gurwitsch dit :

> Supposons un sujet qui ne change pas de point d'observation, et qui conserve son orientation par rapport à la chose perçue, mais qui alternativement ferme et ouvre les yeux. Il vit alors une suite de perceptions qui toutes diffèrent les unes des autres, par le fait même qu'elles se succèdent dans le temps. Liaison de côté pour l'instant la temporalité intrinsèque, c'est-à-dire la durée phénoménale de chacune des perceptions qui font partie de cette suite. A travers chacune des ces perceptions, c'est non seulement la même chose qui est donnée, mais encore elle est donnée du même côté, sous le même aspect, dans la mê me orientation, etc., bref, dans la même présentation unilatérale. Une multiplicité d'actes de perception correspond ainsi à un seul noème perceptif. Celui-ci ne peut donc être identifié à aucun de ces actes multiples. [Gurwitsch, 1957, page 145]

Dans ce texte Gurwitsch trahit la conception husserlienne puisque le noème ne peut pas être exactement le même; laisser de côté la dimension temporelle ne suffit pas car il faudrait aussi laisser de côté les attitudes anticipatoires, vu que le noème change à chaque instant. Je crois d'ailleurs que Gurwitsch contredit en quelque sorte sa propre idée selon laquelle les différentes expériences perceptives servent à découvrir la forme puisque sans même changer de point de vue, le sujet perceptif peut déplacer son attention en fixant plus un partie qu'une autre, comme s'il était en train de faire en quelque sort un zoom sur certaines parties. [29]

D'une certaine façon Gurwitsch a opéré une supression des capacités attribuées originairement au noème pour les transférer au concept de Gestalt.

En outre, les interprétations à la Føllesdal trahissent aussi l'esprit originel de Husserl. Dans un article paru récemment, François Rivenc[30] élève deux types d'objections. Premièrement il critique les raisons qui poussent certains philosophes à une telle démarche. Deuxièmement il essaie de montrer que le parallélisme présupposé entre les notions centrales de la théorie de Frege et de Husserl ne se vérifie qu'au prix d'une phénoménologie transcendantale *au rabais*.

En référence aux motivations, Rivenc avoue ne pas être partisan de la phénoménologie transcendantale mais néanmoins il attire l'attention sur les intentions sous-jacentes que des interprétations fregiennes peuvent avoir. En effet, ce Husserl "revu et corrigé « à la Frege »" permet à ses yeux de légitimer Husserl.

> Il reste quelque chose de curieux dans cette alchimie frégéenne où l'on fait recuire les noèmes: comme si la théorie frégéenne du Sens était si assurée que ré-interpréter l'intentionnalité dans ce cadre conceptuel faisait *ipso facto* de Husserl un auteur respectable! [Rivenc, 1995, page 14]

Ainsi, la phénoménologie transcendantale retrouverait ses lettres des noblesse parce qu'elle n'est qu'une variante intentionnelle de la théorie du sens de Frege. Cette légalisation fait de la philosophie de Husserl un des points d'ancrage pour une "théorie représentationnelle (computationnelle ou émergentiste) de l'esprit".

Le deuxième aspect des critique de Rivenc vise deux aspects différents mais néanmoins connexes. L'analyse intentionnelle que l'on peut développer à partir de concepts fregéens et le réalisme que

[29] Ce qui équivaut à orienter la fovéa sur différents points de l'objet perçu.
[30] [Rivenc, 1995]

cet auteur présuppose constituent selon lui deux des caractéristiques fondamentales qui ne correspondent pas à la phénoménologie transcendantale. Une thèse qui tenterait de tenir compte des deux aspects serait la suivante:

> [...] de même que le noème est une entité abstraite, non perceptible, une « structure» articulée qui opère la médiation entre le sujet et l'objet, et permet de rendre cnmpte de «l'être–dirigé–sur» des représentations, de même le Sens frégéen est une entité intermédiaire (le monde de donation de l'objet, ou ce en quoi le monde de donation est contenu), grâce auquel des objets nous sont présentés. La logique, théorie du Sens, serait donc médiatement théorie des objectivités visées à travers la signification. Ce faisant, on fera simplement jouer aux concepts d'intentioonalité et de noème le rôle qu'on leur donne de manière générale en philosophie de l'esprit contemporaine: rôle de trait d'union entre la conscience et le monde. [...] Bien que le vocabulaire de Searle soit plus brutal (plus naturaliste), la vision de l'intentionnalité selon Searle est profondément analogue: il s'agit de cette propriété qu'ont les représentations « de mettre l'organisme en rapport avec le monde », et le trait commun de toutes ces interprétations est de concevoir l'intentionnalité comme un certain type de relation: une relation peut–être mystérieuse, mais une relation quand même! [Rivenc, 1995, page 21]

C'est cette conception qui doit être contestée. Pour Husserl l'intentionnalité ne pouvait pas être une relation; toute sa démarche consiste à surpasser l'idée naïve de rapport au réel entre le sujet et l'objet.

> La thèse du caractère non relationnel de l'intentionnalité est selon Husserl la clef de toute compréhension ultérieure de la construction de l'objectivité dans la vie subjective. [Rivenc, 1995, page 24]

Et il ajoute

> Que l'intentionnalité ne soit pas une relation au sens où un siècle de culture logico–mathématique nous a habitués à comprendre ce mot, cela découle de la position même du problème que ce concept a pour vocation de résoudre: la constitution dans la vie du sujet de l'idéalité de toute objectivité. Qu'il ait là une énigme pour qui reste dans la position naïve du réalisme scientifique comme philosophie, Husserl en était tout à fait conscient. L'analyse des structures noèmatique doit cependant, selon lui, nous permettre d'y voir plus elair dans ces problèmes de constitution. [Rivenc, 1995, page 24]

L'affirmation que "le noème est généralisation de la notion du Sens" que j'ai déjà exposée plus haut est le fruit d'une interprétation partielle de l'oeuvre d' Husserl. Rivenc démontre que cette idée pouvait être pertinente dans le Husserl de l'époque de la *Première Recherche logique* mais qu'elle fut abandonnée dans les *Cinquième Recherche logique*. Dès lors il ne s'agit pas d'une simple généralisation en *extension* d'un concept strictement déterminé comme celui de Sens fregéen mais de l'enrichissement d'une catégorie auquel le Sens appartiendrait. Le concept de noème surpasse et enrichit celui du Sens.

> Les noèmes ne sont pas des Sens, si l'on entend par là une formule qui ramène l'inconnu au connu; les noèmes sont des Sens, si on se laisse guider par une analogie « scientifiquement » féconde censée enrichir la notion originelle. [Rivenc, 1995, page 27]

En effet, il est important de prendre en compte les trois étapes que l'on reconnaît dans l'œuvre d'Husserl. Au début de sa carrière il soutenait une position favorable au psychologisme dans la fondation de l'arithmétique. La deuxième étape est caractérisé par le tournant qu'il prend dans ses recherches. En se fondant sur la psychologie définie par Hrentano, Husserl abandonne son penchant pour le psychologisme pour adopter une nouvelle discipline, la *phénoménologie* et il soutient une position métaphysique que l'on a décrite sous le nom d'*idéalisme transcendantal*. C'est alors qu'on peut établir le parallélisme le plus étroit entre les concepts fregéens et ceux d'Husserl. Cependant, il y aura une troisième période durant laquelle il transformera la phénoménologie de la période précédente qui était basée sur une espèce de solipsisme méthodologique au sein d'une phénoménologie de l'intersubjectivité et finalement (spécialement dans son œuvre *Crisis*) sur une ontologie du monde vivant, y compris les aspects sociaux de la culture et de l'histoire [Smith and Smith, 1995, page 1]. Finalement, il est étonnant que le prestige et l'importance de Husserl dans la philosophie se trouvent mieux préservés par des philosophes de tradition analytique qui ne sont pas enrôlés dans une tradition phénoménologique que par d'autres qui malgré leurs tendances husserliennes essaient de concilier les principes du réalisme et les apports de la phénoménologie. En ce faisant, ils

réduisent la réduction transcendantale à n'être qu'un moyen au service du réalisme, une fonction ancillaire dont mieux vaut se passer si l'on adbère au réalisme puisqu'on vide ainsi la méthode husserlienne de sa substance.

Je conviens néanmoins qu'il faut saluer toute entreprise ayant pour objectif le développement d'un plan de recberche basé sur un phénoménologie minimalisée. Cette stratégie consisterait en l'exploration de la construction de l'objectivité à partir des données sensorielles comme le fit par exemple Bertrand Russell au cours de sa trajectoire pbilosopbique pour citer un exemple dans le cadre philosophique. Il existerait d'autres parallèles en psychologie comme par exemple la *Gestaltheorie*. Actuellement Barry Smitb poursuit ses recberches dans ce cadre car un de ses buts est de fonder une ontologie formelle des objets tels qu'ils qu'ils sont perçus par le sens commun et je reviendrai sur ses travaux dans le chapitre que je dédie aux modèles de morphogénèse.

4.6 Conclusion

Nous avons commencé ce chapitre par une étude des concepts de représentation et de contenu chez Brentano. J'ai montré que ces concepts ont été annexés par les sciences cognitives représentationnelles dans le cadre d'une interprétation linguistique de l'intentionnalité; interprétation qui selon certains, dont Alain Renaut et Jean Pierre Dupuy, vient tout droit d'une mauvaise lecture de Brentano. Quoiqu'il en soit, il est bien clair que dans les sciences cognitives traditionnelles, le coocept de contenu est devenu l'équivalent du sens dans la philosophie du langage. Le concept de représentation est solidaire du concept du symbole. C'est encore l'béritage de la tradition analytique qui ne permet pas de prendre la totalité du sens du concept de représentation dans l'oeuvre de Brentano. Ainsi, le rôle du sujet dans l'acte intentionnel est aboli. D'ailleurs, on ne parle que très peu dans cette tradition des actes intentionnels et l'on préfère l'expression "attitudes intentionnelles". Tout aspect dynamique du sujet dans des actes iotentionoels n'est guère évoque.

Les objets s'imposent aux sujets. Les sciences cognitives traditionnelles ont pratiqué un réalisme qui souvent n'est pas extrême mais en revanche est facilement passible de naïveté. Néanmoins, les efforts de Drestke pour expliquer la genèse du lien sémantique méritent d'être pris en compte au titre d'efforts pour surmonter ces critiques.

J'ai exposé les problèmes que cette position engendre; il est par exemple très difficile de faire la part des choses entre les aspects purement psychologiques et ceux qui viennent de l'extérieur. Rappelons nous, par exemple de la difficulté pour individualiser les contenus lorsqu'il s'agit des contenus étroits mais aussi les difficulrés pour expliquer la méprise représentationnelle dans le cadre des théories informationnelles.

Ensuite, nous avons présenté deux interprétations de Husserl. Celle qui ressort d'une tradition analytique nous présente un Husserl au rabais, un Husserl qui a besoin de s'aligner sur Frege pour recevoir ses lettres de noblesse. Mais cette opération ne s'avère pas gratuite : la relation caractéristique du *vécu intentionnel* est réduite à une relation mathématique de type sujet–proposition, elle n'est pas dynamique et l'explication que l'on donne des représentation est une explication après–coup.

L'autre tradition, celle qui récuse l'interprétation linguistique de Frege pour mettre en avant la dynamique subjective–objective, celle qui montre que le vécu intentionnel n'est pas une simple juxtaposition objet–représentation interne mais un cheminement inteotionnel entre le sujet et l'objet intentionnel s'avère, à notre avis plus fidèle à la tradition husserlienne.

Cependant, le fait d'être plus fidèle ne la rend pour autant, plus apte à cautionner une démarcbe naturaliste de l'intentionnalité, bien au contraire. Cette conception semble n'exister que pour nous rappeler que lorsqu'on y souscrit, le sujet devient une composante inévitable du vécu intentionnel dont il est le préalable non réductible au niveau physique. Or, si l'on embrasse cette conception fidèle à Husserl, la tbèse de Brentano se voit renforcée dans sa validité pour devenir presque incontournable.

Finalement et dans le cadre général du présent travail, l'exposition des concepts de contenu large, contenu étroit et contenu informationnel est nécessaire à la compréhension des critiques ultérieures que je vais adresser à la théorie representationnelle de Fodor. Cette théorie qui prône

l'existence d'une corrélation entre les propriétés computationnelles et les propriétés sémantiques des contenus a échoué dans son but, étant donné que cette corrélation ne s'avère pas métaphysiquement nécessaire. D'ailleurs, je vise à démontrer que l'hypothèse de la multiréalisation de la cognition, qui est un des arguments centraux invoqués contre la réduction du mental au physique, n'est plus qu'une possibilité contingente dans la théorie fodorienne.

Je vais soutenir dans ma thèse qu'un des problèmes de cette approche est l'ambiguïté du concept de *fonction* qui bien qu' ontologiquement neutre semble très vague pour constituer la corrélation digne de confiance des contenus mentaux comme multiréalisés en eux d'un côté, et les états physiques de l'autre. Dans le but d'éviter le physicalisme réductionniste de type, on a pensé que l'on pouvait bâtir une théorie en faisant plus référence à l'intelligence artificielle qu'aux données des neurosciences.

Il est nécessaire de trouver un autre cadre pour pouvoir décrire la dépendance des états mentaux des états physiques. Je vais proposer une voie de solution dans le modèles de morphodynamique.

En outre, l'exposition de l'usage que l'on fait des concepts de Husserl est nécessaire pour comprendre les critiques que je vais adresser à la théorie phénoménologique et émergentiste du Professeur Jean Petitot.

Ce chapitre clôt la première partie dont le but était l'exposition du problème de la naturalisation de l'intentionnalité et de la conciliation entre les termes du trilemme classique du problème corps–esprit.

On a vu qu'il est difficile de concilier une position physicaliste avec la pertinence causale du mental.

J'ai choisi pour exposer des concepts nécessaires à la suite de mon travail une perspective historique. Cette perspective, même au risque de sembler trop descriptive, a aidé ma propre compréhension du problème et m'a permis de rendre mon exposé différent d'un simple catalogue des concepts qu'on est de toute façon obligé de présenter lorsqu'on se place dans un perspective interdisciplinaire. Néanmoins, en ce faisant j'espère avoir mis l'accent sur quelques caractéristiques et quelques passages intéressants du chemin ou plutôt des *routes* et *déroutes*[31] de l'étude de l'intentionnalité.

[31] Je me permet d'utiliser ces termes empruntés à l'ouvre de [Cayla, 1991]

Partie II

Les modèles des sciences cognitives traditionnelles

Chapitre 5

Le monisme anomal de Donald Davidson

Le cœur à des raisons que la raison ne connait pas.
Vauvenargues

5.1 Introduction

Les explications que nous donnons pour notre propre comportement et pour celui d'autrui dans la vie de tous les jours sont fondées sur des inférences qui ont comme prémisses des attitudes propositionnelles. Le premier cas ne présente pas de problème particulier parce qu'il s'agit des compte–rendus à la première personne. Le second cas est plus controversé parce que nous assignons aux autres des croyances et des désirs. Les uns comme les autres résultent des inférences que nous faisons à partir de leurs dires et de leurs comportements en supposant que le prochain est un agent "rationnel", ce qui veut dire qu'une des caractéristiques de tout système cognitif est la cohérence entre ses différents états mentaux et leurs teneurs. Seulement de cette manière le système peut–il garantir la condition de rationalité, ce qui veut dire être raisonnable, bien–fondé et ne pas être sujet aux critiques d'ordre épistémique.

Cet appel à la rationalité donne à ce type d'explication une valeur normative au pire, nomologique dans un ces plus optimiste. L'affirmation qu'elles constituent aussi les causes des comportements, est plus délicate. On admet en général bien volontiers que les attitudes propositionnelles peuvent jouer le rôle des *raisons* dans l'explication d'une conduite, cependant il est moins évident qu'elles jouent un rôle causal.

Lorsqu'on se place dans une perspective physicaliste on ne peut pas accepter que des énoncés intentionnels telles que les croyances, les désirs entre autres puissent être pris pour des causes par elles mêmes, puisque le causes ne peuvent être que des événements physiques. Cependant outre la nécessité de causalité physique il existe un autre problème: les généralisations que l'on peut faire à partir de ces énoncés ne peuvent pas non plus être considérées comme des lois strictes étant donné qu'elles peuvent souffrir des exceptions ou des violations. Elles sont seulement des règles ou au mieux des lois dites "certeris paribus", dans le sens où elles s'avèrent soumises à un ensemble de faits contextuels si divers qu'elles sont rendues pratiquement inertes pour les prédictions.

Considérer que les raisons sont des causes entraîne un long débat. Nous avons déjà parlé de la position éliminativiste des Churchland qui dénient toute pertinence même normative aux énoncés de la psychologie naïve. Néanmoins la mise en évidence d'une relation qui permette aux attitudes propositionnelles de participer d'une façon ou d'une autre au processus causal du comportement est le défi que doivent relever les non–éliminativistes.

Ludwig Wittgenstein et puis G.E.M. Anscombe, parmi d'autres ne considéraient point que les

raisons fussent des causes. Deux des argumentations employées pour démentir l'identité cause–raison sont les suivantes. Premièrement, la version de Hume de la causalité qui se base sur deux principes:

- **la conjonction constante:** *si C, alors (et seulement alors) toujours E* où E et C sont deux classes de termes. Ceci n'affirme pas une connexion du point de vue de la genèse de la causalité mais seulement une *coïncidence invariable.*

- La succession régulière des événements que l'on considère successifs.

Les deux termes de la conjonction doivent non seulement être *isolés et séparés* mais aussi *associés non connectés* selon le même de Hume.[1] C'est précisément cette dernière condition qui semble de ne pas se vérifier. Cette objection est connue par le nom d' *argument de la connexion logique*. L'argumentation est la suivante: étant donné que les raisons (les croyances et les désirs) présents dans l'antécédent de l'action sont considérés précisément comme leurs raisons, les deux termes de la conjonction humienne entretiennent des relations logiques ou du moins répondent à un critère de raisonnabilité. Or elles ne peuvent pas être considérées comme isolées ou non connectées puisqu'elles jouissent d'une connexion logique ou rationnelle.

La seconde argumentation se réfère au caractère obligatoire des occurrences, à chacune des occurrences de C suit nécessairement E et les relations rationnelles ne peuvent pas garantir une telle régularité sans exception. En effet, l'existence d'une raison pour croire certaines choses ou agir d'une certaine façon ne rend pas pour autant, nécessairement raisonnable l'agent à qui on l'attribue. Il se peut que l'agent ait des désirs que nous ne connaissons pas ou qu'il décide d'appliquer d'autres critères comme d'agir à la lumière d'un autre type de jugement.

Voilà les motivations qui amènent certains auteurs à considérer les *causes* comme différentes et non réductibles aux *raisons*.

Donald Davidson et Jerry Fodor parmi d'autres ont soutenu une position contraire à l'éliminativisme des raisons dans l'explication psychologique. Tous les deux affirment que les raisons sont à un certain degré[2] des causes et ils réhabilitent les attitudes propositionnelles de la psychologie naïve comme une proto–science dont il faut tenir compte dans le but de construire une vraie théorie psychologique. J'exposerai ici la position de Donald Davidson tandis que celle de Fodor sera expliquée dans les chapitres suivants.

Mais avant d'entrer dans le vif du sujet sur la relation entre cause et raison chez Davidson, je voudrais exposer le problème de l'attribution des attitudes à autrui et du rôle que joue la rationalité dans une telle démarche. Davidson et Quine en sont deux des principaux théoriciens.

5.2 Le programme de Davidson

Selon Donald Davidson le problème des attributions des attitudes propositionnelles aux autres sujets va la main dans la main avec celui de l'interprétation du langage [3]. J'exposerai l'approche de Davidson quant à la stratégie de l'interprétation dans la philosophie du langage pour ensuite en tirer sa généralisation à la philosophie de l'esprit. Davidson est fortement influencé par Quine qui fut son maître. J'ai déjà exposé le principe de la double norme 3.4 proposé par Quine comme une manière de sauver le mental au moins du point de vue normatif. Il s'agissait de se rallier à l'irréductibilité du mental selon Brentano mais en conservant un éliminativisme utilitariste. L'application de la stratégie de Quine est illustrée par sa thèse sur l'*indétermination de la traduction* qui implique pour lui l'indétermination intentionnelle. Je vais exposer tout d'abord cette thèse puis je décrirai très sommairement la théorie de la vérité de Tarski afin de réunir les arguments nécessaires à la démonstration de la solution proposée par Davidson.

[1] [Bunge, 1959, cf. page 56 de la traduction espagnole]
[2] L'expression "à un certain degré" sera précisée plus tard.
[3] cf. par exemple [Davidson, 1984a, page 154] et [Davidson, 1984c, page 163]

5.2.1 L'indétermination de la traduction de Quine

L'éliminativisme utilitariste du mental de Quine maintient que les concepts de la psychologie de l'esprit n'ont pas une valeur descriptive et s'ils ont une valeur quelconque elle ne peut qu'être instrumentale. Justement, ce manque de valeur descriptive des termes de la psychologie ordinaire découle de l'indétermination de leur traduction.

Les arguments en faveur de cette indétermination pour Quine sont les suivants. Premièrement, Quine est un des philosophes contemporains qui a le plus critiqué l'utilisation de la notion de "proposition" dans la philosophie de la logique et dans la philosophie du langage.

En tant que défenseur des *ontologies austères*, c'est-à-dire peu peuplées, il dénie toute réalité ontologique aux entités que l'on ne peut individualiser. Ainsi, il met en doute le réalisme ontologique des propositions conçues comme des pensées fregéennes, comme des intentions. Selon lui, on se heurte chaque fois à l'impossibilité d'établir l'identité entre deux propositions.

> De ce point de vue, les propositions sont bien, comme le soutient Quine "des créatures de l'ombre". C'est une autre manière de dire qu'il est difficile d'entretenir une notion unique de proposition, satisfaisant tous les usages plus ou moins préthéoriques de la notion.[Engel, 1989, page 25]

L'indétermination de la traduction et son scepticisme vis-à-vis de l'existence ontologique de la notion de proposition le font récuser toute tentative de définir la notion de proposition à partir de sa signification.

Quine a donné les raisons de ce verdict en développant deux types d'arguments qui mettent en évidence l'indétermination de la traduction. Un de ces arguments, dit "par en bas" invoque l'expérience de la pensée analysée dans le deuxième chapitre de *Word and Objets* [Quine, 1960] qui vise à démontrer l'insrutabilité de la référence aux termes. L'autre argument, dit "par en haut", tend à démontrer l'indétermination de la signification des phrases et le holisme auquel toute interprétation est fatalement soumise.

L'argument "par en haut" reconnaît deux prémisses. D'abord, le premier dogme de l'empirisme dont Quine emprunte la formulation à Pierce[4]:

> "La signification d'un énoncé consiste en la différence que sa vérité introduirait avec l'expérience possible"

La seconde prémisse est le principe holistique que Quine dit avoir emprunté à Pierre Duhem (1861-1916). Selon ce dernier principe, la signification empirique d'un énoncé n'est jamais déterminable prise isolément, il faut faire cette détermination en relation à l'ensemble d'énoncés d'une théorie, voire de la science entière. Comme l'explique Pascal Engel:

> Les deux prémisses mises ensemble conduisent à l'indétermination de la traduction: Pierce + Duhem =l'indétermination [...]. Parce qu'il n'y a pas de manière possible, en raison du holisme épistémologique admis par Quine, de dire en quoi telle phrase exprime telle ou telle croyance, et donc revêt telle ou telle signification, on peut conclure que la signification est indéterminée. Il s'ensuit que la notion de Proposition, entendue au sens de la signification d'une phrase, ou d'identité des significations de deux phrases, ne repose sur aucune base objective. Signification, traduction, synonymie et Propositions sont des notions qu'on ne peut définir individuellement sans recourir aux autres, et qui sont indéterminées. [Engel, 1989, page 26]

Or, l'argumentation "par en haut" met en évidence le caractère holistique auquel est soumis toute interprétation, d'où la difficulté de leur individualisation qui justifie les doutes sur leur ontologie. L'argumentation "par en bas" repose sur l'expérience de la pensée de la traduction radicale. Il s'agit d'imaginer un ethnologue-linguiste dont la tâche consiste à traduire en sa propre langue l'idiome que parlent les indigènes de la tribu qu'il étudie et qui lui est totalement inconnu.

Son but est de construire un dictionnaire. Il est clair que cet ethnologue, au stade initial de son travail ne peut se servir que des réactions des indigènes face aux différents stimuli sensoriels pour percer le secret de leur vocabulaire. Dans un deuxième temps, il pourra essayer des phrases

[4] Quine avait émis en 1953 dans son célèbre article "Two dogmas of empiricism" [Quine, 1963] des critiques sur ce dogme. Il réfute la thèse de l'empirisme logique selon laquelle la signification cognitive d'un énoncé est totalement déterminée par ses conditions de vérification en argumentant que cette signification ne dépend pas seulement de l'énoncé pris de façon isolée. Au contraire, il défend une version holistique de la signification.

dans le langage indigène en présence des stimuli auxquels elles semblaient répondre et tirer des conclusions à partir de l'approbation ou de la désapprobation de ses informateurs indigènes. Ainsi le couplage stimulus–signification est l'unique donnée que notre ethnographe possède pour vérifier la correspondance entre le langage inconnu et le sien. Dès lors on doit renoncer à la traduction de toutes les phrases qui ne se rapportent pas aux stimuli extérieurs, ce qui veut dire qu'on doit renoncer à traduire des phrases pour lesquelles le couplage stimulus–signification ne se vérifie pas. Des phrase du type "Il est célibataire" ou "Deux plus deux font quatre" ne sont pas traduisibles dans ce cadre. Cependant, dans le répertoire de phrases qui vérifient le lien stimulus–signification on rencontre également des problèmes. Supposons que son informateur indigène en signalant un lapin dise "garavaï"; il y a de fortes chances que notre ethnologue traduise cela comme "Voilà un lapin" mais on ne saurait affirmer que "garavaï" signifie lapin. Cela peut signifier une partie intégrante du lapin ou bien au contraire, la désignation universelle pour toute la classe des lapins; disons par exemple qu'il pourrait s'agir d'une expression se référant à la *lapinité*. Aucune méthode ne permet de déterminer laquelle de toutes ces interprétations de "garavaï" sera la bonne. L'absence d'un quelconque procédé pour vérifier la traduction de "garavaï" par "lapin" met en évidence l'inscrutabilité de la référence des termes. Cela démontre aussi la difficulté de trouver uniquement à partir des phrases, la référence des termes qui les composent et ce en absence de toute autre théorie.

La décision de traduire "garavaï" par "lapin" est donc une décision analytique et non empirique et il en va de même pour les autres termes du lexique. On voit qu'il y a une série de traductions possibles et qu'elles sont toutes compatibles avec le couplage stimulus–signification, seule condition nécessaire à la prise de décision analytique. Il s'ensuit qu'on ne construira pas un dictionnaire unique mais une série de dictionnaires incompatibles entre eux.

> Étant donné un manuel de traduction M pour le langage L conforme aux données fournies par le comportement des locuteurs de L, il est toujours possible de construire un autre manuel de traduction M', conforme aux mêmes données de comportement, mais incompatible avec M. En d'autres termes, il n'y a pas de schéma unique de traduction pour une phrase donnée de L, relativement à des données empiriques observables (liées au comportement) données. [Engel, 1989, page 25]

Quine va plus loin dans son hypothèse de l'indétermination de la traduction. Selon lui, celle–ci ne se vérifie pas seulement dans le cas des traductions d'une langue à une autre préalablement inconnue, cet exemple étant tout simplement un cas extrême dont il se sert pour mettre en évidence la pertinence de son hypothèse. Cette situation se vérifie aussi dans le cas des locuteurs d'une même langue. Prenons le cas d'un phrase dont la signification n'est pas directement observable, comme par exemple "il est célibataire", Quine dit:

> The stimulus meaning of a very unobservational occasion sentence for a speaker is a product of two factors, a fairly standard set of sentence–to–sentence connections and a random personal history; hence the largely random character of the stimulus meaning from speaker to speaker.
>
> Now this random character has the effet not only that the stimulus meaning of the sentence for one speaker will differ from the stimulus meaning of *that* sentence for other speakers. It will differ from the stimulus meaning also of any other discoverable sentence for other speakers, in the same language or any other. Granted, a great complex English sentence can be imagined whose stimulus meaning for one man matches, but by sheer exhaustion of cases, another man's stimulus meaning of 'Bachelor'; but such a sentence would never be spotted, because nobody's stimulus meaning of 'Bachelor' would ever be suitably inventoried to begin with.[Quine, 1960, page 45]

On voit que selon Quine, les locuteurs d'une même langue ont recours à des hypothèses analytiques semblables à celles dont use notre ethnologue pour construire son lexique entre les locuteurs d'une même langue. Or de même que l'on ne peut pas affirmer qu'il existe un stock de significations indépendantes propres à servir à la traduction d'un langage à un autre, on doit aussi rejeter l'affirmation qu'une telle situation ait cours au sein d'un même langage.

Néanmoins, le fait que la traduction soit dans tous les cas indéterminée n'exclut pas qu'il y ait des principes normatifs pour parvenir à une traduction sinon unique du moins correcte. Notre ethnologue doit faire la supposition préalable que les indigènes sont des individus rationnels s'il espère réaliser un dictionnaire. Une telle supposition joue le rôle de principe normatif.

Ce principe est connu sous le nom de *principe de charité* et consiste à prendre les attitudes suivantes: d'abord, l'ethnologue doit interpréter les croyances des indigènes de façon à en maximiser

l'accord avec les siennes, ensuite, il doit supposer que l'ensemble des croyances de ses informateurs est cohérent et ne comporte pas de contradiction interne et finalement ses traductions doivent préserver les lois de la logique classique. Il doit de ce fait considérer suspecte toute traduction selon laquelle on attribuerait des croyances absurdes ou illogiques aux indigènes. Ce que Quine [Quine, 1960, cf.] applique à la traduction d'un langage à un autre est aussi valable pour l'attribution des croyances et des contenus intentionnels en général.

> [...] puisqu'assigner une signification aux phrases émises par un locuteur c'est aussi lui assigner certaines croyances et autres états mentaux, il s'ensuit que les contenus mentaux et intentionnels sont tout aussi indéterminés que peuvent l'être les contenus sémantiques.[Engel, 1994b]

Une des données de l'héritage quinnien en référence à l'indétermination de la traduction est que l'on peut trouver une traduction correcte en appliquant le principe de charité qui est un principe normatif et imposé par le holisme auquel toute individualisation des contenus, soient-ils les contenus des phrases d'une langue ou les contenus des croyances, est soumise. Davidson propose une théorie de l'*interprétation radicale* en s'inspirant de Quine. Le fait que la théorie de l'interprétation soit tirée de celle de Quine ne doit pas nous faire croire qu'elles sont identiques et c'est pourquoi nous disons "interprétation" et non "traduction". Davidson même nous met en garde contre cet amalgame dans une note en bas de page de son texte "Radical Interpretation":

> The term 'radical interpretation' is meant to suggest strong kinship with Quine's 'radical translation'. Kinship is not identity, however, and 'interpretation' in place of 'translation' marks one of the differences: a greater emphasis on the explicitly semantical in the former.[Davidson, 1984d, note au pie de page 126]

Jusqu'ici j'ai exposé la position de Quine et le principe de charité; la section suivante sera consacrée à la théorie de l'*interprétation radicale* de Davidson.

5.2.2 La théorie de l'interprétation radicale

De même que Quine, Davidson soutient le caractère holistique de toute interprétation, ce qui veut dire qu'elle doit être faite dans le cadre d'une théorie. Nous allons montrer que ceci signifie que pour lui la valeur de vérité des phrases n'est pas déterminée de façon atomiste mais ne peut s'évaluer au contraire que dans le contexte d'une théorie faisant appel aux autres éléments de la structure.

La théorie de Davidson a deux antécédents: la théorie de la vérité de Tarski et la théorie du sens de Frege.

Les antécédents

En 1933, Alfred Tarski formule sa théorie de la vérité pour les langages formels quantificationnels. [5]

Ce qui nous intéresse ici est de voir jusqu'à quel point la théorie de la vérité de Tarski dédiée aux langages formels s'avèrera applicable à la signification des langages naturels.

Je vais esquisser brièvement la théorie de la vérité de Tarski de manière à mettre en lumière certains concepts nécessaires à la compréhension de la démarche de Davidson.

Soit le langage L et sa syntaxe qui sera définie inductivement par ordre de complexité croissante des expressions de la façon suivante: aux expressions de la logique propositionnelle on ajoute celles des prédicats et des quantificateurs. On aura donc des quantificateurs (universels et existentiels), des variables d'individu ("x", "y", "z" , ...), des constantes d'individu ("a", "b", "c" , ...), des noms de prédicats ("P", "Q", "R" , ...) et des clauses du type:

> Si Px est une phrase et si Qx est une phrase, alors '$P^\frown x^\frown \wedge^\frown Q^\frown x$' est une phrase (où '$^\frown$' est le signe de concaténation qui indique le résultat de la concaténation d'une expression à une autre).

[5] Dans le texte *The concept of Truth in Formalized Languages* paru en polonais en 1933, la traduction allemande en 1936 et la version anglaise en 1956. Il y énonce la Convention C que je vais traiter plus loin et c'est dans *The semantic conception of truth* qu'en 1944 il propose la phrase-T (*T-sentences*).

La théorie de la vérité aura la forme d'une liste de clauses définissant pour chaque expression primitive du langage ses conditions de vérité. Pour satisfaire ce que Tarski a appelé la "condition d'adéquation matérielle" chaque expression devra avoir une forme semblable à l'expression célèbre:

L'expression "La neige est blanche" est vraie si et seulement si la neige est blanche.

La phrase située à gauche de la biconditionnelle est une phrase du langage–objet et la phrase située à droite est une phrase du métalangage auquel appartiennent les énoncés de la théorie de la vérité. Si la phrase à droite est identique à la phrase du langage–objet (comme ci–dessus) alors la théorie est dite *homophonique*, en revanche si c'est une traduction dans le métalangage, la théorie sera dite *hétérophonique*.

La définition de vérité pour le langage L a comme base la notion de *satisfaction* et la vérité sera ainsi un cas–limite de satisfaction.[6] La vérité est définie *récursivement* à partir des conditions de satisfaction des expressions primitives du langage. La stratégie récursive consiste à procéder selon un ordre de complexité croissante : les définitions sont données d'abord pour les expressions primitives, ensuite pour leur composition par emboîtements successifs.

Finalement, la démonstration de toute instance selon la théorie de la vérité doit suivre le schéma métalinguistique du type:

(T) 'S' est vraie si S $(\star)$

où (T) n'est que la "convention T" d'adéquation de toute théorie de la vérité selon Tarski. Dans $(\star)$ 'S' est une désignation métalinguistique de S et cette dernière est une phrase du langage–objet. Or la convention T est obtenue en faisant le remplacement de 'S' par sa traduction dans le métalangage (ou par 'S' elle–même dans le cas homophonique).

> **Une théorie de la vérité prenant cette forme est donc un ensemble d'axiomes, à partir desquels on peut, par substitution, dériver toutes les phrases vraies du langage–objet. Appelons une telle théorie une *théorie–T* (parce qu'elle est conforme à la "Convention T"). La théorie en question est dite *absolue*, parce que la satisfaction des prédicats y est définie par assignation de séquences d'objets définis dans un domaine unique qui est l'ensemble des objets du monde, et non pas relativement à des domaines choisis arbitraires, définis par rapport à des interprétations ou à des modèles. On peut donner une autre théorie–T fondée sur cette notion de satisfaction *relative* à un *modèle*. La théorie–T obtenue est alors dite *relative*. [Engel, 1989, page 82]**

Par exemple, la théorie de la vérité que l'on appliquera dans un système axiomatique modal muni des mondes possibles sera de type relatif. Voilà pour la théorie de la vérité tarskienne.

Cependant, cette théorie dont je viens de donner les grandes lignes ne suffit pas pour bâtir une théorie de la signification pour un langage naturel. La sémantique du système formel est dotée d'un contexte de type extensionel tandis que dans le cas des langages naturels on a affaire à des contextes opaques.

La théorie de la signification de Frege, en revanche tient compte de ce type de situations.

Cette dernière théorie est soumise aux deux principes suivants: le *principe de compositionalité* et le *principe de vériconditionalité*. Pour ce dernier principe, donner le sens d'une phrase c'est donner ses conditions de vérités, tandis que le premier affirme que le sens d'une partie constituante d'une phrase détermine la valeur de celle–ci. La valeur sémantique d'une expression pour Frege est sa référence, c'est–à–dire, pour les noms il s'agit des objets, pour les prédicats des relations. Pour une phrase, la référence est sa valeur de vérité.

La question pertinente pour répondre au problème posé au début de cette section sur l'extension de la théorie de la vérité aux langages naturels est celle–ci: quelle est la relation entre la théorie de vérité tarskienne et la théorie du sens de Frege? De plus, la relation ne semble pas impossible car la théorie tarskienne vérifie à la fois le principe de compositionnalité et le principe de véricondionalité. Comme l'indique Pascal Engel:

[6] Par exemple, aux lettres de prédicat correspondront des clauses comme:

a. un objet α satisfait 'F' si α est F

b. un objet α satisfait 'G' si α est G

> [...] une théorie de la signification pour L est une théorie qui nous montre comment dériver, pour chaque phrase L, un théorème de la forme "S est vrai si p". En d'autres termes, une théorie du sens n'est autre qu'une théorie de la vérité, ou une théorie-T. Les phrases-T en effet établissent, dans un métalangage, les conditions de vérité des phrases du langage-objet L, situées à gauche de la conditionnelle. [Engel, 1989, page 142]

Il faut préciser que ce qu'on entend ici par *théorie de la signification* n'est pas la définition générale qui reviendrait à faire une analyse philosophique du concept de signification en fonction des autres concepts, ou du moins en vue d'établir des relations avec d'autres concepts. Il faut plutôt comprendre cette théorie comme une analyse axiomatique du terme. On vise à bâtir un ensemble d'axiomes dont on puisse dériver la signification de chaque phrase donnée d'une langue naturelle, à partir de théorèmes formels. Tarski se montrait réticent vis-à-vis de toute application de la théorie de la vérité aux langages naturels. [7]

5.2.3 La solution davidsonnienne

Donald Davidson, en revanche a relevé le défi. Il a proposé un ensemble de conditions d'adéquation pour qu'une théorie de la signification conçue pour une langue naturelle arrive à revêtir la forme d'une théorie de la vérité au sens de Tarski.

En fait, lorsqu'on regarde de plus près la forme des phrases-T on se rend compte que la phrase à droite de la biconditionnelle est considérée comme une traduction qui énonce les conditions de vérité de la phrase de gauche.

Tarski partait de contextes transparents ou extensionnels et en prenant la traduction comme phrase primitive il pouvait déduire la vérité. Il n'en est pas de même pour les langages naturels. Davidson propose d'inverser la démarche tarskienne, il prend la vérité comme partie primitive des énoncés et à partir d'elle déduit la signification:

> What I propose is to reverse the direction of explanation: assuming translation, Tarski was able to define truth; the present idea is to take truth as basic and to extract an account of translation of interpretation. [...] Truth is a single property which attaches, or fails to attach, to utterances, while each utterance has its own interpretation; and truth is more apt to connect with fairly simply attitudes of the speakers. [Davidson, 1984b, page 134]

Étant donné que la signification ou la traduction d'une phrase dans un langage donné a la forme :

(S) s signifie (dans L) que p,

il y a deux manières de rendre une telle théorie de la signification correcte. Une d'elles est triviale; il suffit de remplacer "p" par s même.

"Juliette aime Roméo" signifie que Juliette aime Roméo

L'autre consiste à remplacer "p" par une *traduction* de s, mais dans ce cas *il faut déjà connaître* la signification de s et de p:

"Juliette aime Roméo" signifie que "Juliette loves Roméo"

Dans les deux cas, il faut présupposer le concept de signification. La stratégie de Davidson consiste à utiliser le concept de vérité comme l'élément premier ou primitif au lieu de celui de signification.

Ainsi on peut remplacer la proposition S par:

(T) "s" est vrai si p

Cela veut dire qu'on applique la Convention-T aux phrases d'une langue naturelle. Le résultat est l'assimilation d'une théorie de la signification propre à une langue naturelle à une théorie de la vérité pour cette langue.

Observons que même dans le cadre d'une théorie homophonique, la Convention-T n'est pas triviale car p peut être interprétée comme étant une des conditions de vérification. Soit l'exemple célèbre:

[7] [Engel, 1989, cf. page 143].

L'assertion "La neige est blanche" est vraie si et seulement si la neige est blanche

> Connaître le sens d'un phrase, c'est *au moins* connaître ses conditions de vérité. En ce sens, le principe fregéen de vericonditionnalité est justifié pour une langue naturelle.[Engel, 1989, page 146][8]

Il ne serait pas juste d'interpréter le programme de Davidson comme une simple identification des conditions de vérité au sens de la phrase. L'idée de Davidson est de montrer:

> [...] qu'une théorie-T pour une langue naturelle doit nous permettre de représenter *suffisamment* de la relation d'identité de la signification utilisée dans (S) au moyen d'une relation moins fine d'identité de valeurs de vérité utilisée dans (T), pour que moyennant les enntraintes empiriques de l'interprétation, une théorie-T puisse, non pas *constituer* une théorie de la signification, mais *servir* de théorie de la signification.[Engel, 1989, page 148]

Davidson se différencie de Quine sous deux aspects. D'abord comme base d'interprétation, Davidson ne prend pas en compte seulement les données du comportement ou du comportement verbal des locuteurs, car en plus il attribue aux sujets des croyances à partir des phrases que les sujets affirment tenir pour vraies. La deuxième différence tient à l'extension du principe de charité. Rappelons nous que chez Quine, il s'agit en gros de considérer le sujet comme cohérent et rationnel alors que chez Davidson, il faut aussi supposer que les croyances que le sujet professe sont pour la plupart vraies. En effet, de même que Quine, Davidson pense que les significations données peuvent être empiriquement testées. Dans le cadre davidsonien cela veut dire que l'on peut déterminer laquelle de toutes les phrases-T applicables de la théorie-T est pertinente pour trouver la signification des comportements des informateurs. Nous voyons bien que dans une démarche visant à la détermination de la signification il faut présupposer que la vérité est une propriété vers laquelle tendent les assertions auxquelles les informateurs donnent leur assentiment.

Or, la signification telle que la définit le programme davidsonien ne sera pas une signification forte sinon qu'au contraire, elle sera pauvre ou minimaliste puisque elle est fondée sur un concept pauvre de la vérité.

La supposition que les sujets sont rationnels ne doit pas être interprétée chez Davidson comme l'admission d'un *fait*; elle est en revanche une *norme*. Selon Engel [Engel, 1994b, page 74] c'est le caractère distinctif de son approche. L'interprétation ne peut avoir lieu que si l'on n'accepte pas la rationnalité des sujets et donc que l'on caractérise cette interprétation comme "rationalisante".

J'ai commencé cette section en affirmant que Davidson pense que l'interprétation des attitudes propositionnelles vont la main dans la main avec l'interprétation du langage. J'ai montré que le principe de charité est la norme de cette interprétation. S'il existe une interdépendance aussi forte entre l'interprétation du langage et celle des conduites ou comportements, il faudra donc que ce principe de rationalité soit également intégré au comportement.

Je décrirai maintenant le rôle de la rationalité dans l'explication de la conduite. J'essayerai de montrer que la démarche de Davidson l'oblige dans tous les cas à préserver la fonction des raisons dans ces explications, cet objectif restant la clef de la solution qu'il propose au problème corps–esprit: le monisme anomal.

---

[8] Je laisse de côté les critiques que l'on pourrait adresser à tort à cette théorie en argumentant que la biconditionnelle de la phrase-T a seulement besoin d'avoir à gauche et à droite des phrases vraies. Selon cet argument la phrase:

> "La neige est blanche" est vraie si l'herbe est verte

devrait nous "donner la signification" de "la neige est blanche" ce qui est évidemment absurde. Cependant cette critique n'a un sens que si l'on ennsidère la Condition-T comme isolée en faisant abstraction du fait que dans la théorie de la vérité on fournit non seulement cette phrase-T mais tout un ensemble de phrases et une structure généeale pour la théorie même. Supposons que nous considérions la phrase précédente comme appartenant à une langue que je vais appeler F et que le but soit de savoir si la théorie-T telle qu'elle est exposée plus haut est aussi une théorie de signification pour le français. C'est-à-dire, si l'on peut *interpréter* les énoncés de F en français. Cependant en appliquant les règles de dérivation et de substitution prévues dans la théorie ds la vérité tarskienne nous verrions que ce n'est pas le cas.

5.3 Le monisme anomal

Dans [Davidson, 1963a], l'auteur donne une interprétation des structures des rationalisations qui, selon lui comprennent deux aspects: d'un coté elles *justifient* les actions et de l'autre elles expriment le *pourquoi* de l'action survenue.

Le premier aspect distingue les actions des événements qui n'en sont pas, les actions étant intentionnelles. Le deuxième aspect, la clause *parce que* constitue une explication causale. Ainsi dans l'exemple suivant: "Jean a choisi de ne pas partir en vacances pour Pâques parce qu'il veut finir au plus vite son diplôme à l'Université et croit que le temps de vacances lui est nécessaire pour préparer ses examens". Or, pour Davidson l'énoncé "il veut finir au plus vite son diplôme à l'Université" est l'énoncé d'une cause tandis que la conviction qu'il ne doit pas partir en vacances sera la justification de l'action de rester sur place. Davidson appelle les justification rationnelles les *raisons primaires* de l'agent pour réaliser une action, ce qui veut dire que les raisons primaires doivent avoir des justifications prises du point de vue de l'agent. Selon Davidson:

> *R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consist of a pro–attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property. [Davidson, 1980a, page 5]

Ce qui est appelé dans le citation de Davidson *pro–attitude* est la croyance *de l'agent* que le sacrifice des vacances lui permettra de mieux préparer ses examens, de les réussir et ainsi d'éviter de redoubler. Cet aspect introduit un élément *intentionnel* dans la rationalisation ou dans l'explication mentale. Supposons dans notre exemple que Jean ne veuille pas reconnaître vis–à–vis de sa famille qu'il a un certain retard dans ses études, pour justifier l'action de ne pas partir en vacances et que confronté à l'étonnement de sa mère, il donne comme justification qu'il fait des économies en vue d'acheter un nouvel ordinateur. Cette explication justifie sa décision de se priver de vacances, mais ce n'est pas la bonne raison. Or, il faut associer une autre condition pour arriver à différencier les raisons primaires des explications fictives. Pour Davidson ceci justifie la connexion qu'il postule entre les raisons et les causes.

La condition qu'il faut ajouter est la suivante: *la raison primaire pour une action est sa cause.* Cette deuxième condition exclut de la définition de raison primaire l'excuse de l'ordinateur donnée dans notre exemple. Comme le signale [Moya, 1990] ceux qui postulent qu'il y a une différence entre raisons et causes ne peuvent éviter que les fausses raisons d'un acte soient données à la place des vraies. Voilà ce qui justifie dans la structure davidsonienne l'identification des raisons avec les causes.

Néanmoins la question à laquelle Davidson doit répondre est la suivante: comment peut–on dire que les raisons sont des causes sans violer le principe de conservation de la matière? Pour y répondre, il nous faut mettre au clair la caractérisation que cet auteur fait des événements physiques et des événements mentaux.

D'emblée tous les événements sont physiques mais certains parmi eux sont des événements mentaux. La réponse naturelle sera selon Davidson qu'un événement est physique s'il peut être décrit dans un vocabulaire physique et mental s'il existe au moins une description exprimée en des termes mentaux. Une description sera mentale si elle comprend un des verbes catalogués comme essentiellement mentaux.

Cependant il reste encore bien des problèmes à régler; il faut entre autres choses trouver des généralisations qui puissent rendre compte des rapports entre les raisons, les causes et les actions de manière à les concilier avec la loi naturelle. Une des possibilités était d'établir des généralisations entre les raisons et les causes, soit dans le sens de lois strictes, soit comme généralisation *ceteris paribus*. La deuxième solution a été choisie par Jerry Fodor qui a lui même défini sa position dans le passage suivant.

> Now lots of people have argued that you can't have intentional laws –however agreeable they might be –because intentional vocabulary is, per se, unfit for nomological projection (See Skinner, Watson, Quine, Davidson, Stich, interalia) The present paper, however, concerns a different kind of worry: If there are psychological laws, then they must be *nonstrict*; they must be "certeris paribus" or "all else equal" laws. There couldn't, for example, be a mental state whose instantiation in a creature literally *guarantees* a subsequent behavior, if only because the world might come to an end before the creature has a chance to behave. [Fodor, 1991, page 21]

Je vais d'abord présenter les concepts fondamentaux dans la théorie de Davidson pour ensuite définir sa solution.

Le concept d'événement et l'individualisation d'événements Les événements sont des entités singulières, aussi impossibles à "répéter" et ontologiquement irréductibles que les voitures, les chaises et les chats. Il peuvent recevoir différentes descriptions et en particulier des descriptions mentales et physiques. Quels seront les événements mentaux dans ce cadre?

Davidson écrit:

> There are no such things as minds, but people have mental properties, which is to say that certain psychological predicates are true of them. These properties are constantly changing, and such changes are mental events ... Mental events are, in my view, physical (which is not, of course, to say that they are not mental). [Davidson, 1994, page 231]

L'individualisation des événements est conçue de façon lâche:

- Les actions sont des événements et de ce fait elles peuvent avoir différents types de propriétés, et en particulier des propriétés mentales et des propriétés physiques. Il est d'ailleurs indifférent de les décrire mentalement ou physiquement.
- La relation de causalité est une relation entre événements, par conséquent elle est également indépendante de la manière dont ces événements sont décrits. [Engel, 1992, page 307]

Ainsi, cette individualisation lâche de l'événement permet de le définir comme une *action* (pour le différencier de simples événements ou *happenings*) s'il admet au moins une description intentionnelle.

Relations causales et explication causale Pour Davidson, il n'existe pas de loi stricte qui gouverne cette relation cause–raison parce qu'il n'existe pas à ses yeux de lois psychophysique. En plus de l'individualisation lâche des événements il postule l'existence de deux niveaux différents dans les relations de causalité:

(a) le niveau ontologique des *relations* causales entre événements particuliers.

(b) le niveau linguistique des *explications* causales où les éléments sont des phrases de description des événements et des relations qu'ils ont entre eux.

Une *relation causale* est une relation entre des événements particuliers. Par exemple, "il y avait un clou sur le pavé qui a causé la crevaison d'un des pneus de la voiture " exprime une relation causale. Cette relation (la relation "C cause E") est un prédicat diadique qui comprend comme termes deux énoncés particuliers ordonnés ainsi dans le cas choisi: "l'existence du clou sur le pavé" et "la crevaison d'un des pneus de la voiture". Les relations causales ont des contextes transparents, donc si la phrase "C cause E" est vraie, alors elle continuera de l'être quelle que soit la description donnée à C et E.

Les raisons ou justifications des événements seront, par contre, des explications intensionnelles, ce qui veut dire qu'elles ont des contextes opaques et qu'elles relèvent du niveau linguistique. Ceci est dû au fait que l'explication d'un événement peut être effectuée pertinemment pour une description mais non pour une autre tout en gardant dans les deux cas une relation causale.

Le fait qu'il existe une *explication* causale qui justifie "C cause E" n'exclut en rien le fait qu'il y ait *relation* causale entre C et E. Selon [Moya, 1990, cfr. page 108] Davidson s'est toujours rallié à la conception de la causalité inaugurée par Hume en la considérant comme basiquement vraie.

Il faut nous rappeler que selon cette conception l'antécédent et le conséquent d'une relation causale doivent être isolés et non connectés. Cependant, les raisons et les causes semblent conceptuellement ou logiquement reliées. Mais croire cela c'est oublier la différence établie par Davidson entre le niveau des explications ou de la linguistique causale et le niveau des relations causales. Au niveau des explications causales intervient cette relation conceptuelle ou logique qui ne se vérifie que pour relier entre elles les descriptions des événements. En revanche, au niveau de la causalité

proprement dite, elle n'entre pas en jeu parce qu'il s'agit là des relations entre les événements, et la relation causale est maintenue quelque soit la description qu'on donne à ces événements. L'exemple suivant explique cette position : "Un court–circuit a causé le feu". Supposons que soit vrai, "Un court–circuit" et "la cause du feu" sont alors des descriptions différentes du même événement. Par conséquent, comme les énoncés des relations causales ont des contextes transparents, on peut remplacer dans l'énoncé premier *salva veritate* ce qui donne comme résultat "La cause du feu a causé le feu". On convient que cette description–là ne nous donne pas beaucoup d'information sur le sinistre même, mais elle énonce une relation causale; et ceci malgré le fait que les descriptions de l'antécédent et du conséquent ont une relation conceptuelle évidente. En contrepartie, la relation causale reste toujours intacte puisqu'elle se situe à un autre niveau. Il nous reste encore à mettre au clair comment les raisons et les causes s'ordonnent et se combinent afin de pouvoir conclure à l'existence d'une loi stricte.

Les lois strictes et le modèle déductif–nomologique Davidson[9] adopte le modèle déductif–nomologique d'Ernest Nagel qui fut exposé précédemment dans le but de bâtir une théorie scientifique de l'intentionalité. Selon cette théorie il croit pouvoir démontrer comment on affirme des énoncés nomologiques à partir d'événements particuliers.

Le modèle de Nagel soutient qu'il existe dans toute explication scientifique deux types d'énoncés différents: l'*explicans* et l'*explicandum*. Les *explicans* sont composés d'une ou plusieurs lois (généralisations nomologiques), qui peuvent être aussi bien des énoncés existentiels que des énoncés universels décrivant les conditions initiales. Cet ensemble d'énoncés doit non seulement être vrai et contenir au moins une loi, mais aussi il doit avoir un contenu empirique vérifiable. L'*explicandum* est un énoncé existentiel ou une loi qui est une conséquence logique de l'*explicans*.

Lorsqu'on applique ce modèle à des rationalisations intentionnelles, un énoncé causal singulier vrai du genre "ce désir et cette croyance causent cette action" joue le rôle d'un explicandum. Cela veut dire que c'est une conséquence logique de l'ensemble des énoncés d'*explicans*. Donc, il existe des descriptions de ce désir, de cette croyance et cette action telles qu'une fois substituées dans l'énoncé initial, elles fournissent des prémisses justes qui constituent des lois appartenant à l'*explicans*.[10]

L'individualisation lâche d'événements telle qu'elle est conçue par Davidson est fondamentale à cet égard, car les lois strictes ne peuvent être que physiques mais un événement n'est pas borné à sa propre description, soit–elle mentale ou physique. Il faut donc que la croyance, le désir et l'action aient des vraies descriptions physiques, parce que seules les loi physiques (chimiques, neurophysiologiques) sont strictes. Ainsi, si la raison cause l'action selon une certaine description, il y aura une description pour chacune des deux (la raison et la cause) qui prendra la forme d'une loi stricte.

Davidson concilie de cette façon deux aspects contradictoires des états mentaux, d'un coté leur rôle causal et de l'autre leur caractère non–nomologique, ce qui rend impossible de les réduire aux lois strictes lorsqu'ils sont décrits comme des événements mentaux. Cette contradiction *apparente* selon Davidson est exprimée par trois principes que sa théorie parvient à concilier; conciliation qui n'est d'ailleurs pas exempte de critiques, étant donné qu'elle nous ramène vraisemblablement à une conception du mental qui en fait un épiphénomène.

Les trois piliers de Davidson

Les trois thèses que Davidson considère comme simultanément vraies sont les suivantes:

- **Principe d'interaction causale:**

  Tous les événements mentaux sont causalement en relation avec les événements physiques. Ces relations se vérifient aussi bien dans le sens mental–physique que dans le sens physique–

[9] cfr. [Davidson, 1963b]
[10] cfr. [Davidson, 1963b, dans Davidson 1980, page 158]

mental.[11]

- **Caractère nomologique de la causalité:**

  Si deux événements sont en relation de cause à effet, alors il existe des lois strictes auxquelles ils sont soumis. Ces lois appartiennent à un système fermé: tout ce qui peut affecter le système doit être contenu dans ces lois.

- **Caractère non–nomologique du mental:**

  Il n'existe point de lois psychophysiques strictes (c'est–à–dire des lois qui connectent des événements mentaux décrits comme étant mentaux avec des événements physiques décrits comme étant physiques).

Davidson souligne que la conjonction logique des trois conditions implique une position moniste. Les événements mentaux ont des rapports causaux avec les événements physiques selon le premier principe. Pour le deuxième principe la relation causale implique l'existence de lois strictes. Cependant au vu du troisième principe ces lois ne sont pas psychophysiques; en fait toutes les lois strictes ne sont que des lois physiques. L'ensemble des événements mentaux ne peut pas être soumis à des lois strictes qui rendent compte de tous les changements possibles parce qu'il ne constitue pas un système fermé. En plus, le caractère holistique du système cognitif nous oblige à tenir compte de la totalité du système lorsqu'on veut affiner les prédictions. Néanmoins, les conditions de cohérence, de rationalité et de consistance qu'on applique à l'égard des explications mentales n'ont aucun écho dans le cadre de la théorie physique [Davidson, 1980b, cfr. page 231].

Voilà pourquoi Davidson affirme le caractère *non–nomologique* ou *anomal* du mental et caractérise sa position comme un *monisme anomal*. Mais la cohérence de tout les trois se base sur le fait que tous les événements sont physiques (position moniste) même les événements mentaux et c'est en cette qualité qu'ils obéissent à des lois strictes.

La survenance dans le monisme anomal

Nous avons vu que Davidson soutient une position non–réductionniste du mental puisqu'il considére que les caractéristiques typiques des descriptions mentales comme la rationalité et la consistance avec l'ensemble du réseau épistémique de l'agent sont totalement extérieures à la description physique et irréductible à celle–ci. Cependant il défend une position moniste. Il faut donc expliquer la corrélation des propriétés mentales avec les propriétés physiques.

Il est évident que Davidson rejette la solution de l'identité de *type* dans le sens où il n'accepte pas que l'on puisse réduire les propriétés mentales aux propriétés physiques. En particulier, Davidson nie explicitement l'existence de lois psychophysiques strictes dans le postulat du caractère non–nomologique du mental. Ces lois sont indispensables si l'on veut pratiquer le réductionnisme de *types*.[12]

Toutefois Davidson affirme qu'il n'a pas besoin de prôner des lois d'identité entre événements physiques et mentaux et ceci pour deux raisons. Premièrement, sa position ontologique quant aux événements; il sont tous physiques mais ils peuvent être aussi mentaux. Deuxièmement, leur individualisation considérée comme lâche; souvenons–nous que les événements peuvent avoir des propriétés aussi bien physiques que mentales sans que ces dernières soient réductibles aux premières et dans ce sens on peut considérer Davidson comme un dualiste de propriétés.

> Here we have a honest ontology of individual events and can make literal sense of identity. We can also see how there could be identities without correlating laws. It is possible, however, to have an ontology of events with the conditions of individualisation specified in such a way that any identity

[11] Dans l'article [Davidson, 1970], l'énoncé de ce principe est une proposition existentielle tandis que dans [Davidson, 1994] il prend la forme d'une proposition universelle

[12] Comme nous avons vu dans le chapitre précédent, l'identité de type selon laquelle la relation entre "l'éclair" et "la décharge électrique à terre d'un nuage de molécules d'eau ionisées" peut être interprétée comme suit: "Pour tout éclair il existe une décharge électrique à terre d'un nuage de molécules d'eau ionisées qui sont identiques". Ceci équivaut à dire qu'il y a des lois–ponts qui mettent en relation les deux événements plus exactement que ne l'exprime la corrélation entre eux.

implies a correlation laws. Kim, for exemple, suggests that Fa and Gb 'describe or refer to the same event' if and only if a=b and the property of being F= the property of being G. The identity of properties in turn entails that (x)(F(x) ≡ G(x)). [Davidson, 1980a, dans Davidson (1980) page 213]

Cependant, la non–réductibilité davidsonienne des propriétés mentales aux propriétés physiques ne doit pas être comprise comme une sorte d'indépendance totale, voilà pourquoi il fait recours dans le texte *Mental Events* au concept de survenance. Davidson est reconnu comme le premier à avoir utilisé ce concept dans la philosophie de l'esprit.

Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot he two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law and there is good reason to *believe* cannot be done; and we might be able to reduce truth in a formal system to syntactical properties, and this we *know* cannot in general be done. [Davidson, 1980a, dans Davidson (1980) page 214]

La question pertinente est: quel est le type de survenance auquel Davidson se réfère? Tout d'abord, un événement est un changement d'un état dans un autre et tout changement d'état implique un changement des propriétés. Davidson soutient qu'il n'y a pas de changement des propriétés mentales sans changement de propriétés physiques.

Une survenance forte doit énumérer pour chaque changement de la propriété mentale m tous les changements des propriétés physiques p dans tous les mondes possibles, cette règle étant valable pour tous les événements.

Dans le cadre de la théorie davidsonienne on voit que ceci pose un problème étant donné tout d'abord la caractérisation des événements comme des *particuliers* et ensuite le caractère lâche de l'individualisation. L'une empêche les généralisations et l'autre ne permet pas d'avoir une description assez fine des propriétés pour faire état de cette relation.

En effet, s'agissant des *particuliers*, Davidson ne veut pas dire que la même corrélation entre les changements respectifs des propriétés mentales et physiques sont vérifiés dans des événements autres que celui dont on traite. L'impossibilité d'établir cette corrélation est dûe à la dualité entre causalité et explication qui semble admettre deux types d'éclaircissements également pertinents: d'un côté les explications mentales (par des raisons) et de l'autre les explications causales. Cependant les explications fondées sur des raisons ne sont pas réductibles aux niveaux physiques. Or, leur valeur normative et leur connotation rationnelle nous obligent à tenir compte du réseau épistémique de l'agent, ce qui ne permet pas de déterminer une corrélation figée entre les propriétés mentales et les propriétés physiques dans tous les cas.

Ceci pose certes des problèmes pour l'individualisation des événements; Davidson affirme que deux événements sont égaux si et seulement si ils partagent les mêmes causes et les mêmes effets. Plusieurs auteurs lui ont fait remarquer qu'il n'a pas ajouté grand–chose aux critères d'individualisation parce qu'il est évident que lorsque deux choses sont différentes sous le rapport d'une propriété, l'on peut conclure qu'elles sont différentes tout court. Cependant, sans être en désaccord avec cette observation, je pense que Davidson veut mettre l'accent sur le caractère holistique de la cognition dans la causalité. Ainsi dans deux mondes possibles différents un événement peut avoir les mêmes propriétés mentales et des corrélations différentes pour ce qui est de ses propriétés physiques dans chacun de ces mondes puisque le réseau épistémique peut être différent.

De tout que je vient d'exposer il me semble peu probable, voire impossible qu'on puisse assurer une stabilisation des corrélations entre propriétés physiques et mentales dans tous les mondes possibles. En effet, on ne sait pas *quelles* propriétés pbysiques doivent changer avec les propriétés mentales, et c'est pourquoi la théorie ne peut nous dire *en vertu de quoi* les causes mentales causent leurs effets physiques.(cfr. [Engel, 1992]). C'est cette incertitude qui ne permet pas à Davidson de conclure à une survenance forte et l'oblige à proposer une survenance *faible*.

5.4 Les limites de l'approche davidsonienne

La critique fondamentale que m'inspire cette identification des raisons avec les causes est double: les explications causales tenant au caractère non–nomologique du mental n'ont aucune couverture légale et sont donc privées de tout pouvoir *prédictif*. En effet l'absence de lois universelles ne permet de donner aux explications causales qu'un pouvoir *rétrospectif* et retire à cette théorie une caractéristique importante de toute théorie scientifique: le pouvoir de prédiction au moins pour ce qui est des intentions.

La deuxième critique, qui tient à la première, est que les explications causales qui utilisent la raison sont immédiates et évidentes dans les cas à la première personne. Savoir pourquoi nous avons agi d'une certaine façon procède d'une démarche qui a les mêmes caractéristiques que l'étude de la posture actuelle de nos corps par exemple, c'est–à–dire qu'elle est immédiate et qu'on n'a pas besoin d'entreprendre une analyse inductive de nos croyances et de nos désirs. Cependant, lorsqu'il s'agit de trouver les explications des actions d'autrui, on est obligé de les interpréter dans un cadre normatif, d'imaginer des relations épistémiques dans un réseau des croyances et des désirs propres à l'acteur, supposer que ce dernier est un agent rationnel.

Dès lors la tâche d'inférence, à partir des croyances et des désirs que nous attribuons à l'agent fait que le résultat de cette démarche a uniquement une valeur de *bona fide*. On a beau dire que l'explication qu'on trouvera sera valable si et seulement si les raisons sont des causes car étant donné le caractère non–nomologique du mental nous ne disposons pas de généralisations légales qui nous permettent d'accorder meilleur crédit à notre résultat. Dans le cas de la troisième personne, l'identité entre les raisons et les causes est purement théorique car nous n'avons aucun moyen de la vérifier. [13]

Une autre critique (qui est d'ailleurs la plus répandue) [14] au sujet du monisme anomal tient au péril évident d'épiphénoménalisme du mental. Les trois principes proposés par Davidson fournissent une solution élégante au problème de la causalité mentale mais cette solution a ses limites. Les événements mentaux causent des événements physiques mais pas en tant que mentaux sinon en tant qu'événements physiques eux–mêmes. On peut dire qu'il existe un épiphénoménalisme des propriétés mentales puisque ce n'est que lorsqu'on les décrit comme étant d'ordre mental que les événements deviennent des causes. Cette proposition doit être distinguée de celle qui affirme que le mental est inerte du point de vue causal puisque le monisme anomal ne soutient ni n'implique cela. Cependant le monisme anomal dénué du principe de survenance implique comme je l'ai déjà dit, l'épiphénoménalisme des propriétés mentales et cela signifie qu'elles ne peuvent avoir un rôle causal [Kim, 1993a, cfr. page 20].

Pourtant, les propriétés mentales, peut nous répondre un davidsonien, surviennent aux propriétés physiques et il y a donc une corrélation entre elles. Lorsqu'il y a un changement de propriétés mentales il y a un changement des propriétés physiques. Cependant cette corrélation est une survenance faible, ce qui veut dire qu'on ne sait pas quelles propriétés physiques sont à l'origine de quelles propriétés mentales et notre incertitude rend impossible la formalisation nomologique au niveau intentionnel et ne contribue donc point à anéantir le risque de reléguer au rang d'épiphénomène les propriétés mentales. Je suis en accord avec [Kim, 1993a] quand il observe que le monisme anomal seul implique l'épiphénoménalisme du mental; néanmoins lorsqu'on ajoute la survenance, même quand elle est faible, cette objection disparaît. Nous avons exposé dans le deuxième chapitre les divers concepts de survenance et nous avons vu que la survenance faible se vérifie pour un monde possible où les propriétés physiques seraient les mêmes que celles de notre monde mais qui serait par contre vide de propriétés mentales. Je pense qu'il ne faut pas oublier que chez Davidson les prémisses du monisme anomal ne permettent pas une situation de ce type par le *Principe d'interaction causal* puisque ce cas extrême où le monde possible serait vide du mental est éliminé.[15].

[13] Je remercie Ruwen Ogien de m'avoir donné accès à son manuscrit intitulé *L'action* et d'avoir discuté avec moi de ces deux points.

[14] Pour avoir une perspective générale des critiques voir [LePore and McLaughing, 1985]

[15] Pour un compte–rendu des arguments de Davidson et des réponses de ses critiques voir les quatre premiers articles de [Heil and Mele, 1993]: [Davidson, 1994], [Kim, 1993a], [McLaughlin, 1993], [Sosa, 1993]

5.5 Conclusion

Le but de ce chapitre était de présenter la théorie de Davidson et de montrer que bien que ce soit une solution élégante, elle n'est pas satisfaisante pour établir un principe de corrélation forte nécessaire à la solution du problème corps–esprit.

Premièrement, le monisme anomal menace les entités mentales d'être réduites au rang d'épiphénomènes. Secondo, les corrélations entre les propriétés mentales et les propriétés physiques dans ce cadre permettent seulement d'établir une survenance faible. L'existence d'une survenance faible n'est pas satisfaisante pour résoudre le problème corps–esprit. Rappelons–nous[16] que j'ai signalé que la survenance faible n'empêche pas qu'il y ait un monde qui possède les mêmes propriétés physiques que le nôtre tout en étant totalement dépourvu de propriétés mentales.

L'unique réponse valable à mon avis au problème corps–esprit est la proposition, la plus explicite possible, d'une solution du principe de corrélation entre physique et mental. Il est nécessaire de savoir quelles propriétés physiques sont à l'origine ou sont la base de quelles propriétés mentales. Davidson ne peut pas répondre à cette question parce qu'il a choisi une individualisation lâche des événements et des états. Malheureusement, ce type d'individualisation bien que susceptible de nous éviter une démarche réductionniste, ne semble pas être assez stricte pour établir une corrélation forte des propriétés mentales et des propriétés physiques.

Finalement, l'individualisation des états physiques se fait uniquement en relation avec leur pouvoir causal; il me semble que cette caractéristique seule ne suffise pas dans le cas de Davidson. Je vais montrer au chapitre 7 qu'elle ne suffit pas non plus à la théorie de Fodor. Dans le souci de bâtir une théorie non–réductionniste du mental on a eu peur d'en exclure les caractéristiques du substrat en les substituant par des entités plus ou moins neutres comme celles qui sont caractérisées par les rôles causaux/fonctionnels. Mon but, je l'ai déjà dit est de montrer que les démarches de ce type ont échoué.

La thèse que je soutiens tente de surmonter cette difficulté. Elle avance qu'il existe d'autres caractéristiques physiques qui, bien que déterminées par le(s) substrat(s) physiques, n'y sont pas réductibles.

[16] Voir chapitre 2

Chapitre 6

Le fonctionnalisme

I am a realist and a reductive materialist about mind. I hold that mental states are contigency identical to physical - in particular neural - states. ... Like Smart and Amstrong, I am an ex-Rylean. In view of how the term is contested, I do not know whether I am a 'functionalist'.
[Lewis, 1994, page 412]

6.1 Introduction

Les behavioristes avaient raison d'utiliser le paradigme de l'entrée / sortie d'une boîte noire et de considérer l'agent comme un système dont les réponses dépendent de ses entrées. Néanmoins ils avaient tort de négliger le *rôle causal* que cette boîte pouvait jouer. À l'origine de cette faute se trouvait l'éliminativisme radical du mental qui les animait.

Les théories fonctionnalistes, le fonctionnalisme de type et le fonctionnalisme computationnel essaient d'exploiter ce rôle causal. Dans ce chapitre et le suivant je vais présenter ces deux théories fonctionnalistes.

La théorie de l'identité de types que je traite dans ce chapitre contrairement au behavourisme, prend en compte des notions internes. Elle a tort de ne pas les définir de façon relationnelle, c'est-à-dire en relation à leurs propres propriétés, et de les identifier en revanche aux états physiques. Le fonctionnalisme de type veut exploiter les avantages du behaviorisme et de la théorie de l'identité tout en évitant leurs défauts. De même que le behaviorisme, il considère l'agent comme un système d'entrées\sorties mais comme la théorie de l'identité de types il admet l'existence des états internes en leur conférant des rôles causaux.

6.2 Types de fonctionnalismes

On fait souvent remonter les origines du fonctionnalisme en philosophie et en psychologie aux oeuvres d'Aristote. Le terme "fonctionnalisme" est utilisé en d'autres sciences comme l'anthropologie et la critique littéraire mais selon Block [Block, 1980b] on ne pourrait pas dire qu'il s'agit du même concept dans tous les cas. Dans la psychologie et la philosophie entre autres, le fonctionnalisme s'avère être une stratégie de recherche, une méthode d'analyse. Cette stratégie de l'analyse fonctionnaliste consiste à réduire à ses composantes le système que l'on veut expliquer. L'explication sera alors donnée en termes des capacités des composantes et de la façon dont ces parties s'intègrent.

Néanmoins lorsque l'on regarde de plus près l'explication fonctionnaliste on en reconnaît deux formes différentes [Cummins, 1980, cfr. page 186–187].

Ces deux analyses sont inspirées par la correlation entre le concept de *fonction* et le concept de *disposition*. Attribuer une fonction agissante à un objet d'un système implique, ne serait–ce qu'en partie, qu'on lui trouve les dispositions nécessaires pour remplir cette fonction. Au contraire, un objet peut posséder des dispositions pour faire quelque chose sans que ce soit pour autant sa fonction. Par exemple, les personnes portant un stimulateur cardiaque (*pace–maker*) sont interdites de l'utilisation du four à micro–ondes. La raison est qu'un phénomène de couplage électromagnétique se produit entre le stimulateur et le four à micro–ondes de façon telle que celui–là prend les pulsations du four pour les battements normaux du coeur. En conséquence, le stimulateur s'arrête et la personne s'évanouit. Le four à micro–ondes a la disposition d'arrêter le stimulateur cardiaque mais je pense que l'on conviendra que telle n'est pas sa fonction. [1]

Par contre, il y a des bactéries marines qui peuvent survivre seulement dans des zones dépourvues d'oxygène. Ces bactéries sont porteuses de magnétosomes qui ont comme fonction d'indiquer les zones viables pour l'organisme au fond de la mer. Les magnétosomes ont la disposition de signaler la direction du nord magnétique pour les bactéries habitant l'hémisphère nord, mais c'est le contraire pour leurs congénères de l'hémisphère sud. La disposition d'indiquer le nord magnétique est équivalente à la disposition d'indiquer le fond de la mer. Or la fonction d'indication de zones viables pour la survie se manifeste par la capacité d'indiquer les zones dépourvues d'oxygène. [2]

Une fois que l'on a établi la nuance entre les concepts de *fonction* et de *disposition*, on peut passer à l'explication des deux stratégies d'analyse fonctionnelle. La stratégie première se base sur l'analyse des dispositions, l'idée sous–jacente étant que l'existence des régularités dispositionelles se manifeste par la régularité des comportements auxquels ces dispositions sont liées. La deuxième stratégie met l'accent sur les adscriptions des fonctions pertinentes à un système.

1.- La stratégie de l'immersion À la base de cette stratégie on trouve le concept de *disposition*. Supposons que a a la disposition d, cela veut dire qu'il existe une régularité dispositionelle qui fait que certains types causent la manifestation de d en a. Néanmoins, il ne suffit pas que cette disposition d se manifeste en a pour être considérée comme telle car il faut aussi qu'il y ait une connexion entre ces manifestations et une loi générale. Cummins emprunte à Brian O'Shaughnessy l'exemple de la propriété dispositionelle d' "élévation" suivante:

> Consider the disposition he calls *elevancy*: the tendency of an object to rise in the water of its own accord. To explain elevancy, we must explain why freeing a submerged elevant object causes it to rise. This we may do as follows. In every case, the ratio of an elevant object's mass to its nonpermeable volume is less than the density (mas per unit volume) of water. Archimedes' principle tells us that water exerts an upward force on a submerged object equal to the weight of the water displaced. In the case of an elevant object, this force evidently exceeds the weight of the object by some amount f. Freeing the object changes the net force on it from zero to a net force of magnitude f in the direction of the surface, and the object rises accordingly. Here, we subsume the connection between freeings and risings under a general law connecting changes in net force with changes in motion by citing a feature of elevant objects which allows us (via Archimedes' principle) to represent freeing them under water as an instance of introducing a net force in the direction of the surface.[Cummins, 1980, page 186]

Ainsi la propriété dispositionelle de flottaison sera expliquée par la connexion entre la montée à la surface et l'immersion due à un changement des forces produit par une propriété de l'objet en question. La force dirigée vers la surface qui produit la montée en surface résulte d'une application particulière d'une loi plus générale, la loi d'Archimède.

2.- La stratégie analytique Dans cette stratégie on parle des capacités au lieu des dispositions bien que les dispositions et les capacités ne soient pas exactement la même chose. On peut assigner une fonction à quelque chose dès lors qu'elle aura la *capacité* mais pas forcément la *disposition* d'accomplir cette fonction. Cependant et en laissant de côté ces différences, la notion de capacité se trouve être plus appropriée à l'analyse fonctionnelle. L'analyse fonctionnelle est

[1] Je remercie Luc Tissot, fondateur d'Intermedics qui m'a suggéré cet exemple.
[2] Exemple tiré de [Drestke, 1986]

habituellement utilisée dans le domaine technique. Par exemple dans un diagramme eo électronique, chaque symbole représente un objet physique qui a une certaine capacité. L'analyse d'un engin complexe consiste en l'analyse du tout en fonction des analyses de ses parties. Dans l'analyse orientée objets en informatique chaque objet est une unité de programme qui a une ou plusieurs fonctions. L'ensemble de ces objets forme un tout doot l'intégration remplit les tâches imparties par les concepteurs du système. [3]

Quel est l'intérêt de l'explication analytique fonctionnelle?

Selon Cummins [Cummins, 1980, cfr. page 189] l'intérêt de ce type d'explication peut se définir selon les critères suivants:

1. l'étendue des capacités à analyser s'avère être d'un type très différent de celles qui servent d'explication.
2. l'étendue des capacités utilisées dans l'explication s'avère être moins complexe que celles que l'on veut expliquer.
3. la relative sophistication des programmes ou paradigmes sert à démontrer la complexité de la relation entre les parties et le processus qui les composent.

Selon Cummins, plus la distance est grande entre les niveaux respectifs de complexité des propriétés à analyser et celles que servent à les expliquer, plus complexe sera le programme proposé. Il donoe l'exemple de la tbéorie des automates.

> Automata theory supplies us with extremely powerful techniques for constructing diverse analyses of very sophisticated tasks into very unsophisticated tasks. This allows us to see how, in principle, a mechanism such as the brain, cousisting of physiologically unsophisticated components (relatively speaking), can acquire very sophisticated capacities. It is the prospect of promoting the capacity to store ones and zeros into the capacity to solve logic problems and recognize patterns that makes the analytical strategy so appealing in cognitive psychology. [Cummins, 1980, page 189]

Ces deux manières de conçevoir la stratégie fonctionnaliste sont à l'origine de deux positions différentes dans la psychologie et la philosophie de l'esprit.

L'analyse fonctionnelle des concepts mentaux développée par Armstrong et D. Lewis qui l'appelaient *analyse du rôle causal.*

La théorie empirique computationnelle de l'esprit de Jerry Fodor et d'Hilary Putnam.

Le fonctionnalisme orienté vers l'analyse du rôle causal suit le chemin défini par la première de ces stratégies car à la base de l'explication on trouve les dispositions du comportement. La deuxième en revanche s'inspire de l'analyse fonctionnelle et je vais exposer les différents paradigmes qui ont été utilisés pour formuler l'explication fonctionnelle.

Dans ce chapitre je présenterai le fonctionnalisme métaphysique tandis que dans le suivant je tenterai de décrire la théorie empirique computationnelle.

6.3 Le fonctionnalisme selon Lewis et Armstrong

6.3.1 Les antécédents

L'analyse du rôle causal des états mentaux admet deux méthodes dictées par des inspirations différentes: l'analyse du behaviorisme logique de Gilbert Ryle (1900–1982), et la notion de la *description neutre du mental* (*neutral topic mental description*) prônée par J. J. C. Smart (1920–) qui était un partisan de la tbéorie de l'identité.

[3] Je me réfère ici au concept fonctionnel selon une conception plus large que celle utilisée pour differentier les types d'analyse, de conception d'un système informatique. Dans ce cadre le terme fonctionnel est perçu comme différent et même opposé à la conception orientée objets. Ces dénominations sont valables seulement dans ce cadre très précis et restreint. Lorsque l'on prend le concept fonctionnel dans un seus plus large et plus général, on peut convenir de considérer la stratégie fonctioonelle comme celle que l'on applique aussi dans la conception orientée objets.

Les dispositions du comportement

Gilbert Ryle publia en 1949 son livre *The concept of the mind* qui eut une énorme influence. A la différence de Watson ou de Skinner qui préconisaient un éliminativisme radical du mental, le projet de Ryle implique que le behaviorisme traite le mental en termes de comportements. Ce type de behaviorisme a été appelé behaviorisme logique. Ryle n'a pas eu une position radicalement éliminativiste. Bien au contraire, il entreprit de faire une description des concepts courants du mental tels que ceux qu'utilise la psychologie ordinaire. Selon Ryle on commet des "erreurs de catégorie" lorsqu'on conçoit l'esprit et le corps comme deux choses différentes.

> ... Idealism and Materialism are answers to an improper question. The 'reduction' of the material world to mental states and processes, as well as the 'reduction' of mental states and processes to physical states and processes, presupposes the legitimacy of the disjunction 'Either there exist minds or there exist bodies (but not both)'. It would be like saying 'Either she bought a left-hand and a right-hand glove or she bought a pair of gloves (but not both)'[Ryle, 1949, page 30]

Il ne faut pas considérer les termes mentaux comme désignant des objets mais comme des propriétés dispositionelles du comportement. Ces dispositions du comportement sont comparées aux propriétés dispositionelles des autres objets, par exemple la fragilité d'un verre. Dans ce cas-là, selon le behaviorisme, il y aurait un ensemble de propriétés contrafactuelles qui exprimeraient cette disposition à la fragilité du type : *Si le verre tombe sur une surface dure, il va se briser.* Mais Ryle attire notre attention sur le fait qu'une propriété dispositionnelle n'est pas un état interne à l'objet, mais nous indique seulement son comportement sous de telles conditions. Le mérite du travail de Ryle est d'avoir souligné le fait que chaque événement mental peut être vu comme un état dispositionnel dans le réseau des comportements et des autres états mentaux. [4]

L'analyse neutre du comportement

J. J. C. Smart [Smart, 1959] voit dans le fonctionnalisme une possibilité de réfuter l'objection émise par les dualistes de propriétés à l'encontre de la théorie de l'identité de types dont j'ai déjà parlé au chapitre 2 (§2.5.2). Rappelons nous que les dualistes de propriétés récusent l'identité entre les propriétés mentales et les propriétés neurologiques. Ils accordent aux physicalistes que chaque état mental est identique à un état neurologique mais que malgré cela, on peut dénier l'identité entre leurs respectives propriétés. Cela veut dire que les états physiques ont des propriétés phénoménales qui sont irréductiblement mentales.

L'explication fonctionnaliste est ontologiquement neutre. Les états mentaux sont caractérisés par leur rôle causal dans un réseau qui comprend d'autres états mentaux, des stimuli et des réponses de comportement mais elle ne se prononce point sur le statut ontologique des objets qui remplissent ces rôles. Le fonctionnalisme est compatible, à la limite, avec les deux types de considérations ontologiques possibles que l'on peut émettre pour ces entités, soit physique soit mentale.

En profitant de la neutralité ontologique dont jouissent les entités mentales dans le cadre fonctionnaliste, Smart signale que l'objection *a priori* à l'existence des propriétés mentales irréductibles formulée par les dualistes peut être levée, car le fonctionnalisme ne se prononce pas sur leur statut ontologique.

La position de Smart conduit à une rupture avec le behaviorisme car elle admet non seulement l'existence des états internes mais aussi qu'on leur attribue des pouvoirs causaux.

Je pense que cette position choisit une issue intermédiaire pour résoudre l'objection des dualistes de propriétés. Elle accepte, en effet de tenir compte des états internes sans se prononcer sur leur identité ontologique, donc sans affirmer qu'ils sont irréductibles comme le disent les dualistes de propriétés.

Lewis et Armstrong n'en restent pas là; ils utilisent aussi cette neutralité pour prouver la thèse de l'identité même.

---

[4] Hilary Putnam [Putnam, 1965] signale que les dispositions au comportement ne sont pas suffisantes pour assurer un comportement. Il cite plusieurs expériences de pensée subies par des "super-spartiates" que leur éducation et culture ont conditionné de telle manière qu'aucun signe ne nous laisse percevoir leur souffrance.

6.3.2 L'approche de Lewis et d'Armstrong

D. M. Armstrong (1926–) fait l'éloge du concept ryléen de disposition du comportement dans son texte *The Nature of Mind* [Armstrong, 1970]. Néanmoins, il signale que le behaviorisme échoue lorsqu'il s'agit de donner un compte–rendu des événements mentaux qui n'ont pas une corrélation dans le comportement, comme par exemple "il réfléchit à la démonstration du théorème de Fermat". C'est pourquoi selon lui les behavioristes doivent faire la correction suivante dans le concept de disposition du comportement:

> ... in order to reach the correct view, I am suggesting, they would have to conceive of these dispositions as actual *states* of the person who has the disposition, states that have actual power to bring about behaviour in suitable circumstances. But to do this is to abandon the central inspiration of Behaviourism: that in talking about the mind we do not have to go behind outward behaviour to inner states.[Armstrong, 1970, page 196]

David K. Lewis (1941–), pour sa part et indépendemment d'Armstrong arrive à la même conception des états mentaux. Pour Lewis, un état mental M est aussi l'occupant d'un rôle M. La théorie de la description neutre du mental de Smart lui permet d'affirmer que la détermination de l'état qui occupe le rôle (qu'il soit physique ou neuronal, constitué de puces de silicium ou d'un ensemble de verres de bière) ne pose un problème pertinent qu'*a posteriori*. Après avoir acquis plus d'information on devrait pouvoir dire de quel type d'activité neuronale il s'agit. [Lewis, 1994, cf. page 418]

L'analyse du rôle causal prônée par Lewis et Armstrong perpétue le clivage établi antérieurement entre le behaviorisme et la théorie de l'identité de types au sujet du type de réduction possible. Rappelons–nous que le behaviorisme pratiquait une réduction analytique linguistique car il considérait que les énoncés mentaux et les dispositions du comportement avaient une relation de synonymie; le physicalisme de type par contre pratique une réduction synthétique *à posteriori*.

Lewis signale que la réduction analytique ou de synonymie ne peut pas être appliquée dans ce cas car un rôle fonctionel ne correspond pas invariablement à un même nom. Ils ne sont ni l'un ni l'autre des désignateurs rigides. Il s'agira en revanche d'identification contingente ou occasionnelle.

Lewis propose une méthode pour justifier cette identité occasionnelle, c'est–à–dire qu'il tente de montrer le chemin pour une réduction du mental à des entités ontologiquement neutres mais existantes. Il s'agit de trouver les référents des entités qui occupent les rôles causaux du comportement. Ces référents dont l'ontologie est neutre ont une existence métaphysique et ils existent dans le paradis platonicien.

Selon l'avis de Lewis, si une théorie physiologique convient, elle pourrait impliquer des entités psychophysiques. L'implication chargée de faire le pont devra prendre la forme suivante:

> Mental state M = the occupant of causal role R (by definition of M) Neural state N = the occupant of causal role R (by the physiological theory) Therefore, mental state M = neutral state N (by transitivity of =)
>
> If the meaning of the names of mental states were really such as to provide the first premise, and if the advance of physiology were such as to provide the second premise, the the conclusion would follow. Physiology and meaning of words leave us no choice but to make the psychophysical identification. [Lewis, 1972, page 207]

Cette structure théorique responsable de la réduction nécessite une théorie psychologique. La question que l'on se pose à ce point est: à partir de quelle théorie psychologique définit–on la clé de cette structure théorique, c'est–à–dire les rôles?

La théorie que Lewis a choisie est la psychologie ordinaire.

> Applied to common–sense psychology – folk science rather than professional science, but a theory nonetheless – we get the hypothesis ... that mental state M (say, an experience) is definable as the occupant of a certain causal role R – that is, as the state, of whatever sort, that causally connected in specified ways to sensory stimuli, motor reponses, and other mental states.[Lewis, 1972, pages 207–208]

Résumons, jusqu'ici nous avons les données suivantes: d'un côté la théorie –comme Lewis la considère– de la psychologie ordinaire et de l'autre les rôles causaux qu'elle détermine. Chaque rôle causal détermine un état mental et ce dernier occupe une place dans le réseau causal. La tâche

est justement de trouver l'état qui satisfasse ce rôle. Il s'agit à la fois de mettre en évidence un nom ou un référent pour ces rôles et de montrer leur existence métaphysique.

Pour y parvenir, Lewis utilise la méthode de Ramsey. Nous allons expliquer cette méthode née de l'échec du vérificationisme et qui se révèle être un outil de choix pour déterminer les identifications entre les rôles et les états *a posteriori*.

6.3.3 La méthode de Ramsey

Un peu d'histoire

Le vérificationisme strict avait échoué à cause de la critique popperienne [5]. Les empiristes logiques entreprennent le "tournant linguistique" dont j'ai déjà parlé dans le premier chapitre. A cette fin, ils décident de faire la différence entre deux types de langage existants dans toute théorie scientifique selon la nature des termes que composent ses assertions. D'un côté nous avons le *langage observationnel* contenant des énoncés composés du vocabulaire logique et des termes désignant tous des entités publiquement observables. De l'autre côté, il y aurait un *langage théorique* s'appliquant à des entités qui ne sont pas observables comme par exemple "proton" ou "nucléaire". Comme le signale Pierre Jacob:

> Cette distinction repose [donc] sur deux présuppositions: qu'il est possible de distinguer le vocabulaire purement logique du vocabulaire extra-logique (ou descriptif) des sciences; qu'à l'intérieur du vocabulaire descriptif on peut distinguer les termes qui ne désignent que des entités observables de ceux qui n'en désignent jamais. Le principe sémantique de l'empirisme affirme qu'un énoncé observationnel, ne contenant (outre le vocabulaire logique) que des termes désignant des entités observables, se comprend plus facilement qu'un énoncé théorique. Le but de l'empirisme est alors de mettre à jour les relations systématiques qui unissent, dans le langage de théories scientifiques, les énoncés théoriques aux énoncés observationnels, censés leur conférer leur "signification empirique". [Jacob, 1980, page 129]

Or, le but est trouver un mode d'identification théorique entre les termes appartenant au vocabulaire théorique et leurs correspondants du vocabulaire observationnel. Il s'agit d'établir des lois-ponts, comme avons déjà vu dans le premier chapitre.

Le problème est de définir le statut de ces règles dont on peut dire qu'elles sont arbitraires ou conventionnelles, mais ceci n'est pas tout-à-fait le cas parce qu'elles relient un terme théorique à plusieurs procédures expérimentales, qui sont des conséquences de l'adhésion hypothétique à l'existence de l'entité dénotée par ce terme théorique: si vous croyez à l'existence d'un champ gravitationnel, vous observerez telle interaction entre la lune et un satellite passant dans son voisinage, selon l'exemple emprunté à Pierre Jacob.

Entre les années 1930 et 1940 on s'aperçut que l'objectif d'éliminer totalement le langage théorique au bénéfice d'un langage exclusivement observationnel n'était qu'un mirage. Face à cet acquis, il fallut imaginer des méthodes de vérification. On a pensé ainsi que les propriétés purement logiques des théories, une fois réduites à l'état d'un système d'axiomes et à leurs conséquences déductibles, pourraient tout simplement laisser la place à l'ensemble de leurs conséquences formulées en termes observables.

On a proposé alors deux méthodes. L'une est dûe à W. Craig, l'autre a été empruntée par Carnap à Frank Ramsey (1903-1930). Nous traiterons ici seulement de la deuxième.

Soit l'exemple suivant [Jacob, 1980, cf. page 153]

---

[5] Entre autres choses parce que la vérification des énoncés universels empiriques n'est pas possible dans le cas où la théorie s'avère être vraie.

> Si j'énonce la théorie que "Tous les cygnes sont blancs", à moins de savoir que j'ai examiné exhaustivement la totalité des cygnes de l'univers, la théorie demeurera invérifiable, quel que soit le nombre de cygnes examinés. En revanche, dès que j'observe un cygne noir, ma théorie se trouve réfutée.[Jacob, 1980, page 127]

EXEMPLE 2

$$\forall x(Px \supset Ix) \tag{6.1}$$
$$\forall x(Px \supset Ax) \tag{6.2}$$
$$\forall x(Px \supset Tx) \tag{6.3}$$
$$\forall x(Px \supset Vx) \tag{6.4}$$
$$\forall x(Px \supset Ex) \tag{6.5}$$
$$\forall x(Px \supset Bx) \tag{6.6}$$
$$\forall x(Ix \supset Fx) \tag{6.7}$$

La première équation est le postulat théorique qui affirme que

> Le phosphore blanc a une température d'ignition de 30^0 centigrades.

Les termes P et I sont des termes théoriques. Les règles qui suivent (6.2 à 6.6) relient P, un des termes théoriques, à des termes du vocabulaire observationnel.
Dans l'ordre, à partir de 6.2 jusqu'à 6.6 elles expriment:

- Le phosphore blanc a une odeur d'ail.
- Le phosphore blanc est soluble dans la térébenthine.
- Le phosphore blanc est soluble dans l'huile végétale.
- Le phosphore blanc est soluble dans l'éther.
- Le phosphore blanc brûle la peau.

La dernière formule est la règle de correspondance suivante,

> Si un objet a une température d'ignition de 30^0 C, alors quand il est environné d'air ayant une température au moins égale à 30^0, il s'enflamme.

La méthode de Ramsey a le mérite de dissoudre toutes les questions métaphysiques dangereuses sur la "réalité" des entités inobservables : elle substitue les termes suspects, c'est–à–dire les termes théoriques, des assertions par des énoncés de propriétés quantifiées existentiellement. Il procédera comme suit.

Soit T le postulat premier qui contient les termes suspects P et I. Soit C l'ensemble des énoncés observationels pour P et la règle de correspondance pour I.

Alors l'ensemble des postulats est la conjonction TC. On peut former l'*énoncé de Ramsey* à partir de l'ensemble TC en remplaçant toute occurrence des termes suspects P et I par une variable de propriété (Ψ et Φ), précédée d'un quantificateur existentiel:

$$(\exists\Psi)(\exists\Phi)(\forall x)\,(\Psi x \supset (Ax \wedge Tx \wedge Vx \wedge Ex \wedge Bx)) \wedge (\Phi x \supset Fx) \wedge (\Psi x \supset \Phi x)$$

L'énoncé de Ramsey a fait disparaître les termes suspects P et I. Cet énoncé affirme qu'il existe deux propriétés Ψ et Φ qui jouent exactement le rôle qu'un chimiste attribuerait à P et I.

Cette dernière équation exprime le même contenu que l'ensemble des précédentes. Cependant, elle est différente à deux égards: d'un part, elle n'utilise pas de termes suspects et d'autre part, les variables liées aux quantificateurs ne sont plus du même type logique. Cette dernière expression permet ainsi d'économiser une partie théorique du vocabulaire descriptif tout en enrichissant le vocabulaire théorique. En effet, au lieu que le domaine des variables liées contienne seulement de simples individus, le domaine des variables liées par les quantificateurs contient des propriétés.

L'utilisation de la méthodes de Ramsey par le fonctionnalisme de types

Lewis utilise la méthode de Ramsey pour déterminer les instantiations des rôles. Les termes mentaux ne sont pas spécifiés comme étant ceux qui instancient ou qui réalisent la théorie mais une fois les référents ou les noms trouvés ce sont ces derniers qui instancient la théorie. De façon analogue au cas précédent tiré de la physique, les états suspects, mentaux dans le cas présent, sont introduits en utilisant des termes observables de stimuli – réponses. Les énoncés observables expriment des relations causales entre eux et les termes suspects, c'est-à-dire entre eux et ces états de nature non spécifiée.

Soit l'énoncé suivant de la psychologie populaire:

- Si Jean croit qu'il est en retard, il va vouloir téléphoner pour l'annoncer.
- Si Jean croit qu'il est en retard, le temps estimé du voyage est supérieur au laps de temps écoulé entre le moment où il commence a acquérir cette conviction et l'heure d'arrivée prévue.
- Si Jean croit qu'il est en retard, il va a commencer à rouler plus vite que d'habitude.
- Si Jean croit qu'il est en retard, il ne va pas s'arrêter pour acheter des cigarettes.
- Si Jean veut téléphoner pour annoncer son retard, alors s'il trouve sur son chemin un téléphone il va s'arrêter pour téléphoner

Ce que l'on peut représenter comme suit:

$$\forall x(Rx \supset Tx) \tag{6.8}$$
$$\forall x(Rx \supset Vx) \tag{6.9}$$
$$\forall x(Rx \supset Px) \tag{6.10}$$
$$\forall x(Rx \supset Ax) \tag{6.11}$$
$$\forall x(Tx \supset Cx) \tag{6.12}$$
$$\tag{6.13}$$

Les termes suspects qui appartiennent au postulat théorique sont la croyance que l'on attribue à Jean d'être en retard et son désir de l'annoncer.

Les énoncés 6.9 à 6.11 sont des énoncés observables qui mettent en relation un des termes suspects avec des faits vérifiables. Tandis que 6.12 est la règle de correspondance entre le terme suspect restant et une autre condition observationelle. Si l'on applique le méthode de Ramsey à cette théorie on obtient:

$$(\exists\Psi)(\exists\Phi)(\forall x)\,(\Psi x \supset (Vx \wedge Px \wedge Ax)) \wedge (\Phi x \supset Cx) \wedge (\Psi x \supset \Phi x)$$

Les termes mentaux et suspects sont éliminés et on trouve à leur place certaines entités qui occupent certains rôles causaux dans la théorie. Ce que ces entités sont en réalité n'est guère important pour le moment. L'important est que chacune d'elles a un rapport causal avec une autre et aussi un rapport avec des termes observables, c'est-à-dire avec les rôles qu'elles occupent.

> Because we understand the O-terms [6], and we can define the T-terms [7] from them, theories are fully meaningful; we have reason to think a good theory true, then whatever exist according to the theory really *does* exist.[...] [Lewis, 1972, page 211, italiques dans le texte original]

Les identifications théoriques ainsi obtenues ne sont pas imposées par la volonté de quiconque, elle sont en revanche le fruit d'inférences logiques. En vertu des définitions proposées, les noms que l'on donne aux termes suspects seront non seulement les occupants des rôles causaux auxquels ils s'appliquent mais aussi leur référents.

[6] Les termes observables.

[7] Les T-terms sont ceux que l'on a appellé les termes suspects en raison d'être des termes théoriques.

Cette démarche est celle que va nous permettre un jour d'inférer que les états mentaux M_1, M_2, ... sont les états neuronaux N_1, N_2, ...

Quel est le problème implicite à cette approche? Lewis utilise la méthode de Ramsey pour procéder à l'élimination des termes théoriques mais cette méthode était avant lui appliquée dans le cadre des théories scientifiques. Lewis l'applique en revanche à la psychologie ordinaire et, ce faisant, il établit que le fonctionnalisme qu'il prône va de pair avec elle. Si une des deux théories s'avère fausse, l'autre aussi doit être récusée. Or, soit les termes mentaux sont des entités existantes (bien que l'on puisse ignorer leur nature) en conséquence de la validité de la psychologie ordinaire, soit leur existence et la validité de la psychologie ordinaire sont toutes deux des mirages.

> Here it is important that, on my version of causal definability, the mental terms stand or fall together. If common–sense psychology fails, all of them are alike denotationless.

Il existe encore un autre problème causé par la vocation de neutralité ontologique du fonctionnalisme du rôle causal. Dans ce cadre, les définitions des termes mentaux se font en fonction de la place qu'ils occupeot dans une théorie empirique. En plus, les définitions de ces termes sont *fonctionnellement implicites*. Les deux caractéristiques que je viens de citer peuvent être considérées comme des problèmes. Oo pourrait argumenter du manque de précision pour faire une caractérisation détaillée de la difficulté de n'avoir que des définitions implicites. Cependant, tout compte fait, elles présentent au moins une vertu puisqu'elles laissent la porte ouverte à des spéculations postérieures lorsqu'on arrive à les préciser.

6.4 Conclusion

J'ai exposé ici les caractéristiques principales du fonctionnalisme de Lewis. Lorsqu'on pose la question : quels sont alors les états mentaux? Les partisans de cette théorie vont vous répondre que ce sont des états fonctionnels. La question suivante est : Ces états fonctionnels sont–ils des états matériels? Pour les fonctionnalistes à la Lewis la réponse sera oui. Lewis et Armstrong pensent que le fonctionnalisme doit amener de l'eau au moulin du physicalisme. Pour eux, si le fonctionnalisme est vrai alors le physicalisme de types l'est aussi. Pour la théorie fonctionnaliste représentationnelle on verra que c'est le contraire. Si le fonctionoalisme est vrai alors ceci va prouver que le physicalisme de types est une théorie fausse. Ces contradictions naissent de ce que l'on a considéré comme une des vertus du fonctionnalisme: la neutralité ontologique. Dans la même ligne d'argumentation, il est difficile de déterminer si le fonctionnalisme des rôles causaux a opéré une réduction des termes mentaux. La réponse à la question est très nuancée. La démarche de Lewis et le processus de Ramsey qu'il applique permettent d'obtenir une définition des états mentaux par des expressions qui ne contiennent aucun terme mental. En définitive on les a éliminés mais seulement d'une certaine manière.

Chapitre 7

Le fonctionnalisme computationnel

"... you'll sing again, and you'll sing it again, and you'll sing it again until you sing it right!"
-punch line of an old opera joke
[Fodor, 1994]

7.1 Introduction

Contrairement à Lewis et à Armstrong pour qui le fonctionnalisme était la preuve de l'identité de types, aux yeux des fonctionnalistes computationnels il constitue justement la réfutation de cette identité. Dans le chapitre précédent on a déjà vu que Lewis et Armstrong identifient les états mentaux aux *choses* dotées des propriétés qui définissent des rôlrs causaux. La stratégie du fonctionnalisme computationnel en revanche, consiste en l'identification non pas aux *choses* qui ont une propriété déterminée mais à la *propriété fonctionnelle* en question. [1]. A la place de l'identité de types ces derniers défendent l'identité occasionnelle des entités mentales avec les entités physiques.

Le fonctionnalisme computationnel connaît deux formes différentes selon le type de paradigme dont il s'inspire.[2] En premier lieu le fonctionnalisme turingien qui se base sur l'bypothèse de l'algorithmicité de la pensé établie par la thèse Turing–Church et ensuite, le fonctionnalisme représentationel ou syntaxique fondé sur l'hypothèse que la pensée peut être considérée comme un système de traitement symbolique.

Dans ce chapitre je vais exposer ces deux approches. Je vais aussi montrer les difficultés que Putnam et Fodor ont rencontrées en voulant individualiser les états mentaux à partir des seules propriétés fonctionnelles, c'est–à–dire sans aucune référence au niveau physique. Je conclurai que le problème de l'isomorphisme fonctionnel auquel une telle démarche aboutit est loin d'être résolu. En outre, ces difficultés mettent en doute non seulement l'hypotbèse de multiréalisabilité de la cognition mais elles suggèrent aussi que le concept de fonction adopté par les fonctionnalistes representationnels s'avère trop libéral pour expliquer l'individualisation des états mentaux.

Ce dernier aspect est central dans la tbèse que je soutieos. En effet, afin de bâtir une solutioo physicaliste non–réductionniste du mental, on a soutenu que l'individualisation des états mentaux pouvait se faire en fonction de leurs rôles fooctionnels et dans le même but on a proposé l'bypothèse de la multiréalisation. En ce faisant, nous avons négligé une source essentielle d'information telle que les résultats des neurosciences, en reléguant ceux–ci au deuxième plan. Ce chapitre essaie de

[1] J'ai déjà signalé dans le chapitre précédent que le concept de *fonction* diffère dans les deux théories. Pour Armstrong et Lewis la fonction se base sur les dispositions du comportement tandis que selon les fonctionnalistes computationnels elle relève d'une analyse fonctionnelle ou téléologique

[2] Pour la dénomination des paradigmes je suis la classification suggérée dans [Copeland, 1994]. Je me réfèrerai à l'*hypothèse de l'algorithmicité de la pensée* (ma traduction de emphthe algorithmicity assumptionet à l'*hypothèse du système de traitement symbolique* (ma traduction de *symbol system hypothesis*).

souligner les limites de ces approches, particulièrement le fonctionnalisme turingien et la théorie proposée par Jerry Fodor.

7.2 Le fonctionnalisme turingien

Les fonctionnalistes computationnels récusaient l'identité de types tout en restant physicalistes. Ils avaient donc besoin de trouver une méthode d'individualisation des états fonctionnels sans faire appel au domaine physique. [3] Il s'agissait donc de trouver un méthode abstraite de spécification des états mentaux mais le risque était de tomber dans des explications pseudo–fonctionnelles. Bref, il fallait trouver un vocabulaire canonique constituant une grille d'explication compatible avec le physicalisme sans y appartenir.

Voilà pourquoi Putnam s'est tourné vers la Machine ou Automate de Turing. Cette formalisation semblait fournir le vocabulaire canonique que l'on cherchait. En ce faisant, Putnam établit implicitement une identité entre les fonctions mathématiques et les fonctions téléologiques telles que ces dernières sont conçues en biologie.

7.2.1 La machine de Turing

Alan Mathison Turing (1912–1954) était un mathématicien et logicien anglais qui imagina aux alentours de 1936 les machines ou automates qui portent son nom et qui se sont révélés être d'excellents modèles abstraits des ordinateurs réalisés à partir de 1946. Turing conçoit la machine qui porte son nom pour répondre à une des questions posées par David Hilbert (1862–1943) sur la solubilité des problèmes d'une classe donnée, cette question étant : est–il possible de trouver un processus mécanique (par exemple un algorithme) pour résoudre tous les problèmes mathématiques issus d'une classe donnée? La réponse de Turing était négative mais pour la justifier il eut recours à cette abstraction mathématique connue sous le nom de machine de Turing.

Il est utile de rappeler le cadre historique de l'évolution des mathématiques à cette époque. David Hilbert avait proposé un programme de démonstration basé sur une position philosophique : le *formalisme*. Selon ce credo formaliste, les mathématiques doivent être analysées comme une activité sans signification, semblable à un jeu. Elles consistent en des règles formelles fixées à l'avance et permettant de construire certains assemblages de symboles, à savoir les énoncés mathématiques et leurs démonstrations.

L'élément essentiel de la pensée était le *mécanisme*. Au contraire des *intuitionnistes* qui réclamaient pour le mathématicien le rôle de sujet pensant, le formalisme le réduit à la dimension d'un robot. L'intuition n'est que ce qui permet de compenser en partie notre infériorité par rapport aux vraies machines.

Néanmoins, l'ontologie de Hilbert n'est pas vide; elle est habitée par certains objets, certaines propriétés et certaines démonstrations qui ont du sens et qui existent. Il s'agit de constructions finies sur des nombres naturels qu'Hilbert appelle objets ou propriétés *élémentaires* et qui ont la forme:

$$\forall x_1 \ldots \forall x_n, f(x_1, \ldots, x_n) = 0$$

où f est une fonction calculable, c'est–à–dire récursive.[4]

Les démonstrations élémentaires sont celles qu'utilisent les énoncés élémentaires et des principes particulièrement immédiats comme l'induction ou la récurrence sur des données élémentaires. Jean–Yves Girard caractérise l'ontologie hilbertienne de la façon suivante :

> Tout ce qui ne fait pas partie de l'élémentarité, Hilbert l'appelle *obstrait* (qu'il s'agisse d'objets, de propriétés ou de prédicats): pour lui ces objets idéaux ne sont que des abstractions commodes permettant à la mauvaise machine qu'est le mathématicien de fonctionner plus efficacement; mais ces objets n'ont pas de sens en eux–mêmes. Le programme de Hilbert est une tentative pour *démontrer*

[3] Rappelons nous qu'une des conditions fondamentales pour qu'une entité puisse revendiquer le statut d'existence ontologique est qu'il existe un moyen d'identification, selon une des maximes de Quine les plus respectées.

[4] Je reviendrai sur la définition de récursivité plus loin dans ce chapitre, pour le moment, prenons comme définition provisoire récursif = calculable.

la justesse de l'ontologie hilbertienne: *Étant donné une démonstration D d'un énoncé élémentaire R par des méthodes abstraites (par exemple, à l'aide de l'axiome du choix en théorie des ensembles)* montrer que *R* peut être obtenu directement dans les mathématiques élémentaires. [Girard, 1985, page 1102]

Le programme d'Hilbert a été réfuté par les théorèmes de limitations que je présenterai dans la prochaine section. Pour raisonner sur la question d'Hilbert, Turing utilise deux outils mathématiques utilisés auparavant par Georg Cantor (1845–1918) et par Kurt Gödel (1906–1976). L'outil de Georg Cantor est l'argumentatioo diagonale pour démontrer l'existence des nombres irrationnels. La stratégie consistait à montrer qu'aucune suite oe pouvait comprendre tous les nombres réels. Chaque suite possible ne pouvait servir qu'à définir un autre nombre à une infinité de décimales qui avait été oublié. Par ailleurs, Cantor démontra ainsi qu'il existe plus de nombres réels que de nombres rationnels.

L'autre outil est le codage de Gödel qui permet à Turing de formaliser tous les énoncés bien formés (*ebf*) dans une théorie axiomatique comme des numéros et en ce faisant d'établir des suites. Le codage de Gödel consiste à définir une application $x \mapsto \ulcorner x \urcorner$ de l'ensemble des symboles, des formules et des suites de formules dans des nombres naturels (N). Cette application doit:

1. être injective;
2. capable de calculer effectivement $\ulcorner x \urcorner$, étant donné x;
3. étant dooné un entier n, établir effectivement s'il existe un x tel que $n = \ulcorner x \urcorner$ et, dans ce cas, déterminer effectivement cet x.

Le nombre $\ulcorner x \urcorner$ est appelé le nombre de Gödel (*n.g.*) de x. Il y a différentes façon de définir ce codage; l'une d'elles est la suivante: Soit L l'alphabet à partir duquel les énoncés sont formés. L'opération a deux étapes:

1. Attribution d'un nombre de Gödel à chaque symbole de l'alpbabet de L.

   Il s'agit de définir une applicatioo g qui à chaque symbole de l'alphabet de L attribue un nombre univoquement détermioé:

   $$g : L \to \mathbf{N} \tag{7.1}$$

2. Attribution d'un nombre de Gödel à chaque *ebf* de L.

   Toute *ebf* se présente comme la concaténation de symboles s_i de l'alphabet de L. Soit $A = s_1 \ldots s_k$. Il s'agit de tenir compte des deux informations suivantes: 1.- Chaque symbole possède un place bien déterminé dans une *ebf*. 2.- Indépendamment de la place qu'il occupe, chaque symbole possède un nombre de Gödel univoquement déterminé.

   $$s_1 \mapsto g(s_1) \tag{7.2}$$
   $$\vdots \tag{7.3}$$
   $$s_k \mapsto g(s_k) \tag{7.4}$$

Il est évident qu'à l'aide des informations 7.1 et 7.4 il est possible de déterminer le nombre de Gödel d'une expressioo. [Miéville, 1991, cf. page 37–38]

Muni de ces deux outils Turing a transformé la questioo de Hilbert en la dotant d'une perspective plus opérationnelle. Pour cela il a eu l'idée de concevoir une machine idéale capable de décider, pour toute assertion qui lui serait présentée, si elle était démontrable ou non. Une fois que la machine abstraite est proposée, Turing paraphrase la question de Hilbert comme suit: Est-il possible de décider si une machine de Turing donnée qui agit sur des données spécifiques va s'arrêter ou non? Turing, en utilisant la méthode diagonale de Cantor montre qu'il n'existe aucun algorithme qui puisse répondre à cette question de façon systématique.

> Il avait démontré que les mathématiques ne pourraient jamais être épuisées par un ensemble fini de procédures diverses. Il était allé au coeur même du problème et l'avait réglé par une constatation, à la fois simple et élégante. [Hodges, 1988, page 96]

Pour arriver à ce résultat Turing avait conçu en plus l'idée de la machine de Turing universelle, cette dernière pouvant simuler n'importe quelle autre machine de Turing. Voilà pour la genèse de la Machine de Turing.

Automate fini Je vais maintenant définir ladite machine mais auparavant il est nécessaire de présenter le concept d'automate à nombre fini d'états. Un automate à nombre fini d'états:

$$S = \{S_i | 0 \leq i \leq n\}$$

possède une *tête de lecture* devant laquelle défile un ruban divisé en cellules, chacune des cellules pouvant contenir un symbole, c'est-à-dire un des symboles appartenant à un langage; l'automate est muni d'un nombre de règles du type:

$$M_i, S \rightarrow S_k$$

Cette formule doit se lire: "Si devant la tête de lecture il y a le symbole M_i et l'automate est dans l'état S alors il passe à l'état S_k".

Un automate de Turing est composé d'un automate à nombre fini d'états, mais il est doté d'une tête de lecture–écriture devant laquelle se trouve un ruban divisé en cellules et qui peut se déplacer de droite à gauche et de gauche à droite. Cet automate est muni d'un nombre fini de règles qui peuvent être de trois types:

$S_iQ_jP_mS_k$ signifie que l'automate est dans l'état S_i et si la cellule située sous la tête de lecture–écriture contient Q_j l'automate remplace Q_j par P_m et passe dans l'état S_k.

$S_iQ_jDS_k$ dans les même conditions, on déplace le ruban d'une cellule vers la droite et l'automate passe à l'état S_k.

$S_iQ_jGS_k$ dans ce cas, on déplace le ruban d'un rang vers la gauche et l'automate passe à l'état S_k.

La machine universelle (U), rappelons-le, est capable de simuler n'importe quelle autre machine de Turing. Comme les machines de Turing sont des énoncés bien formés (*ebf*) on peut donc appliquer la numérotation de Gödel, ainsi que celle auquel correspond le nombre n, notée T_n. Pour que la machine U agisse comme la machine T_n, tout ce dont on a besoin est de donner comme entrée dans la première cellule de la table le nombre n et ensuite (à droite d'un symbole qui code un marqueur) on peut donner l'entrée que la machine T_n est supposée lire. Le résultat d'une action de U sur toute la bande s'avère identique à celui obtenu par T_n dans la partie droite de la bande.

La thèse Turing–Church

Turing a montré que les fonctions calculables au moyen de cette machine sont exactement les fonctions récursives et il en déduit l'hypothèse que toute fonction calculable par quelque machine que l'on puisse imaginer est calculable par sa machine. Mais qu'est-ce qu'une *fonction calculable* et qu'est-ce qu'une fonction récursive?

La notion de l'effectivement ou du mécaniquement calculable est d'ordre intuitif et méta-mathématique alors qu'en revanche le concept de fonction récursive se veut une formalisation de celle-ci.

L'ensemble des fonctions calculables a été identifié avec divers formalismes dans le but de rendre formelle cette notion intuitive. Ainsi elles ont été identifiées aux ensembles suivants:

- L'ensemble des fonctions récursives. La définition de fonction récursive due à Jacques Herbrant (1908–1931) et aussi à Kurt Gödel.

- Alonzo Church (1903–) propose d'identifier l'idée intuitive de fonctions effectivement calculables aux fonctions λ–définissables. Ces dernières furent définies conjointement par Church et Stephen Kleene (1909–). Ainsi Alonzo Church propose une thèse qui porte son nom:

  THÉORÈME 1 (THÈSE DE CHURCH)
  Toute fonction effectivement calculable est récursive ou λ–définissable.

- En 1936 Turing identifie les fonctions calculables avec celles que l'on a calculées avec la Machine de Turing.

- Andrë Andreïevitch Markov (1856–1922) identifie les fonctions calculables en utilisant la théorie des algorithmes.

La question est de voir si tous ces ensembles ainsi définis sont co–référentiels. Miéville nous donne la chronique des résultats

> **En 1936, Kleene fournit une première réponse: l'ensemble de fonctions calculables par λ– définissabilité est équivalent à celui des fonctions récursives. L'année suivante, Turing démontre que toute fonction λ–définissable est calculable par une machine de Turing et que toute fonction calculable par une telle machine est une fonction récursive. Enfin, V. Detlovs (1953) a établi l'équivalence entre l'ensemble des fonctions récursives et celui de fonctions calculables à l'aide des algorithmes de Markov. Ainsi quatre ensembles ont la même extension et on n'a pas trouvé de définition qui fournisse une classe plus étendue.**
>
> **On peut donc énoncer la thèse généralisée de Church:**
>
> THÉORÈME 2 (THÈSE GÉNÉRALISÉE DE CHURCH)
> ***Une fonction est effectivement calculable si c'est:***
>
> - ***une fonction récursive, ou***
> - ***une fonction λ–définissable ou***
> - ***une fonction calculable par la machine de Turing, ou***
> - ***une fonction calculable par un algorithme de Markov. [Miéville, 1991, page 91–92]***

Cette équivalence entre les quatre ensembles et l'arithmétisation de Gödel appliquée à ces ensembles permet aux résultats de la théorie des fonctions récursives d'être appliqués aussi aux autres ensembles. L'inscription des fonctions récursives au sein d'un système formel permet d'établir des thèses métalogiques fondamentales.

Par exemple supposons que l'on cherche à savoir si un énoncé bien formé (*ebf*) qui peut avoir la forme d'une fonction calculable par une machine de Turing, ou par un algorithme de Markov ou définissable comme une λ-fonction, appartient ou non à l'ensemble des théorèmes ou même, en allant plus loin, étant donné une suite d'*ebf*, on veut savoir si elle appartient à l'ensemble des preuves. Grâce à l'équivalence des quatre ensembles et à l'arithmétisation de Gödel, on peut procéder de la manière suivante:

1. identifier l'ensemble des formules bien formées pour chacun des formalismes cités ci–dessus avec des ensembles récursifs,

2. relations binaires qui sont exprimées par la fonction propositionnelle du type "v est un nombre de la preuve de n" avec des relations récursives.

On peut transposer les deux questions sur les *ebf* en termes d'appartenance ou de non–appartenance d'un nombre à un ensemble de nombres. Ainsi les deux questions précédentes deviennent équivalentes à la question: une *ebf* (ou une suite d'*ebf* arithmétiquement codée), appartient–elle à tel ou tel ensemble défini par une propriété ou une relation, elle aussi arithmétiquement codées? De cette façon on peut faire des calculs sur l'image de la fonction mais pour répondre à la question posée il est nécessaire que la procédure utilisée soit récursive. [Miéville, 1991, cf. page 103]

Je vais définir cette notion de fonction récursive.

DEFINITION 4 (RELATION PRIMITIVE RÉCURSIVE (RPR))
Soit R, une relation n-aire: $R(x_1, \ldots, x_n)$ et soit $f_R : (x_1, \ldots, x_n) \mapsto \{0, 1\}$ la fonction caractéristique suivante :

$$f_R(x_1, \ldots, x_n) = \begin{cases} 1 & \text{si } x_1, \ldots, x_n \text{ ont entre eux la relation } R \\ 0 & \text{n'ont pas entre eux la relation } R \end{cases} \tag{7.5}$$

Une relation n-aire est primitive récursive [RPR] si et seulement si sa fonction caractéristique associée est une fonction récursive primitive (FRP).

Définition des fonctions récursives: Les fonctions récursives ont été introduites par Julius Wilhelm Richard Dedekind (1831–1916) en 1888. On appelle *fonction récursive primitive* le plus petit ensemble de fonctions de $\mathbf{N}^n$ dans $\mathbf{N}$ qui satisfait aux deux exigences suivantes :

1. Il contient la fonction $S(x)$, *successeur de*, la fonction *zéro(x)* qui prend la valeur 0 quels que soient ses arguments et les *fonctions d'identification* $Id_i^n(x_1, \ldots, x_n) = x_i$ qui prennent pour valeurs l'un de leurs arguments.

2. Il est fermé[5] sur l'opération de composition de fonctions et sur l'opération de récursion primitive, opération décrite par le schéma de récursion primitive.[6]

   On appelle *fonction récursive générale* les fonctions générées comme les précédentes, mais avec l'aide d'une sur l'opération supplémentaire, à savoir la *minimisation* qui produit à $n-1$ places à partir d'une fonction à n places. [Gochet and Gribomont, 1990, cf. pages 83–84, Vol I]

Bien que l'on puisse démontrer que l'ensemble des fonctions récursives primitives a une extension plus petite que les fonctions récursives générales, on ne saurait dire si l'ensemble de ces dernières réunit toutes les fonctions calculables[7]. Plusieurs résultats se montrent convergents mais il est impossible de les justifier par une démonstration véritable. [Miéville, 1991, cf. page 113 Vol. II]

Thèse Turing–Church De cette supposition sur l'identification de l'ensemble des fonctions calculables avec l'ensemble des fonctions récursives et l'ensemble des fonctions calculables au moyen d'une machine de Turing on peut formuler l'hypothèse de l'algorithmicité de la pensée comme suit :

> Combinées, les thèses de Church et de Turing sont donc exprimées par la double égalité : récursif = calculable par l'homme = calculable par (tout) ordinateur. [Andler, 1985, page 192]

Je désignerai désormais cette relation d'equivalence sous le nom de thèse de Turing–Church.

Lors de l'introduction des fonctions récursives nous avons signalé, ne l'oublions pas, que leur importance réside dans le fait que leur inscription au sein d'un système formel permet d'établir des thèse métalogiques fondamentales.

Soit le langage de premier ordre L_{ar} celui qui est couramment utilisé pour l'arithmétique; il comprend un symbole de relation binaire $\approx$, deux symboles d'opérations binaires $+$ et $\times$ et deux symboles constants 0 et 1. Si l'on ajoute en plus l'opération unitaire s, alors toute mention d'une formule, d'une théorie d'une structure sous-entend la référence à ce langage.

---

[5] Une opération est fermée sur un ensemble lorsque les résultats de cette opération sont tous à l'intérieur de l'ensemble considéré.

[6] Le *schéma de récursion primitive* est le suivant : soient f, g et h des symboles des fonction et S la fonction *successeur* :

a $\forall x_1, \ldots, x_n \in N_0, h(x_1, \ldots, x_n) = f((x_1, \ldots, x_n)$,

b $\forall x_1, \ldots, x_n, y \in N_0, S(x_1, \ldots, x_n, S(y)) = g(x_1, \ldots, x_n, y, h(x_1, \ldots, x_n, y))$.

Ce schema formalise la notion de calcul fait pas à pas.

[7] *calculable* dans le sens large du terme.

DEFINITION 5 (LE MODÈLE STANDARD DE L'ARITHMÉTIQUE)
Le modèle standard de l'arithmétique est la structure $\mathfrak{n}$ *d'univers* $\mathbf{N}$ *munie de l'addition et de la multiplication usuelles, de la fonction successeur* $\bar{s}^{\mathfrak{n}} = suc : n \mapsto n + 1$ *et des éléments distingués* $\bar{0}^{\mathfrak{n}} = 0$ *et* $\bar{1}^{\mathfrak{n}}$.

Aussi, toute fonction récursive est arithmétique.

Théorèmes de limitation Le programme d'Hilbert était basé sur la possibilité de rendre toute démonstration réalisable directement par les mathématiques élémentaires.

Or ce programme a échoué et j'énumérerai quelques résultats que l'on connait, en général, en tant que théorèmes de limitation pour discuter à la fin la portée de l'un d'eux sur les théories computationnelles de la pensée: le premier théorème d'incomplétude de Gödel.

Ces théorèmes visent à réfuter l'hypothèse hilbertienne sur l'aritbmétique formelle. Ils peuvent être formulés de manière approximative comme suit:

T. de Church–Rosser Toute théorie $\mathfrak{n}$ est un modèle indécidable.

T. de Tarski L'ensemble des énoncés vrais dans $\mathfrak{n}$ est indéfinissable arithmétiquement.

Premier théorème de l'incomplétude de Gödel Toute théorie axiomatisable [6] même rudimentaire est incomplète [9].

Second théorème de l'incomplétude de Gödel Toute théorie axiomatisable suffisamment développée ne permet pas de prouver sa propre cohérence. [Andler, 1985, cf. page 193]

Tous ces théorèmes montrent les limites des capacités démonstratives et descriptives des systèmes formels, quoiqu'ils fassent référence à l'arithmétique. Il montrent aussi la nécessité pour les symboles d'avoir une signification; on ne peut pas réduire la tbéorie de la démonstration à une simple manipulation syntaxique des symboles. Or, le formalisme ne peut plus être tenu pour la fondation des mathématiques.

Le premier théorème de l'incomplétude de Gödel dépasse le domaine mathématique et il est souvent invoqué par les adversaires de la thèse mécaniste de la pensée pour arguer que l'esprit humain ne fonctionne pas comme une machine de Turing.

Les arguments contre l'approche computationnelle de la pensée Les réfutation de l'approche computationnelle de la pensée, de même que la démarche contraire, celle qui soutient la validité de la thèse de Church–Turing souffrent toutes les deux d'un handicap majeur. Il y a un *hiatus philosophique* comme le note John Searle. Pour lui l'biatus philosophique signale la division existante entre une théorie des fonctions récursives très élégante et enrichie par les théorèmes de Turing et Church, et d'autre part les engins électroniques de notre époque. [Searle, 1994h, cf. 205]

A mon avis, le hiatus se produit entre cette théorie mathématique et la conception même de la cognition. Lorsqu'on essaie d'exposer la portée de la thèse Turing–Church, quelles capacités cognitives désigne–t–on chez les hommes? Faut–il prendre en compte par exemple les activités dite conscientes? Faut–il prendre en compte la possibilité d'élaborer des métareprésentations des agents humains? A partir de quel lien conceptuel peut–on faire l'extrapolation d'une théorie mathématique à la cognition?

En général, les argumentations aussi bien dans un sens que dans l'autre ne répondent pas à cette question.

La présence du programme de Hilbert dans le contexte historique semble y être pour beaucoup.

[8] Une théorie est *axiomatisable*, si elle admet un ensemble décidable d'axiomes

[9] Une théorie est *complète* si elle permet de prouver ou de réfuter tout énoncé. Plus strictement et pour le calcul propositionnel un système est *complet* si:

1. Toute formule prouvable dans le système est tautologique.
2. Toute formule tautologique est prouvable.

Les partisans de la pertinence du modèle de Turing pour la cognition gardent une certaine nostalgie du pouvoir que le formalisme hilbertien attribuait aux systèmes formels.

Les opposants par contre, mettent l'accent sur les théorèmes de limitation qui attestent l'échec du formalisme hilbertien, notamment le premier théorème de Gödel. Ces résultats démontreraient que la tâche du mathématicien ne peut pas être identique à celle d'un robot comme Hilbert voulait le démontrer. En effet, il existerait au moins un aspect de la cognition, l'activité du mathématicien où l'homme surpasserait les possibilités d'un mécanisme automatique tel que celui de Turing. La question est de savoir tout d'abord si cette thèse est défendable et ensuite, si c'est bien le cas, quel est l'aspect de l'activité du mathématicien qui ne peut pas être assumé par la machine de Turing.

Jacques Paul Dubucs dans un intéressant article fait une étude très fine des objections à la thèse mécaniste de la pensée découlant du premier théorème de l'incomplétude de Gödel. Dubucs donne l'énoncé suivant du premier théorème d'incomplétude Gödel :

> Tout système formel S suffisamment riche[10] contient, s'il est cohérent[11], un énoncé élémeotaire G_S improuvable mais vrai (« élémentaire » signifie ici : $\forall x : \Phi x, \Phi$ récursif). [Dubucs, 1992, page 73]

Selon Dubucs les arguments qui visent à établir la supériorité de l'esprit humain sur les machines reposent sur deux types différents, bien que complémentaires d'erreurs. Soit ils surestiment les capacités de la cognition, soit ils sous-estiment les capacités de la machine.

Avant de continuer à citer les objections proprement dites, rappelons comment l'on utilise l'arithmétisation de Gödel pour les preuves dans un système formel donné.

On assigne à chaque axiome ou à chaque formule obtenu à partir d'une ou plusieurs formules antérieures par l'application d'une règle d'inférence un nombre de Gödel, comme nous l'avons déjà expliqué dans des section précédentes. Maintenant que les formules, les axiomes et les règles sont codés, il s'avère qu'une preuve dévient une séquence de nombres.

Chacune des suites ainsi obtenue et qui constitue une preuve est aussi codée par un nombre. Une fois que l'on a procédé ainsi, on peut énoncer le prédicat suivant "$\ulcorner y \urcorner$ est le nombre gödelien de la preuve y" et à partir de lui le prédicat binaire suivant "x qui est le nombre gödelien d'une démonstration est la preuve de la formule y" qu'en général on note $Pr_S(x, \ulcorner y \urcorner)$. [12]

Mais la thèse du théorème : "G_S est improbable" et qui est représentée par le prédicat $\Phi(x)$ peut être exprimée comme $\neg Pr_S(x, \ulcorner G_S \urcorner)$.

Alors $\Phi(x)$ est la formule qui peut être énoncée comme : "x n'est pas le numéro d'une preuve dans S de la formule G_S".

Mettons en clair ce que signifie "G_S est improuvable". Cela veut dire que tout entier n qui ne satisfait pas Φ est bien le numéro de la preuve de G_S.

Ce qui coovainc un agent humain de la vérité de la formule G_S est le raisonnement *modus tollens* suivant : Supposons que G_S soit faux, alors en vertu de la richesse de S [13] il existera un

---

[10] L'expression *suffisamment riche* signifie que le système formel est de force au moins égale à R où R est la théorie qui comporte les axiomes suivants:

- $\forall x \forall y(sx \approx sy \rightarrow x \approx y)$
- $\forall s \neg(sx \approx 0)$
- $\forall(\neg(x \approx 0) \rightarrow \exists y(x \approx sy)$
- $\forall x(x + 0 \approx x)$
- $\forall x \forall y(x + sy \approx s(x + y))$
- $\forall x(x \times 0 \approx 0)$
- $\forall x \forall y(x \times sy \approx (x \times y) + x)$

L'observation que l'on peut faire est que les lettres x et y désignent des variables formelles. En plus lorsqu'on observe le système on voit qu'il est plus faible que n puisqu'il ne comprend aucune forme de récurrence et que la commutativité ne peut être dérivée de ces axiomes.

[11] Un système S est *cohérent* s'il n'existe pas de formule F telle que $\vdash_S F$ et $\vdash_S \neg F$.

[12] Où $\ulcorner y \urcorner$ représente rappelons-nous le nombre de Gödel de y

[13] Les relations récursives que le système S est capable de représenter nous permettent d'affirmer que si Φ est un prédicat récursif, alors pour chaque entier n énoncer $\Phi(x)$ est prouvable si n satisfait Φ et réfutable a contrario.

entier n tel que $\neg\Phi(n)$ soit prouvable, ce qui réfutera G_S, mais étant donné que tout entier n qui ne satisfait pas Φ est bien le numéro de la preuve de G_S, alors G_S serait prouvable, puisqu'il existe de toute façon un nombre n qui est bien une preuve de G_S. D'où il s'ensuit que S serait incohérent. Donc G_S est vraie.

Est-ce que la performance dont l'agent humain vient de faire preuve est une réfutation de l'hypothèse mécaniste de la pensée? Dubucs relativise cet argument en disant:

> **L'antimécaniste soutient que la performance cognitive dont je viens de faire preuve n'est pas à la portée de la machine $T(S)$. Il a tort. Car ou bien il pense que je viens de prouver que G_S est vrai, et alors il sur-estime cette performance: j'ai simplement montré que *si* S est cohérent, alors G_S est vrai. Ou bien il reconnaît que j'ai seulement prouvé que si S est cohérent, alors G_S est vrai, et dans ce cas il sous-estime les capacités de $T(S)$, puisque l'énoncé qui affirme que la cohérence de S entraîne la vérité de G_S *est* un théorème de S.** [Dubucs, 1992, page 74]

Ainsi, selon Dubucs l'argument anti-mécanique n'est pas valable parce qu'il s'avère absurde en son raisonnement.

Cependant on peut être en désaccord avec Dubucs. On peut soutenir que quoi qu'il en soit le raisonnement précédent a pour conséquence la vérité de G_S, en faisant valoir la cohérence de S. Nous pouvons dire que les caparités sémantiques bumaines *sont au rendez-vous* dans ce raisonnement. En effet, on peut faire la part des choses entre la formule du théorème "Si S cohérent alors G_S est vrai" et la conséquence que l'on donne pour vraie, puisqu'elle est ioférée de façon valide si l'on accepte la méthode de *modus tollens*.

Néanmoins, cette argumentation a ses limites lorsqu'il s'agit de réfuter la thèse mécaniste de la pensée. En effet, elle ne prouve pas que je ne suis pas un machine, puisque la marbine $\tilde{S} = S \bigcup Coh(S)$ serait en mesure de prouver G_S. Or, l'antimécaniste a du mal à soutenir l'argument que l'agent humain est supérieur à toutes les machines. [14]

L'argument de la supériorité des agents humaios face à $T(S)$ échoue face à $T(\tilde{S})$. De ce fait l'argumentation de réfutation prendra le format suivant:

> La machine de Turing T ne peut pas, tout en restant cohérente, énumérer toutes les vérités arithmétiques que je suis capable de justifier. [Dubucs, 1992, cf. page 76]

Peut-on dire, donc que $T(\tilde{S})$ reproduit mes capacités? Selon Dubucs, et je suis de son avis, la réponse est non. Bien que la quantité des formules que peut prouver $T(\tilde{S})$ soit égale au nombre de celles que je puis démontrer, j'ajoute néanmoins qu'il existe une différence qualitative. J'arrive à prouver que G_S guidée par la référence *attendue* aux symboles que je traite, "c'est-à-dire capable d'invoquer le fait que certains de leurs regroupements reflètent éventuellement les propriétés de la structure mathématique que *je vise*" (cf. page 78) et il continue:

> **[...] l'aptitude de $T(\tilde{S})$ à inscrire G_S sur son ruban de sortie est le pur résultat de son architecture combinatoire $T(\tilde{S})$, et c'est là différence cruciale d'avec moi, est *intentionnellement inerte*.** [Dubucs, 1992, cf. page 76]

Revenons sur les aspects cognitifs de l'agent humain qui s'avèrent être différents de ceux que possède la machine $T(S)$.

Dubucs catégorise ces aspects eo deux types selon la stratégie que l'agent humain adopte pour obtenir la vérité de G_S.

1. Raisonnement visant la vérité de G_S

   Consiste à prendre un à un les énoncés $\Phi(n)$ et conclure que $\forall x : \Phi x$. C'est justement le passage que j'opère de la vérification de tous les n à la généralisation à tous les éléments de l'univers qui me permet d'ajouter la formule $\forall x : \Phi x$.

2. Raisonnement visant la cohérence de S

   Consiste à prouver par induction sur la longueur des preuves la cohérence se S, et en ayant prouvé ceci à déduire la vérité de G_S.

[14] L'argument de ce type dirait qu' "Il existe un énoncé arithmétique que je suis capable de justifier, mais qu'aucune machine de Turing cohérente ne peut reconnaître."

La première stratégie n'est pas véritablement une objection à l'hypothèse mécaniste de l'esprit. D'abord, l'évaluation de chaque instance de $\Phi(n)$ pour tous les n est une tâche mécanisable. En second lieu, il est empiriquement impossible pour un agent humain, même en lui concédant un temps illimité de vérifier un ensemble innombrable d'éléments $\Phi(x)$. Bien que le passage à la généralisation universelle soit la prérogative des seules capacités humaines, les conditions requises pour son application ne seront jamais obtenues.

La deuxième démarche constitue une véritable objection.

> Elle consiste à vérifier que les axiomes de S sont satisfaits dans le modèle attendu (donc que les énoncés dont la preuve dans S est de longueur 1 sont vrais dans ce modèle) et que les règles d'inférence de S préservent la vérité. Dans ces conditions tous les énoncés prouvables dans S sont vrais dans le modèle attendu, $\bar{0} = \bar{1}$ n'est pas prouvable dans S, $Coh(S)$ est donc vrai, et l'on peut en déduire G_S. Un raisonnement de ce type, qui ne comporte aucun processus inférentiel « actuellement infini», est visiblement capable de conduire à une classe de réponses *effectivement discriminable* de celles de la machine $T(S)$ dans le test construit [...] précédent[ement]. La seule étape non- mécanisable est la première, dans laquelle est évaluée la correction des axiomes de S dans le modèle visé : consistant à accepter ou à rejeter une classe d'énoncés *selon leur référence* (VRAI ou FAUX) dans l'interprétation attendue, elle n'est pas plus effectuable par une machine que ne le serait une instruction du genre : « Si le symbole observé est une réalisation du type T, faire A; sinon faire B ». [Dubucs, 1992, page 83–84]

De tout ce que l'on vient d'exposer, on voit que l'argumentation du premier théorème d'incomplétude de Gödel contre l'hypothèse mécaniste de l'esprit est loin d'avoir été bien utilisée dans la plupart des cas. La version du théorème de Gödel dont nous nous sommes servis jusqu'ici est la version *sémantique* selon laquelle l' énoncé G_S, bien qu'improuvable est vrai.

Il y a une autre version, dite *syntaxique* qui énonce ainsi son théorème :

> Chaque système formel suffisamment riche contient, s'il est cohérent, un énoncé indécidable.

Pour certains auteurs comme Shanker et aussi Rosser l'énoncé pertinent est celui que je viens de donner. Rosser l'a exprimé en 1936 comme suit :

> Pour toute théorie arithmétique, S, axiomatisable, cohérente et de force au moins égale à N, il existe un énoncé ϕ tel que ni ϕ ni $\neg\phi$ ne sont prouvables dans S.

En général, ceux qui défendent l'énoncé syntaxique disent que le défaut majeur de l'énoncé sémantique est qu'il externalise la connexion entre une proposition mathématique et sa preuve car il est incohérent d'affirmer que G_S est à la fois vraie et improuvable.

Cette conception syntaxique du théorème de Gödel permet de réfuter le dernier argument contre la conception mécaniste de la pensée. Cet argument peut être énoncé comme suit : la machine de Turing ne peut pas être à la fois cohérente et capable d'énumérer tous les énoncés qu'un agent *qui vise le modèle standard de l'arithmétique* (c'est-à-dire n) est capable de reconnaître comme vrais.

Selon la conception syntaxique cet argument se heurte à un certain scepticisme inspiré par Wittgenstein sur l'existence d'une manifestation qui puisse valider la reconnaissance faite à partir du modèle standard. Selon ce scepticisme il serait nécessaire de manifester cette attitude de reconnaissance de façon publique, c'est-à-dire être en mesure de produire une classe d'énoncés assez riche pour pouvoir déterminer sans ambiguïté les propriétés du modèle standard. Cependant, cette possibilité est réfutée par le premier théorème de Gödel qui démontre précisément qu'il est impossible de définir (à un isomorphisme près) ce modèle à l'aide d'une classe récursivement axiomatisable d'énoncés (du premier ordre). En effet, comme l'ensemble d'axiomes dont on part constitue la théorie standard cet ensemble s'avère être récursif; néanmoins l'ensemble des théorèmes dérivables de ces axiomes n'est pas lui même récursif mais plutôt récursivement énumérable. Ainsi les notions d'entier ou de nombre fini ne sont pas complètement caractérisables à l'aide d'une classe récursivement énumérable d'énoncés.

Je pense, avec Dubucs que ceci n'est pas un vraie objection.

En d'autres termes, bien qu'il nous soit impossible d'obtenir, par le biais d'accord sur une classe « présentable » d'énoncés, une *garantie publique* de l'identité de nos notions « privées » d'« entier» ou de « nombre fini », nous avons de bonnes raisons de supposer que nous venons à nous entendre lorsque nous disons d'une preuve dans un système formel qu'elle est un objet « fini » (même si nous ne spécifions pas de borne supérieure à sa longueur). Or si une telle convergence est possible, qui ne se résume ni à une disposition communément partagée à asserter les énoncés constitutifs d'un certain corpus ni, a *fortiori*, à une propension commune à entretenir avec certaines *occurrences* de ces énoncés une relation matérielle déterminée, c'est qu'il existe au moins une notion dont la psychologie computationnaliste est incapable d'expliquer la maîtrise.[Dubucs, 1992, page 86-87]

Résumons: l'objection à la thèse mécaniste de la pensée doit se baser sur la supériorité cognitive de l'agent humain à pouvoir développer une stratégie d'induction sur la longueur des formules. Néanmoins cet argument peut être mis en doute à partir d'une perspective de scepticisme wittgensteinien.

7.2.2 Le fonctionnalisme de Putnam dans les années soixante

Hilary Putnam croyait trouver dans le formalisme de la machine de Turing le vocabulaire canonique que les fonctionnalistes computationnels cherchaient. Les motivations pour ce choix sont multiples.

Premièrement, la thèse Turing–Church identifie les possibilités de calcul des agents humains avec cette machine abstraite.

Deuxièmement, l'indépendance entre la description logique de la machine de Turing et le niveau physique de sa réalisation et que l'on appelle le *principe de multiréalisabilité* de la cognition. Le modèle basé sur la machine de Turing permet deux types de descriptions, la description physique à laquelle Putnam fait référence comme étant le point de vue de l'ingénieur ou point de vue *structurel* et la description qui fait référence seulement aux tables des états de la machine, le point de vue du logicien [15]. Ainsi, poser la machine de Turing comme modèle de la cognition permet de conserver les avantages du behaviorisme et de la psychologie classique tout en évitant leurs difficultés.

It is interesting to note that just as there are two possible descriptions of the behaviour of a Turing machine – the engineer's structural blueprint and the logician's 'machine table'– so there are two possible descriptions of human psychology, The 'behavioristic' approach (including in this category theories which employ 'hypothetical constructe', including 'constructs' taken from physiology) aims at eventually providing a complete physicalistic description of human behavior, in terms which link up with chemistry and physics. This corresponds to the engineer's or physicist's description of a realised Turing machine. But it would also be possible to seek a more abstract description of human mental processes, in terms of 'mental states' (physical realization, if any, unspecified) and 'impressions' (they play the role of symbols on the machine's tapes) – a description which would specify the law controlling the order in which the states succeeded one another, and the relation with verbalization (or, at any rate, verbalized thought). This description, which would be the analogue of a 'machine table', it was in fact the program of classical psychology to provide! Classical psychology is often thought to have failed for *methodological* reason; I would suggest, in the light of this analogy, that it failed rather for empirical reason – the mental states and 'impressions' of human beings do not form a causally closed system to the extent to which the 'configuration' of a Turing machine do. [Putnam, 1960, page 372–373]

Objections de Putnam à l'identité de types Une des stratégies de Putnam contre l'identité de types soutenue par les courants fonctionnalistes des rôles causaux est basée sur la récusation de l'identification *a posteriori* ou *synthétique* des vocabulaires théoriques différents que ces théories proposaient. Pour Putnam, avant 1967, les identifications de ces types doivent être *analytiques*, c'est–à–dire qu'elles ne doivent pas faire appel à l'expérience. Autrement dit, des prédicats P et Q correspondent à la même propriété si et seulement si ils sont analytiquement co-extensifs.

Mais à partir de 1967 il abandonne cette position et il admet en revanche l'identité synthétique des propriétés . Pour établir cette identité il suffit de démontrer qu'elle simplifie l'explication des phénomènes de la théorie.

[15] Comme le signale Joëlle Proust, cette dualité ne fait que refléter deux points de vues courants que, dans la biologie contemporaine on considère complémentaires. L'un de ces point de vues est *structurel* et met en évidence les différences anatomiques pour classer les êtres. L'autre est fonctionnalle et permet au biologiste de mettre l'accent sur l'unité des fonctions (respiratoires, digestives, etc.) qui peuvent masquer les particularités de la réalisation physique. (cf. [Proust, 1993])

Dans cette ligne de pensée Putnam propose en [Putnam, 1967a] une identification empirique entre les états mentaux et les états fonctionnels. Il prend l'état mental d'"avoir de la douleur".

Les partisans du fonctionnalisme des rôles causaux affirmaient que : "avoir de la douleur " est un état physique du cerveau, ceci étant un énoncé synthétique qui requiert une réduction empirique. Étant donné que les deux propriétés ("avoir de la douleur" et "état physique du cerveau") ne sont pas non plus des synonymes selon l'usage, l'identification empirique doit être démontrée. Pour les partisans du fonctionnalisme de types il y a là une analogie avec l'équivalence des expressions "température" et "mouvement cinétique moléculaire" dans la physique mais selon Putnam ces deux cas ne sont pas au même niveau de validité. La réduction empirique opérée dans le deuxième est valide tandis que la première est contestable car les deux états à identifier (mental et physique) ne partagent point la même région d'espace-temps. [Putnam, 1967a, page 49]

La métaphore de l'automate probabiliste Cette identification avec les états physiques est erronée et il faut donc trouver un autre type d'identification. Il s'agit d'identifier l'état mental "d'avoir de la douleur" non pas avec des états physiques mais avec des *états fonctionnels*, ainsi il émet l'hypothèse suivante:

> Tous les organismes capables d'avoir de la douleur sont des automates probabilistiques. [Putnam, 1967a, cf. page 51]

Un automate probabilistique est semblable à la machine de Turing mais il a les caractéristiques suivantes:

1.- A la différence de la machine de Turing qui a une potentialité de mémoire infinie, l'automate probabiliste a une capacité finie et fixe. ($\star$)

2.- Les transitions entre les états se font de façon aléatoire et non déterministe. Aussi cet automate est muni d' organes moteur et d'organes sensoriels et certains de ses états correspondent aux possibles 'entrées' et 'sorties'. ($\star\star$)

La réduction empirique se fait de la façon suivante:

> 1 Every organism capable of feeling pain possesses at least one Probabilistic Automaton Description (specifying the function states of the Automaton and the transition probabilities between them) of a certain kind (i.e. being capable of feeling pain *is* possessing an appropriate kind of functional organization).
>
> 2 No organism capable of feeling pain possesses a decomposition into parts that separately possess Probabilistic Automaton Descriptions of the kind referred to in [($\star$)] (This rules out a society of organisms, or a person in a room running a program)
>
> 3 For every Probabilistic Automaton Description of the kind referred to in [($\star\star$)] , there exists a subset of the sensory inputs such that an organism with that Description is in pain when and only when some of the sensory inputs are in the subset. [Putnam, 1994, page 509]

Pour Putnam à cette époque, cette description bien que vague, était toutefois plus précise et meilleure que celle qui voulait identifier la douleur à des états physiques-chimiques. Putnam a soutenu ces idées, convaincu que le rôle de la psychologie et le but de tout programme de recherche dans ce cadre est de donner un format canonique pour et par la description de *tous* les organismes. Mais plus tard il s'adresse l'autocritique suivante.

> [...] not just for the psychological description of human beings, please note (as if that were not Utopian enough!), but a normal form for the psychological description of an arbitrary organism! [Putnam, 1994, page 509]

7.2.3 L'abandon de l'hypothèse turingienne

Putnam réfute l'hypothèse qui affirmait que tout agent cognitif était *littéralement* une machine de Turing vers 1973.

Block et Fodor [Block and Fodor, 1972] avait déjà adressé leurs critiques au formalisme turingien en signalant qu'il s'avérait trop austère pour les descriptions en psychologie. Les difficultés du modèle turingien peuvent être énumérées comme suit:

1. **L'apprentissage et la mémorisation ne peuvent être représentés que de façon restreinte**

   Bien que l'acquisition des nouvelles informations puisse être représentée par des symboles nouveaux dans la bande, l'acquisition des nouveaux états n'est pas possible puisque le nombre d'états, selon le formalisme de Turing, doit être fini. Cela veut dire que les états psychologiques (voire les états de la machine) sont toujours les mêmes et ceci indépendamment des nouvelles que l'on acquiert.

2 **La machine de Turing est séquentielle**

   Il n'est pas possible pour l'agent d'être simultanément en plusieurs états à la fois ou qu'un comportement soit le résultat de plusieurs états simultanés.

3 **L'identité des états turingiens n'est pas suffisamment abstraite**

   Deux états sont identiques s'ils ont les mêmes états successeurs ou les mêmes sorties, ce qui est très limité pour l'identité des états psychologiques. Il suffit que la sortie verbale produite pour le même état psychologique soit différent pour que les deux états de la machine soient différents. Supposons qu' à la douleur causée par une brûlure l'agent A dise "Zut! " et l'agent B dise "ouch! ".

Cependant la critique plus importante faite par Block et Fodor et acceptée par Putnam lui-même est que la machine de Turing n'est pas une bonne formalisation canonique des états fonctionnels. Putnam affirme dans [Putnam, 1967a] qu'il a été contraint d'utiliser la machine de Turing pour échapper à la définition du concept d'*isomorphisme fonctionnel.*

Si Putnam voulait faire valoir son idée de l'existence des états fonctionnels, j'ai déjà signalé plus haut qu'il devait montrer un moyen de les individualiser. Postuler la notion d'isomorphisme fonctionnel est problématique parce que lorsqu'on n'a pas un formalisme comme la machine de Turing, le concept d'isomorphisme doit faire une pétition de principe. En effet, le concept d'isomorphisme fonctionnel contient en soi la supposition que les états fonctionnels existent. Comme l'explique Putnam:

> The concept which is key to unravelling the mysteries in the philosophy of mind, I think, is the concept of *functional isomorphism*. The systems are functionally isomorphic if *there is a correspondence between the states of one and the states of the other that preserves functional relations*. To start with computing machine examples, if the functional relations are just sequence relations, e. g. *state A is always followed by state B*, then, for F to be a functional isomorphism, it must the case that state A is followed by state B, then in system 1 if and only if state $F(A)$ is followed by state $F(B)$ in system 2. If the fonctional relations are, say, data or print–out relations, e.g. *when print π is printed ont the tape, system 1 goes into state A*, these must be preserved. *When print π is printed on the tape, system 2 goes into state F(A)*, if F is a functional isomorphism between system 1 and system 2. More generally, if T is a correct theory of the functioning of system 1, at the fonctional or psychological level, then an isomorphism between system 1 and system 2 must map each property and relation defined in system 2 in such a way that T comes out true when all references to system 1 are replaced by references to system 2, and all property and relation symbols in T are reinterpreted according to the mapping. The difficulty with the notion of functional isomorphism is that it *presupposes the notion of a thing being a functional or psychological description*. It is for this reason that, in various papers ont this subject, I introduced and explained the notion in terms of Turing machines. [Putnam, 1973, page 291–292]

Une autre raison qui pousse Putnam à abandonner non seulement le fonctionnalisme de Turing mais aussi tout type de fonctionnalisme est le fait que celui–ci n'est pas compatible avec la théorie linguistique préconisée par Putnam lui même.

J'ai déjà exposé en des chapitres précédents[16] sa position externaliste sur la signification que l'on a caricaturée selon l'énoncé: *les significations ne sont pas dans la tête*. Or les significations et aussi les contenus ne peuvent pas être seulement déterminés par des 'entrées sensorielles', des transitions entre des états et des 'sorties motrices' comme c'est le cas selon les thèses fonctionnalistes. Rappelons nous que pour Putnam les contenus sont déterminés par l'environnement et

[16] Voir Chapitre 4 §4.4.4

aussi par les relations avec d'autres agents. En définitive, Putnam arrive à la conclusion que le fonctionnalisme en général est incompatible avec son externalisme sémantique.[17].

7.2.4 Conclusion

Le recours de Putnam au formalisme de la machine de Turing pour individualiser les états fonctionnels de façon à satisfaire l'exigence physicaliste tout en démontrant leur réalité ontologique a échoué. J'ai exposé les raisons de cet échec. Cependant, certains concepts développés dans cette première théorie computationnelle sont toujours d'actualité.

L'idée de différents niveaux de description de la cognition a fait son chemin ainsi que le concept de multiréalisabilité de la cognition. En relation avec le premier concept, la machine de Turing ne permet que deux niveaux, le niveau logique et le niveau physique équivalents respectivement au niveau de programmation et au niveau du matériel. A mon avis, le problème fondamental du fonctionnalisme turingien est qu'il ignore l'intentionnalité. Le formalisme de la machine de Turing ne nous permet pas de représenter les concepts de la psychologie ordinaire tels que les croyances et les désirs. Il n'y a pas de place pour attribuer aux contenus des attitudes propositionnelles un rôle quelconque. De ce point de vue le fonctionnalisme de Lewis est plus conséquent avec la psychologie ordinaire car, rappelons-le, chaque état mental est pour lui un état physique (identité de types, p. ex. "avoir mal est *vérifier la stimulation des fibres de type C*") et c'est par sa condition d'état physique que la causalité se justifie en fonction du type d'état mental ou de son contenu .

Selon Turing, par contre, les états mentaux sont des états fonctionnels, d'où leur pouvoir causal (ce sont des états qui font partie du registre de la machine de Turing) mais comment faire le lien entre ce pouvoir causal et les contenus de ces états comme le voudrait la psychologie ordinaire?

C'est alors qu'on a pensé à un autre paradigme computationnel; les avances en intelligence artificielle et les recherches de David Marr ont permis de changer de paradigme. Ainsi, au lieu d'avoir deux niveaux d'analyse comme dans les cas de la machine, on en aura trois dont le troisième sera celui des niveaux representationnels.

Je vais maintenant faire état du fonctionnalisme computationnel représentationnel du MIT[18] qui comprend ces trois niveaux d'analyse. Mais auparavant je vais présenter les niveaux d'analyse et je ferai une description de l'intelligence artificielle durant ces dernières années.

7.3 L'intelligence artificielle representationnelle entre en scène

La définition de l'Intelligence Artificielle (IA) est due à Marvin Minsky, fondateur du laboratoire de IA au MIT en tant que science de réaliser des machines capables de choses pour lesquelles l'intelligence humaine serait nécessaire. [19]

Néanmoins la dénomination de l'Intelligence Artificielle remonte à John McCarthy qui organise en 1956 la conférence appelée *Dartmouth Summer Research Project on Artificial Intelligence* où le nom de la discipline nouvelle est apparu pour la première fois.

Le premier programme informatique d'IA fut construit par Arthur Samuel qui dans les années cinquante, montre un programme de jeux de dames. La particularité de ce programme était qu'il comportait un mécanisme d'apprentissage et qu'il était de nature heuristique. [20]

En 1956, un deuxième programme nommé *Logic Theorist* a été développé par Allen Newell, Cliff Shaw et Herbert Simon. Ce programme prouve 38 des 52 théorèmes des *Principia mathematica* de Whitehead et de Russell. Les théoriciens de l'intelligence artificielle se sont empressés de donner à la science nouvelle le niveau d'une science à part entière. Ainsi une grande partie de l'activité

[17]Pour une discussion de cette incompatibilité voir [Putnam, 1988, chapitre V]

[18]MIT Massachusetts Institut of Technology

[19]cf. [Minsky, 1968]

[20](du gr. *heurskein*, trouver). Qui consiste ou tend à trouver: *La méthode heuristique.* [...]. Discipline qui se propose de dégager et de formuler les règles de la recherche et de la découverte" Dictionnaire Larousse. L'heuristique peut être considérée comme une méthode foncièrement opposée aux processus algorithmiques dans la mesure où ceux-ci consistent en étapes strictement déterminées comme dans un programme informatique traditionnel.

de Newell et Simon entre autres, consista à illustrer et à défendre les ambitions de la nouvelle discipline qui à mon avis souffrait déjà d'un dualisme.

Ce dualisme vient du clivage dans la conception de la nouvelle science. D'un côté sa valeur instrumentale hors pair, elle pourrait servir comme un outil très puissant au service de l'industrie, de l'armée. D'autre part les ambitions de la nouvelle science, comme science de l'*artificiel* devaient aussi contribuer à la connaissance empirique de la nature:

> [...] construire une machine que incarne une hypothèse sur la réalité, ou construire un modèle de celle-ci, et mettre à l'épreuve cette hypothèse ou ce modèle en faisant fonctionner la machine, c'est être fidèle à la méthode expérimentale qui prévaut dans les sciences de la nature.[Dupuy, 1994, page 94]

Newell et Simon ont établi des hypothèses sur la cognition selon un format plus proche des recommandations ou directives pour l'analyse et conception d'un système informatique que d'une théorie psychologique à proprement parler.

Les chercheurs en IA ont très souvent tendance à penser que si un système effectue une performance remarquable pour réaliser une tâche considérée comme complexe il donne aussitôt lieu à une théorie sur l'aspect de la cognition qu'ils modélisent sur le champ malgré le fait que le système ait été créé de façon *ad hoc*.

L'hypothèse du traitement symbolique

L'hypothèse du traitement symbolique est à la base de toute l'IA traditionnelle. Newell et Simon ont donné les bases pour l'analyse des problèmes de façon à aboutir à la construction d'une machine aussi intelligente que l'être humain.

On peut énumérer les éléments-clés suivants:

- **Environnement:** C'est la caractérisation du problème en fonction des différents états potentiels des affaires , des actions possibles pour changer les états, des buts à partir desquels des actions peuvent être gérées.
- **Représentation interne:** Il faut donner à l'environnement du problème une représentation. Cette représentation est une collection de "structures symboliques". Elles correspondent de façon systématique à l'environnement du problème.
- **Recherche:** Il faut faire une recherche entre toutes les possibilités d'action, afin de repérer celles qui peuvent conduire au but désiré.
- **Choix:** Finalement, la sélection des actions qui permettent de réaliser le mieux possible les buts recherchés.

La façon de faire le choix est en concordance avec le principe de rationalité: si un acteur sait que l'une de ses actions va conduire à l'un de ses buts, alors il choisira cette action. Alors, selon l'hypothèse de traitement symbolique un ordinateur muni d'une mémoire suffisante et programmé selon les règles ci-dessus va acquérir une intelligence semblable aux êtres humains. L'idée sous-jacente à cette hypothèse est que l'intelligence est une affaire de manipulation adéquate de symboles.

Mais le principe de rationalité n'est pas un principe appartenant au niveau symbolique, mais un principe metasymbolique. En effet le principe de rationalité est une meta-règle pour les calculs, en quelque sort il se trouve être "la loi de conduite" de ces derniers. Ainsi donc, Alan Newell introduit la notion de niveau de connaissance (*knowledge level*) comme une autre composante du système [Newell, 1982].

Traditionnellement les niveaux d'un système informatique sont:

- Niveau de logiciel (symbolique) (*Program (symbol) Level*)
- Sous-niveau de structure d'implantation (*Registrer transfer sublevel*)
- Sous-niveau de circuit logique (*Logic circuit sublevel*)

- Niveau des circuits (*Circuit level*)
- Niveau d'appareil (*Device level*)

Chaque niveau peut être défini de deux façons différentes:

1. De façon autonome, sans faire référence aux autres niveaux.
2. En le réduisant aux niveaux inférieurs.

et chacun a les caractéristiques suivantes:

1. des moyens que l'on utilise dans les processus (i.e.: bits, symboles).
2. des composantes qui fournissent les primitives pour les processus (i.e.: registre, mémoire).
3. des lois de composition qui permettent de montrer les composantes du système.
4. des lois de conduite qui déterminent la façon dont la conduite du système dépend de la composante, de la conduite et de la structure du système (i.e.: opérations logiques).

Cependant, ces niveaux donnent seulement la possibilité de décrire les systèmes informatiques mais ne peuvent pas décrire l'environnement. C'est pourquoi Newell propose la création d'un nouveau niveau, placé au niveau du logiciel, qu'il appelle "niveau de connaissance" ("knowledge level"). Ce niveau a comme composantes des objectifs, des actions et un corps, c'est-à-dire qu'il est l'*agent*. Or, il doit déterminer les actions à faire en concordance avec des objectifs, en suivant la loi de rationnalité déjà exposée.

L'hypothèse du niveau de connaissance peut être exprimé comme suit:

> Il y a un niveau informatique différent, qui existe directement sur le niveau symbolique, qui est caractérisé par ses moyens de calculs et par le principe de rationalité comme loi de conduite.

Le niveau de connaissance doit assurer que les calculs se font en fonction des contenus intentionnels, mais pour y arriver il est nécessaire pour ce niveau d'avoir les deux caractéristiques suivantes: d'abord qu'il puisse représenter les contenus qui rendent compte des caractéristiques de l'environnement et ensuite que ses représentations aient un pouvoir causal. C'est justement sur ce point que l'IA rejoint les problèmes de la philosophie de l'esprit, car en fait les conditions citées plus hauts décrivent le problème du renvoi tel qu'on l'a vu dans des chapitres précédents.

Il est possible que devant la tâche immense que représente la résolution du problème du renvoi on renonce au but de construire des programmes pour résoudre tous les problèmes possibles, c'est-à-dire des programmes qui puissent résoudre les problèmes en général et qu'on se tourne vers les systèmes experts. Un système expert est un programme capable de résoudre des problèmes bornés à un domaine particulier.

Sémantique, informatique et langage Dans le cadre des systèmes experts, nombreux sont ceux qui cherchent la représentation du langage naturel et on trouve une discussion intéressante de cette question dans [Woods, 1975]. Woods dit qu'il y a plusieurs concepts de sémantique mélangés dans les sciences cognitives. Les linguistes aiment à expliquer le fait qu'une phrase possède plusieurs significations et à démontrer un processus selon lequel on peut déterminer qu'une phrase au contraire n'a pas de sens. Certaines philosophies s'attachent à spécifier la signification d'une notation formelle plutôt que d'un langage naturel et la notation qu'elles utilisent est dépourvue d'ambiguïté. Les philosophes en fait s'intéressent surtout aux règles de vérité. Comme le dit Woods:

> La signification pour le philosophe n'est pas définie en termes d'une autre notation à partir de laquelle on représente différentes interprétations possibles d'une phrase, mais il est intéressé par les conditions de vérité d'une représentation qui est déjà formelle.[Woods, 1975, page 38]

Woods dit que les chercheurs en intelligence artificielle et les psychologues doivent avoir une vision plus globale du phénomène sémantique.

L'utilisation du mot sémantique dans le contexte de l'intelligence artificielle tend à être abusive car le concept sémantique fait référence, non seulement à la relation entre la forme linguistique et la signification, mais aussi à toutes les capacités d'inférence du système.

Cet abus provient de l'utilisation d'informations sémantiques pour la détermination d'un objet dénoté et aussi des inférences par rapport à l'objet. D'aucuns nient la différence entre syntaxe et sémantique, ce qui est dû au fait que dans les logiciels destinés à l'analyse du langage naturel, les processus d'inférence auxquels on a recours pour choisir parmi plusieurs possibilités celle qui est correcte font partie du module syntaxique. C'est pourquoi on a tendance à penser que le processus sémantique n'existe pas.

Pour ce qui est de la théorie du langage de programmation, les spécifications sémantiques ont un rapport direct avec les actions que la machine doit réaliser. C'est pourquoi beaucoup pensent que dans les langages de programmation la syntaxe joue aussi le rôle de la sémantique mais maints critiques disent que cette équation entraîne à une forme de vérificationisme si on l'interprète hors de contexte.

A partir de la théorie logique des "mondes possibles" on peut trouver beaucoup de logiciels qui sont faits et qui représentent un partie très restreinte de la réalité. Cela est le cas du "monde des blocs" de Winograd, ou SPEECHLIS : un logiciel de compression du langage oral sur la géologie lunaire qui n'a recours qu'à 250 mots (cf. [Winograd and Flores, 1989]).

Le problème est celui-ci: comment attribuer les valeurs de vérité à toutes les propositions qu'on peut représenter dans le monde possible qu'on a choisi? On peut le faire à partir des fonctions ou des processus qui attribuent les valeurs de vérité aux propositions représentées. Un tel processus est appelé "Procedural Semantic" par Woods et se situe au niveau de la connaissance prôné par Newell. Dans la ligne de Woods, David Marr veut attirer l'attention sur l'ambiguïté existante non dans le concept de sémantique mais quant au rôle de l'intelligence artificielle. En effet, selon lui, elle ne doit pas être utilisée pour reproduire les processus cognitifs comme une autre méthode empirique de recherche. Je vais présenter les idées de Marr en relation à ce problème.

7.3.1 L'intelligence artificielle selon Marr

David Marr étudia les mathématiques à Cambridge puis il fit son mémoire de diplôme dans le département de psychologie de cette université. Il s'agissait de proposer un modèle systémique pour le fonctionnement du cervelet [Marr, 1969].

En 1973 il est invité par Papert et Minsky au Laboratoire d'Intelligence Artificielle du MIT où il développe les techniques de modélisation informatique de la vision. C'est donc au MIT qu'il fit connaissance de Tomasso Poggio qui était attaché au *Max-Plank-Institut für Biologische Kybernetik*. C'est avec lui que Marr commence la recherche qui ouvrira une ligne d'étude très riche et très importante par la suite. A cette époque, Marr forma des étudiant qui, ainsi qu'il l'a dit lui-même , devinrent très vite des collègues. Parmi eux: Keith Nishihara, Shimon Ullman, Kent Stevens. Marr est mort d'une leucémie en 1980, à l'âge de 35 ans. Malgré sa courte carrière, il a laissé des textes fondamentaux dans le domaine de la vision du point de vue représentationnel. Son livre *Vision* [Marr, 1980] est une référence irremplaçable pour toute recherche dans ce domaine.

Marr établit sa position vis-à-vis de l'intelligence artificielle par un texte paru dans la revue *Artificial Intelligence* [Marr, 1977]. Il y engage une discussion sur le danger d'utiliser l'intelligence artificielle, comme l'ont proposé Simon et Newell, pour reproduire des processus cognitifs au lieu de les expliquer et de les modéliser.

Pour éviter cet écueil, il établit très clairement la différence entre théorie et algorithme dans l'intelligence artificielle. Tout problème en intelligence artificielle a deux aspects:

1. La théorie computationnelle : cette partie est la plus abstraite de la formulation et elle répond à deux questions: que va-t-on calculer et pourquoi? Clarifier le but.

2. Les algorithmes : comment on va-t-on procéder?

Les algorithmes sont en relation avec l'équipement employé, et pour une même théorie, il peut y avoir plus d'un algorithme pour la résoudre. Selon Marr, la différence entre traitement parallèle et sériel existe aussi au niveau algorithmique, tandis que la théorie a comme unique déterminant la nature du problème.

Pour mieux expliquer cette différence il dit que la théorie computationnelle est semblable à la notion de compétence du langage dans la théorie chomskienne, tandis que le niveau des algorithmes correspond à la performance dans le langage.

Cependant pour établir les niveaux intervenant dans l'*IA*, il reprend les concepts de Newell et Simon:

1. Le niveau de la théorie computationnelle.

2. Le niveau des algorithmes.

3. Le niveau de l'application par rapport à l'équipement (*hardware*).

Chaque niveau peut être représenté selon les termes de son niveau supérieur mais l'inverse n'est pas vrai.

Concernant les méthodes de l'IA en relation avec la vision, les représentations servent à décrire l'environnement et ont par conséquent une signification basée dans le monde physique. Or, comme l'exprime Ullman [Ullman, 1979, cf. page et 11], les différences entre les niveaux proposés sont aussi dues au fait que plus le niveau où nous nous trouvons est haut, plus nous sommes proches du domaine représenté (environnement). Que le langage de la machine soit capable de représenter la réalité n'est pas un avis partagé par tout le monde.[21]

Cette référence au monde est un des principes que Marr reconnaît, parmi d'autres, comme essentiel dans la mise au point des processus symboliques complexes dans les programmes de computation. Ces principes que l'on trouve dans [Marr, 1976] sont les suivants:

1. **Le principe de la nomination explicite**

   L'acte de nomination fait pour Marr le caractère distinctif de la computation symbolique; le langage LISP est une des plus hautes expressions de cette capacité.

2. **Le principe de conception modulaire**

   La nécessité d'écrire les programmes de façon modulaire, de façon à pouvoir les changer facilement, s'il le faut, sans affecter tout le reste.

3. **Le principe d'engagement moindre**

   Ce principe établit qu'il faut toujours éviter une action qu'on devra défaire ensuite.

4. **Le principe de bonne dégradation des données**

   Cela va dans le sens qu'il faut toujours essayer de ne pas dégrader des données avant d'avoir épuisé toutes les possibilités d'utilisation pour obtenir des informations.

Marr fait aussi une classification des théories entre ce qu'il appelle "Type *I*" et "Type *II*".

Les théories de type *I* sont celles dont on sait avant de les concevoir quelle est l'information qu'on souhaite obtenir. Les théories de type *II* sont celles qui comprennent un grand nombre de processus et dont l'interaction est l'unique description atteignable.

Pour Marr cette classification n'est pas dichotomique. Cela veut dire qu'il existe un ensemble de problèmes d'IA qui peuvent entre classés comme plus proches soit d'un extrême soit de l'autre.

[21] Pour Fodor, la capacité des langages de machine de représenter la réalité de l'environnement est nulle. Il admet que le langage machine est à la fois sémantiquement et syntaxiquement interprétable, mais récuse l'idée que ce langage puisse modéliser le langage universel de représentation postulé par Chomsky et repris par lui- même dans *The language of the thought*[Fodor, 1975]. En fait, pour le langage interne de la machine, l'unique environnement qu'elle peut représenter est l'environnement des états de la machine et rien d'autre. [Fodor, 1978].

7.4 Les théories representationnelles du MIT

L'hypothèse du traitement symbolique a engendré un nouveau paradigme qui, bien que mathématiquement équivalent à celui de la Machine de Turing est toutefois plus éloquent. La théorie représentationnelle développée au MIT à partir de la fin des années soixante se base sur ce paradigme.

Les développements de la *grammaire transformationnelle ou générative* de Chomsky retracés dans son livre *Syntactic Structure* et les développements en IA conjointement avec les théories de Turing ont constitué le bouillon de culture de la théorie représentationnelle.

Les motivations philosophiques telles que l'hypothèse Turing–Church, l'hypothèse du traitement symbolique (bien que cette dernière soit encore en état embryonnaire) et la division entre sémantique et syntaxe, bien tranchée à partir des travaux des Chomsky et de Jerry Fodor[22] sont arrivés à un état de maturité tel qu'elles ont permis la postulation de ce modèle représentationnel dont la version fodorienne repose sur l'existence du *Langage de la pensée*.

7.4.1 Représentation et parallélisme syntaxico–causal

Dans le premier chapitre de *The language of the thought* [Fodor, 1975], Fodor fait l'énumération des points de son programme :

1. Les uniques modèles psychologiques des processus cognitifs qui paraissent même de loin plausibles, paraissent computationnels.

2. La computation présuppose un moyen de faire des calculs : un système de représentation.

3. Avoir des théories "de loin plausibles" est mieux que de n'avoir aucune théorie.

4. Nous sommes, de façon provisoire, obligés d'attribuer un système de représentation à des organismes.

5. C'est un but raisonnable de recherche que d'essayer de caractériser le système de représentation que nous sommes, de façon provisoire, obligés d'attribuer aux organismes.

6. C'est une stratégie raisonnable de recherche que d'essayer d'inférer cette caractérisation à partir de détails appartenant à ces théories psychologiques qui semblent prouvées.

7. Cette stratégie peut être réellement efficace : il est possible de trouver un type d'inférence comme il est décrit dans le point 6, lequel, s'il n'est pas exactement apodictique,[23] aurait, prima facie, l'air d'être plausible.

Mais que représentera le système d'un agent donné?

D'abord il y aura une représentation de la situation S, ensuite de l'ensemble d'options disponibles dans S et aussi des conséquences probables de chaque option qu'il a ordonné eu égard à ses préférences. Comme l'agent fait les calculs de façon computationnelle, et que l'action qu'il exécute est le résultat des calculs qu'il vient de faire à partir des options possibles représentées, il faut qu'il ait accès à un système de représentation assez riche.

Tout ce processus caractérise l'agent comme ayant des attitudes propositionnelles, c'est-à-dire comme étant, en quelque sorte, en rapport computationnel avec sa propre représentation interne. De cette façon le cœur de l'explication en psychologie cognitive tient à l'organisme et à ses attitudes propositionnelles. Pour chaque attitude propositionnelle que l'organisme peut avoir,

[22] Jerry Fodor et Jerrold Katz en 1963 ont les premiers essayé de faire de la sémantique une partie systématique de la description linguistique. En effet, la première proposition de Chomsky (1957) ne tient pas compte d'un module sémantique; c'est seulement lorsqu'il propose sa théorie dite "standard" que la composante sémantique est admise [Chomsky, 1965].

[23] Apodictique: logiquement nécessaire, par opposition à l'assertorique et au problématique. Ces termes ont été répandus par Kant, qui en fait les trois divisions de la modalité des jugements -Lalande, A.(1985) Vocabulaire Technique et Critique de la Philosophie).Paris, Editions Etudes Vivantes.

il y a une représentation interne et une relation avec elle qui détermine de façon univoque une attitude propositionnelle.

Les relations entre l'organisme O et les représentations F s'appellent états mentaux et, ceux-ci ont entre eux des relations causales, c'est-à-dire qu'il y a une succession entre eux en concordance avec les principes de calcul qui sont appliqués aux représentations. Mais il n'est pas possible d'avoir un processus de calcul sans une représentation, et s'il n'y a pas de processus de calcul il n'y a pas de modèle. Ainsi le modèle (ou le schéma de modèle) présenté par Fodor a besoin d'un langage.

D'abord parce qu'il doit pouvoir être capable de représenter une infinité de situations possibles, c'est-à-dire avoir la "capacité" d'une langue naturelle. Ensuite la nécessité de représenter des aspects saillants de ces situations à l'agent même suppose une familiarité avec les propriétés sémantiques de vérité et de référence.

Le système de représentation est un langage, mais il ne peut être un langage naturel parce que c'est, par exemple, un moyen d'expliquer les comportements et il y a des comportements chez les bébés avant même l'apprentissage de la langue.

On peut résumer l'idée de Fodor, jusqu'ici, de la façon suivante:

1. Les modèles disponibles des processus cognitifs caractérisent ces processus comme fondamentalement computationnels et de là on présuppose un système de représentation dans lequel les calculs tournent.

2. Ce système de représentation ne peut pas être un langage naturel, bien que:

3. Les propriétés sémantiques de tout langage naturel qu'il est possible d'apprendre doivent être exprimables dans le système.

Pylyshyn affirme que "l'idée que les processus mentaux sont computationnels est en effet une hypothèse empirique sérieuse plutôt qu'une métaphore." [Pylyshyn, 1984, page 55] Il considère qu'il y a trois niveaux d'explication: le niveau physique (biologique), le niveau symbolique ou syntaxique, qui est quelquefois appelé le niveau fonctionnel, et le niveau sémantique ou intentionnel.

Dans l'ordinateur on peut faire référence aux objets dans un domaine (interprétation intentionnelle ou sémantique) ou bien aux sujets du calcul, i.e. le domaine abstrait des nombres. Pour Pylyshyn et Fodor, l'ordinateur n'interprète pas les symboles comme étant nombres, ils sont seulement patrons formels qui font tourner la machine. Dans le niveau physique, les règles et les représentations sont codifiées et les calculs sont gouvernés par tous les deux.

Tout cela pose le problème du dualisme de la nature du fonctionnement mental: ou bien elle est fonctionnelle (causale) ou bien elle est intentionnelle. Pour Pylyshyn, la clef de la solution semble se trouver dans la séparation, à l'intérieur du domaine cognitif, des aspects sémantiques et des aspects syntaxiques. Cependant, il y aura une relation entre eux:

> Parce qu'un processus computationnel n'a pas d'accès au domaine réel lui même..., il est nécessaire, si les règles doivent continuer d'être interprétables de façon sémantique...que toute distinction sémantique importante puisse être reflétée par des distinctions syntaxiques...Ces caractéristiques doivent être reflétées en différences fonctionnelles dans la mise en oeuvre du procédé. C'est à cela qu'on fait référence quand on dit que le procédé représente quelque chose. Bref, seuls les aspects syntaxiques codifiés (mais tous ces aspects) du domaine représenté peuvent affecter la façon dont les processus sont orientés. [Pylyshyn, 1980, pages 113-114 ma traduction]

Fodor le dit de façon plus claire:

> On permet que les représentations mentales affectent la conduite en vertu de leurs contenus, mais on soutient que les représentations mentales sont distinctes en contenu seulement si elles sont distinctes aussi dans leurs formes.
>
> La première clause est nécessaire à la plausibilité de la relation des états mentaux avec les représentations mentales et la deuxième rend seule plausible l'identification des états mentaux à des calculs (Les calculs sont justement des conséquences causales en vertu de leur forme). [Fodor, 1980, page 99 ma traduction]

Mais Pylysbyn affirme que personne n'a la moindre idée sur la manière dont le processus ainsi décrit serait compatible avec la loi naturelle. Fodor pense que les processus mentaux n'ont accès qu'aux propriétés formelles (non–sémantiques) des représentations mentales, bien qu'ils aient la possibilité de les représenter.

> Mon avis, donc, n'est pas, bien sûr, que le solipsisme [méthodologique][24] est vrai, c'est seulement que la vérité, la référence et le reste des notions sémantiques n'ont pas de catégories psychologiques. [Fodor, 1980, page 71 ma traduction]

L'hypothèse que toute différence sémantique doit être reflétée par une différence syntaxique s'inspire clairement de l'hypothèse de l'isomorphisme de Carnap.

> Carnap (1947) defined a relation of intensional isomorphism between expressions such that two expressions are intensionally isomorphic just in case both have the same internal structure, and all corresponding parts of the two expressions have the same intension. [Fodor, 1977, page 46]

Le pas sémantique qui permet d'associer les symboles au langage se ferait avant la représentation mais il n'y a aucune hypothèse sérieuse qui explique comment cela doit se passer. Cette hypothèse aide aussi à résoudre le problème de la causalité parce qu'elle permet de rendre cette dernière formelle, autrement dit, elle permet qu'elle soit causale et formelle à la fois. Elle est formelle parce que les causes de l'action sont calculées à partir de processus qui agissent seulement au niveau symbolique, et elle est causale parce que les symboles relèvent d'un langage qui permet de refléter les caractéristique sémantiques en tant que caractéristiques syntaxiques

Or, pour Fodor, les causes de la conduite mentale ne sont pas sémantiques du point de vue des processus de calcul, mais le sont pour la représentation.

Toutefois Fodor ne reste pas figé sur ces positions. Il a subi une évolution depuis la publication de *The langage of the thought*. Les critiques qu'on lui a faites sur le concept du *contenu étroit* et dont j'ai traité dans des chapitres précédents[25] l'on amené à revoir cette notion. Néanmoins, il ne faut pas croire qu'il soit facile à convaincre, bien au contraire. Fodor est un polémiste brillant dans les controverses philosophiques. Écouter Fodor argumenter après une de ses conférences évoque l'image d'un grand maître d'échec jouant des parties simultanées : un petit tour et tout le monde est "échec et mat". L'analogie établie par Dan Dennet pour le décrire est très précise :

> [...] most philosopher are like old beds: you jump on them and sink deep into qualifications, revisions, addenda. But Fodor is like a trampoline: you jump on him and he springs back, presenting claims twice as trenchant and outrageous. If some of us see further, it's for jumping on Jerry [Fodor]. [Loewer and Rey, 1991, page ix]

Je vais continuer à exposer la conception fodorienne de l'esprit à la lumière de ses derniers écrits. Notamment, les conférences en Sciences Cognitives "Jean Nicod" organisées par le Centre National de la Recherche Scientifique français dans le cadre du programme "CogniSciences" dont Fodor était le premier orateur invité.

7.5 La conception fodorienne de l'esprit

Contrairement à Donald Davidson pour qui les explications basées sur des états mentaux ne méritent pas la caractérisation de nomologique, Jerry Fodor soutient que les généralisations en psychologie sont des lois. Il en découle que les états intentionnels sont dotés de pouvoirs causaux et que ces pouvoirs sont en rapport direct avec leur contenu.

Fodor peut revendiquer une position physicaliste parce qu'il soutient que les propriétés mentales se *multiréalisent* ou *sont implantées* par l'intermédiaire du niveau computationnel.[26] La définition des propriétés mentales sera donnée en termes fonctionnels et non plus strictement physiques.

La postulation que les propriétés mentales sont implantées au niveau computationnel fait d'elles des propriétés de deuxième ordre en termes de niveaux d'implantation vis-à-vis du niveau physique et permet de combiner la position fodorienne avec l'hypothèse de la multiréalisation de ces mêmes propriétés.

La multiréalisabilité de la cognition dans cette théorie joue un double rôle. Premièrement elle garantit le caractère non-réductionniste de la solution de Fodor. En deuxième lieu, elle permet de

[24] Le solipsisme méthodologique pour Fodor doit être compris comme une théorie empirique sur l'esprit qui en affirme la nature computationnelle.

[25] Chapitre 4 §4.4.1 – 4.4.3.

[26] Dans les prochaines sections je discuterai le concept de *réalisation* en détail.

donner un compte–rendu du caractère non–strict des généralisations en psychologie sans démentir le caractère strict des lois de la physique. Les exceptions à ces lois en psychologie seront expliquées non comme des exceptions aux lois fondamentales du niveau physique mais comme des exceptions au niveau computationnel.

Bref, la conception fodorienne de l'esprit tient en trois phrases:

- Les lois en psychologie sont intentionnelles.
- La sémantique est purement informationnelle.
- La pensée est computationnelle.

J'ai déjà expliqué le troisième postulat dans la section précédente tout en exposant l'hypothèse du langage de la pensée.

Je vais à présent expliquer les deux premiers et indiquer les difficultés qui découlent de la conjonction des trois.

7.5.1 Les lois en psychologie sont intentionnelles

Selon Fodor, les généralisations en psychologie se basent sur des phrases qui expriment des croyances et des désirs. Ces généralisations qui appartiennent à la psychologie ordinaire ont les deux caractéristiques fondamentales suivantes: elles font référence aux états mentaux et l'on détermine leur pouvoir causal en fonction de leur contenu.

> So, typical intentional generalisation might be of the form: 'If you want to [...], and you believe that you can't [...] unless you [...] then, ceteris paribus, you will perform an act that is intended to be [...]'. E.g.: If you want to *make an omelette*, and you believe that you can't *make an omelette* unless you *break some eggs*, then, ceteris paribus, you will perform an action that is intended to be *an egg breaking*. (Whether the action actually *succeeds* in being an egg breaking depends, of course, on whether the world cooperates and the eggs break.) Notice that the beliefs and desires and actions subsumed by such generalisations are picked out by reference to their contents; to what they are belief *that*, desires *for* and intentions *to*. Patently, then, if you propose to take it seriously that psychological explanations are intentional, you had better have a theory of content up the sleeve.[Fodor, 1994, page 4, le texte en italique appartient au texte original]

Ce premier postulat sur le caractère nomologique des généralisations en psychologie pose deux problèmes. D'abord, dans quelle mesure est–il possible d'appliquer le concept de loi aux généralisations basées sur des énoncés intentionnels? A cet égard, je discuterai le concept de science spéciale selon Fodor pour pouvoir caractériser le concept de généralisation dans ces sciences.

Le second problème est de rendre compatible le caractère causal octroyé aux états mentaux avec leur contenus.

Dans cette section je discuterai le premier problème alors que j'aborderai le second lors de la discussion du caractère informationnel des contenus dans les sections suivantes du chapitre.

Les sciences spéciales Fodor partage l'idée selon laquelle tout ce qui est réel est sous l'empire des lois de la physique fondamentale; en revanche, certains objets et événements, en fonction de leurs spécificité tombent, en outre, sous les lois de telle ou telle science spéciale. [27]

Au contraire de la physique fondamentale qui se caractérise par l'universalité de son domaine d'application, les sciences spéciales régissent un domaine restreint d'objets et d'événements. Ce domaine restreint auquel ces dernières s'appliquent limite aussi la validité de leurs lois.

Selon [Kistler, 1995b] la conception fodorienne des sciences les divise en deux catégories binaires différentes. La première catégorisation sert à refléter le clivage des sciences selon le domaine d'applicabilité : on aura la physique d'un côté et les sciences spéciales de l'autre. La seconde catégorisation, en revanche tranche selon le caractère des lois : d'un côté les lois strictes, de l'autre les lois qui ne valent que sous certaines conditions *ceteris paribus*. L'analyse de Fodor va encore plus

[27] cf. [Kistler, 1995b]. Je voudrais remercier Max Kistler de m'avoir facilité ce manuscrit si clair et si édifiant et je lui suis gré également de nos discussions au sujet des clauses *ceteris paribus*.

loin quand il affirme que les catégories obtenues par l'application de n'importe lequel des deux critères sont les mêmes; c'est à dire que seule la physique a des lois strictes.[28]

Les généralisations en psychologie Certains auteurs nient que des généralisations nomologiques puissent être établies en psychologie. Rappelons-nous par exemple de la position de Donald Davidson[29] qui nie le caractère nomologique des énoncés mentaux et réduit ainsi la psychologie ordinaire à une valeur instrumentale plutôt que scientifique.

Une discussion de l'exemple de la généralisation que Fodor propose dans la citation sur le désir de manger des omelettes nous permettra d'exemplifier les limites de telles prédictions. On voit que la généralisation donnée à partir de mon désir de me cuire une omelette n'est valable que si je n'ai pas une crise de foie qui ne me permettrait même pas de sentir la pression d'une plume sur le côté droit du corps justement à la hauteur du foie, que si j'ai des oeufs à la maison, ou que le jour n'étant pas férié je peux aller les acheter, et ainsi de suite.

Les détracteurs disent que les généralisations de la psychologie ordinaire ne peuvent pas servir de prédictions car elles doivent être bornées (*hedged*) par des conditions ou clauses dites *ceteris paribus*. Ces clauses priveront fatalement ces types de généralisation de tout pouvoir prédictif. Pour les détracteurs des généralisations en psychologie il y deux chemins possibles. Soit ces types de généralisations sont vides de tout pouvoir prédictif, soit elles sont tout simplement fausses.

Les défenseurs des généralisations en psychologie en revanche, disent que l'on pourrait se mettre à l'abri des toutes les contrefactuelles[30] en ajoutant aux généralisations la clause "toutes les autres choses étant égales par ailleurs". Ceci éliminerait les cas où j'éprouve le désir de manger un omelette lorsque j'ai une crise de foie ou que je n'ai pas d'oeufs à la maison, etc. en les réduisant à cette seule expression.

Selon Fodor, toutes les sciences spéciales qui ont des schémas explicatifs empiriques possèdent des clauses *ceteris paribus* mais à la différence de la psychologie leurs clauses sont explicites. La spécificité des généralisations en psychologie, quant à elle tient au caractère implicite de ces clauses.

Fodor nous propose d'analyser une généralisation d'une autre science spéciale. Prenons une loi simple de la géologie : "un fleuve sinueux ronge sa berge extérieure". [Fodor, 1987, cf. page 4]

Si l'on suit la démarche des détracteurs des généralisations en psychologie on devrait dire si l'on prend cette expression comme une généralisation stricte qu'elle s'avérera sûrement fausse. En effet, elle ne se vérifiera que si le climat ne change pas en faisant geler le fleuve, si la fin du monde n'a pas lieu, ou si l'on ne construit pas un barrage [...].

Pour les partisans de l'existence des généralisations en psychologie, il serait erroné de les comprendre de cette façon parce que si l'on peut réduire l'énoncé précédent au suivant : "un fleuve sinueux ronge sa berge extérieure à moins qu'il ne la ronge pas."

Une autre façon plus adéquate d'énoncer la généralisation précédente serait "un fleuve sinueux ronge sa berge extérieure, toutes autres choses étant égales par ailleurs", ce que l'on peut paraphraser comme suit "un fleuve sinueux ronge sa berge extérieure dans tout monde nomologiquement possible où les idéalisations de la géologie se vérifient."

Or, ce que les détracteurs voient comme un défaut majeur des généralisations en psychologie n'est qu'une des caractéristiques incontournables de toute généralisation dans toute science spéciale. Néanmoins, le fait que ces dernières doivent être bornées par des clauses *ceteris paribus* n'enlève en rien leur pouvoir informatif.

J'ai déjà signalé que Fodor explique l'existence des exceptions au niveau psychologique sans pour autant soutenir l'existence d'exceptions au niveau plus fondamental de la physique. Il y parvient en assumant que les propriétés psychologiques se multiréalisent dans des niveaux inférieurs. C'est justement le fait d'être multiréalisées qui explique l'existence des exceptions. Je discuterai ceci dans les sections suivantes.

---

[28] Le fait de considérer les lois de la physique comme étant strictes a été mis en doute récemment par Carl Hempel [Hempel, 1988]. Or la taxonomie fodorienne ne fait pas l'unanimité.

[29] cf. Chapitre 4 de ce texte.

[30] J'emprunte la définition de *contrefactuelle* à Frank Döring: "une contrefactuelle est une phrase affirmative dont l'énonciation dans un contexte donné et dans des circonstances normales peut s'avérer aussi bien vraie que fausse". [Döring, 1995]

La réalisation de lois: Au contraire de la relation de réalisation *des propriétés*[31] qui est compatible avec une position physicaliste du mental, la réalisation *des lois* s'avère difficilement compatible avec le physicalisme si l'on prend une position forte pour la réalisation.

L'analyse des réalisations de lois peut se faire à partir de deux perspectives différentes. La première est celle qui prend en considération une conception de la réalisation forte[32]de Kim qui rendra le mental inerte du point de vue causal. La seconde choisira la conception faible et c'est celle empruntée pac Fodor.

Le problème que l'on se pose est exprimé dans le scbéma 7.6 qui montre le schéma du mécanisme de bas-niveau servant à implanter une loi de plus haut niveau.

On dira que F (par exemple une propriété physique) est l'implantation ou réalisation de M_F (par exemple une propriété mentale).

$$\begin{array}{ccc} M_F & \longrightarrow & M_G \\ \downarrow & & \downarrow \\ F & \longrightarrow & G \end{array} \tag{7.6}$$

En illustration prenons la loi de la psychologie ordinaire suivante: "Tout honnête homme, lorsqu'il croit avoir fait tort à quelqu'un (c'est à dire tout honnête homme ayant la propriété mentale M_F) a le désir de présenter ses excuses à cette personne (a la propriété mentale M_G)"

La croyance d'avoir fait tort à quelqu'un est la cause du désir de s'excuser. Cependant, selon le compte-rendu physicaliste qui nous occupe, les propriétés M_F et M_G sont réalisées ou implantées sur les propriétés physiques F et G respectivement. Selon la définition de réalisation de Kim que nous avons donnée dans le chapitre 2, la propriété physique F et la propriété physique G sont des conditions suffisantes pour la réalisation de M_F et de M_G respectivement. En outre, l'existence de la propriété physique G dans le système (dans l'exemple, par "système" il faut comprendre notre honnête homme) est assurée par la relation nomologique causale de base que nous avons exprimée dans notre schéma 7.6 comme $F \longrightarrow G$. Il s'avère que la conjonction de F et de cette dernière loi suffit à la réalisation de M_G.

Cependant, la loi de la psychologie ordinaire sigoifie que la propriété M_F (la croyance d'avoir fait tort à quelqu'un) est aussi la cause suffisante de M_G. De tout ce que je viens d'exposer il ressort que la propriété M_G (le désir de s'excuser) est surdéterminée. Il n'est pas possible de plaider en faveur d'un postulat que les deux événements sont nécessaires pour causer M_G parce que chacun d'eux par lui même suffit pour manifester ladite propriété.

Dans ce cas il y a deux chemins possibles. Le premier qui adopte une conception forte de la réalisation en récusant la dualité entre *explications* et *nécessité nomologique* menace de priver de tout pouvoir causal la propriété M_F. L'autre, qui part d'une conception faible de la réalisation semble au moins *sauver quelques meubles* quant à la pertinence causale du mental. Si l'on choisit une conception forte de la réalisation, alors Kim affirme que l'unique solution permettant de sauver la pertinence causale du mental est d'accepter le principe suivant. Il s'agit du *Principe causal de la réalisation.*

> *The causal Realization Principle* If a given instance of S occurs by being realized by Q, then any cause of this instance must be a cause of this instance of Q (and of course any cause of this instance of Q is a cause of this instance of S)[Kim, 1993b, page 208]

A mon avis, le principe causal de la réalisation qui est une pétition de principe ne rend pas indiscutable la pertinence causale du mental. En effet, pour que la pertinence causale du mental

[31]Pour une définition du concept de réalisation voir le chapitre 2 §2.3.4.

[32]Dans le chapitre 2 (§2.3.4.) j'ai appelé *réalisation forte* la conception selon laquelle le niveau d'implantation est le niveau de base ou physique. La conséquence de cette condition est que toutes les relations nomologiques établies entre les propriétés de niveau supérieur et celles de base sont d'emblée explicatives parce qu'elles contiennent des explications en fonction de la microstructure. Cette condition implique le caractère épiphénomenal du mental. En effet, nous savons que l'on peut laisser une place au moins explicative au concept de réalisation si l'on peut considérer la dualité causale explicative.

puisse être préservée, il faut qu'une propriété mentale soit comptée comme une cause de la propriété S. En outre, ce principe entre en conflit avec la définition de réalisation forte car si la propriété physique suffit à l'instantation de la propriété mentale on ne voit pas pourquoi on devrait tenir compte d'autres causes. Finalement, la réalisation prise dans le sens fort semble condamner les propriétés mentales à l'épiphénoménalisme.

La seconde solution me semble plus acceptable. Il s'agit de trouver une solution sans rejeter le cadre suivant : M_F est réalisé par F et $F \rightarrow G$ pour une loi de niveau inférieur à celui de M_F et étant donné que G réalise M_G alors on a M_G.

Selon Fodor la conclusion d'inertie causale du mental vient d'une confusion amenant à croire que s'il y a une loi de niveau supérieur (p. ex. une loi en psychologie) et qu'il y a aussi un mécanisme d'implantation pour cette loi alors les propriétés projetées par cette loi se révèlent inertes. [Fodor, 1989a, cf. page 142]. [33] La situation telle que Fodor nous la présente est un peu différente. Tout d'abord les généralisations en psychologie sont des lois qui n'appartiennent pas au niveau de base. Les causes mentales ont leurs effets du fait qu'elles sont soumises à ces lois en psychologie; mais étant donné que les lois en psychologie ne sont pas des lois de base, elles ont besoin d'un mécanisme de médiation ou d'implantation. Fodor continue :

> However, it seems to me that to admit that mental causes must be related to their effects (including, notice, their *mental* effects) by physical mechanisms *just is* to admit that mental causes are physical. Or, if it's not, then it's to admit something so close that I can't see why the difference matters. [Fodor, 1989a, page 156, les italiques font partie du texte original]

C'est une chose de dire que les entités mentales sont liées causalement à leurs effets par des mécanismes physiques, et d'admettre que les causes mentales sont des causes physiques et une autre de conclure qu'elles n'ont aucune pertinence causale. Fodor propose le concept de *responsabilité causale* qui, tout en étant différent de celui de cause *stricto sensu* laisse place à la causalité du mental. Le concept de responsabilité causale dans le cas du mental est en rapport foncier avec le caractère non-strict des lois en psychologie. Si l'on a une loi en psychologie qui dit que "$M_F \longrightarrow M_G$ ceteris paribus" alors le fait que celle-ci soit une loi non stricte et non une relation contingente est dû au caractère de responsabilité causale que l'on peut octroyer à M_F [Fodor and LePore, 1992, cf. page 152].

La différence fondamentale entre l'approche de Fodor et de Kim est que le second considère que toutes les relations métaphysiquement nécessaires sont aussi explicatives tandis que pour Fodor, ces relations bien que nécessaires doivent faire référence explicitement au mécanisme d'implantation. Pour Kim, le niveau réalisateur est le niveau basique (physique); ainsi l'implantation revient au même qu'une description en termes de microstructure, alors que Fodor admet l'éventualité que le niveau réalisateur ne soit pas le niveau physique, d'où le besoin d'éclaircir le mécanisme de réalisation qui dans le cas du mental est un mécanisme syntaxique chargé du rôle explicatif.

---

[33] En effet, la définition de la propriété projetée de Fodor doit être comprise de la manière suivante : Si "l'instantiation des $F's$ est suffisante (causalement et non dans le sens de réalisation) à l'instantiation des $G's$" est une loi causale, alors on dira de toutes les paires d'événements qui vérifient cette loi qu'elles sont *couvertes* par cette loi. On dira aussi, que cette loi *projette* les propriétés en vertu desquelles ces événements individuels s'avèrent être couverts par cette loi.

Soient les paires d'événements individuels suivants, (F, G) qui sont des événements physiques et (M_F, M_G) qui sont des événements mentaux. Supposons maintenant que les deux paires sont couvertes par la loi citée ci-dessus. La loi va faire la projection des propriétés en vertu desquelles ces événements individuels sont couverts par cette loi. Jusqu'ici tout va bien mais qu'arriverait-il si, comme la notation que j'utilise le suggère, les événements M_F et M_G étaient réalisés par les événements F et G respectivement ?

Si les événements M_F et M_G sont couverts par la loi citée mais sont aussi réalisés par les événements F et G, les propriétés projetées par la loi ne sont pas celles de M_F et M_G mais plutôt celles projetées par leur réalisation, c'est-à-dire les propriétés de F et G. Selon notre hypothèse physicaliste tous les événements mentaux sont réalisés par des événements physiques (en définitive ils sont mentaux mais ils sont aussi physiques), alors si ce n'est pas en vertu des propriétés de F et G, il résulte que les propriétés projetées le seront par d'autres événements physiques (dues à la possibilité de multiréalisation), disons F' et G'. Ces derniers événements, bien que différents des précédents sont néanmoins tout aussi physiques.

Voilà pourquoi l'hypothèse de réalisation des états mentaux au niveau physique rend inertes les propriétés mentales projetées par des lois.

Dans le cadre de la théorie de Fodor le schéma précédent 7.6 peut être remanié de la manière suivante [Fodor, 1994, cfr. page 10–figure 1.3]:

$$\begin{array}{ccc} M_F & \longrightarrow & M_G \\ \uparrow & & \downarrow \\ F & \longrightarrow & G \end{array} \tag{7.7}$$

Toutes les flèches représentent des relations suffisantes. Les flèches horizontales exprimeront des relations suffisantes causales, les flèches verticales exprimeront des relations suffisantes de réalisation.

Pour rejeter le caractère epiphénomenal du mental de M_F, il suffit de réclamer l'existence d'une relation suffisante de M_G à G. Cette condition n'exprimera point une relation causale mais elle exprimera des mécanismes selon lesquels M_G comporte une responsabilité causale. En définitive, cette dernière relation sera la condition d'implantation.

Ces mécanismes justifient le fait que $M_F \longrightarrow M_G$ soit une loi non–stricte plutôt qu'une relation contingente.

En effet, Fodor fait une différence entre les concepts de responsabilité causale d'une macropropriété (ou de niveau supérieur) et celui de spécification du mécanisme d'implantation. A mon avis, c'est justement l'existence de ce mécanisme d'implantation qui rend la macropropriété causalement responsable.

Fodor fait du concept de réalisation des lois un outil très élégant du physicalisme non–réductionniste. Selon sa conception des sciences spéciales, le compte–rendu de la réalisation des lois appartenant à ces dernières est une démarche indispensable.

Si les lois appartiennent à des sciences fondamentales comme la physique, l'expression d'une loi du type "*Fs causent Gs*" n'a pas besoin d'explication tout simplement parce que l'on est déjà au niveau le plus fondamental. Au contraire, s'il s'agit des lois des sciences spéciales il faut démontrer le mécanisme d'implantation. Cette démarche est différente de celle qu'implique une réduction.

Le concept de réalisation se démarque de celui de réduction au moins dans le sens nagelien. Il ne s'agit pas, comme il en est pour une réduction, d'un changement de vocabulaire basé sur des identités proposées par les lois ponts, mais plutôt d'un changement motivé par les termes requis pour expliquer le mécanisme d'implantation dans le cas le plus général. Comme Fodor nous le dit:

> If you want to talk laws of inheritance, you talk recessive traits and dominant traits and homozygotes and heterozygotes; if you want to talk mechanisms of inheritance, you talk chromosomes and genes and how the *DNA* folds. If you want to talk psychological law, you talk intentional vocabulary; if you want to talk psychological mechanism, you talk syntactic (ou maybe neurological) vocabulary. If you want to talk geological law, you talk mountains and glaciers; if you want to talk geological mechanism, you talk abrasion coefficients and cleavage planes. If you want to talk aerodynamic law, you talk airfolds and lift forces; if you want to talk aerodynamics mechanism, you talk gas pressures and laminar flows. It doesn't follow that the property of being a belief or an aircraft or a recessive trait is causally inert; all that follows is that *specifying the causally responsible macroproperty isn't the same as specifying the implementing mechanism* [Fodor, 1989a, page 146, les parties en italiques appartient au texte originale]

Finalement la réalisation des *lois à la Fodor* a la vertu de plaider pour l'existence d'une réduction épistémologique via le concept de réalisation sans opérer, de ce fait une réduction ontologique du mental.

Dans la section suivante je discuterai le concept de multiréalisation et l'utilisation que Fodor en fait pour expliquer les exceptions aux lois en psychologie.

La multiréalisation des propriétés et des lois: L'argument de la possibilité logique de multiréalisation des propriétés mentales a été utilisé dans le début des années soixante par Hilary Putnam. Il voulait répondre au physicalistes de types qui prônaient l'identité entre états mentaux et états physiques qui avaient les mêmes rôles fonctionnels[34]. La multiréalisabilité est une arme contre le physicalisme réductionniste.

[34] Voir chapitre 6.

Un état physique, par exemple la douleur que je vais représenter par M_D, peut être réalisé par différents types de structures neurologiques. Supposons qu'il soit représenté par N_H chez les hommes, N_M chez les martiens, N_V chez les vaches, etc. L'état mental M_D pourra être réalisé par chacun des états N_X de la liste qui est d'ailleurs exhaustive.

On peut dire que l'état M_D est réalisé par la disjonction suivante: $N_H \vee N_M \vee N_V \ldots$ Cette disjonction des états garantit l'impossibilité d'une réduction parce qu'on ne pourra pas dire, par exemple que l'état M_D n'est que l'état N_H. L'identification d'un état mental avec un autre état physique n'est pas possible.

Il en est de même pour les propriétés.

En effet, il y a une différence entre *réduction* et *réalisation multiple* pour une propriété P_L du niveau L:

- La réduction est obtenue toutes les fois que la propriété P_L du niveau L est *identifiable* avec une des propriétés du niveau $L-1$.
- La réalisation multiple est obtenue s'il existe une disjonction de propriétés du niveau $L-1$ telles que:
  1. l'instantiation d'un des éléments de la disjonction est suffisant pour l'instantiation de P_L.
  2. l'instantiation de P_L est suffisante pour l'instantiation de la disjonction mais non pour l'instantiation d'un de ses éléments. [Fodor, 1994, cf. page11]

La seconde condition assure le caractère suffisant et elle exclut explicitement le caractère nécessaire de la relation de réalisation multiple; en effet la réalisation de la propriété d'ordre supérieur suffit à l'instantiation de la disjonction et non à l'instantiation d'un de ces membres. Le fait que la propriété d'ordre supérieur soit réalisée chaque fois, à l'instant t par un seul membre de la disjonction ne doit pas être confondu avec la possibilité logique que dans d'autres mondes possibles il s'agisse d'un autre membre de ladite disjonction.

En outre, l'individualisation des états mentaux se fait en termes fonctionnels.

> In the classic case of multiple realisation, the higher-level property is said to be 'functionally defined', and the realizing disjunction includes all and only the mechanisms that can perform the defining function. (So, for exemple, there is presumably some disjunction of mechanisms any of which might perform the defining function of a carburator, and such that every nomologically possible carburator is an instance of one of the disjuncts or other).These days, most philosophers of mind suppose that most psychological properties are multiply realized. [Fodor, 1994, page 11]

La figure 7.1 nous montre la propriété G qui est multiréalisée par la disjonction $M1_G \vee M2_G \vee M3_G \vee \ldots$ Par conséquent n'importe quelle occurrence de la disjonction est une condition suffisante pour une occurrence de G.

J'ai déjà montré comment la multiréalisation des propriétés milite contre une conception réductionniste. En ce qui suit, j'illustrerai comment la multiréalisation sert à expliquer les exceptions des lois des sciences spéciales.

Multiréalisation et exceptions des lois des sciences spéciales: Rappelons-nous que, pour Fodor toutes les lois des sciences spéciales sont soumises à des clauses *ceteris paribus*. La différence entre les généralisations et les lois strictes est que les premières ne supportent pas les contrefactuelles tandis que les dernières leurs sont indifférentes. Or, dans les sciences spéciales il y a des exceptions.

Les exceptions à une loi servent à la réfuter s'il s'agit d'une loi stricte; en revanche s'il s'agit des énoncés de lois *ceteris paribus* [Fodor, 1991, page 27] la multiréalisabilité est au centre de l'explication des exceptions. En effet Fodor explique la possibilité des exceptions à une loi par la réalisation de propriétés multiples mises en jeu par cette loi.

Pour Fodor une loi peut avoir des exceptions si et seulement si les propriétés qu'elle implique sont d'ordre supérieur par rapport au niveau nomologique fondamental. Dans le cadre de la loi

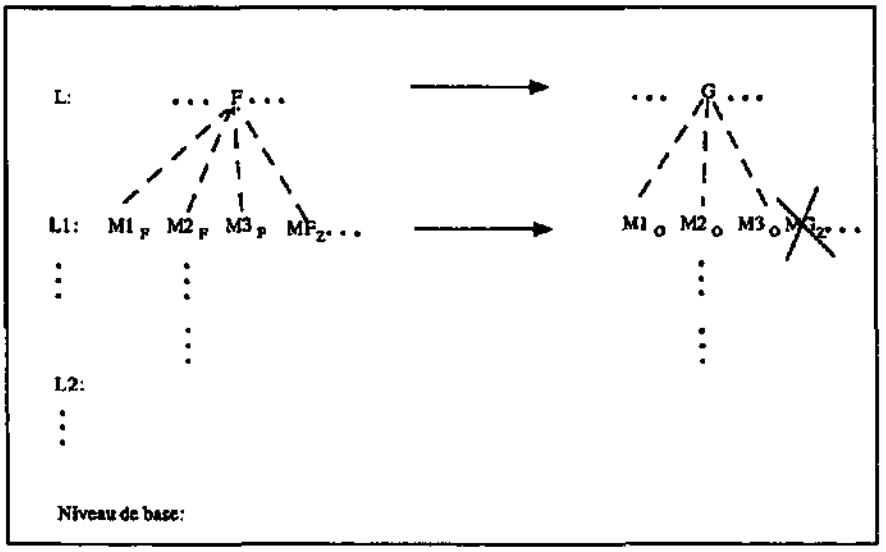

Figure 7.1: Le schéma montre la relation entre l'implantation d'une loi (F cause G) et le mécanisme d'implantation. Dans ce cas on suppose que les deux propriétés F et G sont multiréalisées par le niveau immédiatement inférieur ($L1$). Par exemple, la loi "F cause G" est implantée par le mécanisme qui obéit à la loi $M1_F$ cause $M1_G$. Cette dernière loi est à la fois implantée par des mécanismes du niveau inférieur, et ainsi de suite jusqu'au niveau de base. Aussi on peut voir que l'état mental de MF_Z qui est une des instantiation de la propriété F n'a pas de correspondant MG_Z comme contrepartie dans la loi "$F \rightarrow G$"

"$F \rightarrow G$" ceci serait expliqué par une occurrence de MF que j'appellerais MF_Z telle qu'il n'existe aucune loi qui relie MF_Z avec une des occurrences de G du niveau réducteur (voir figure 7.1). Cela veut dire que MF_Z, bien qu'étant une instantiation de M_F ne fait pas partie de la disjonction qui sert d'antécédente à la loi. A ce moment-là toute instantiation de F par MF_Z sera une exception à la loi de niveau supérieur "$F \rightarrow G$".

J'aimerais donner un exemple d'exception à des généralisations en utilisant un concept de la psychologie expérimentale animale: le concept d'inhibition latente. L'inhibition latente de l'attention est un phénomène étudié dans le cadre du behaviorisme qu'il dépasse pour s'intégrer à la psychologie cognitive contemporaine. Selon Robert Lubow qui est une des autorités dans ce domaine l'inhibition latente a trois caractéristiques: premièrement les conditions qui la produisent, deuxièmement les conditions que l'on utilise pour mesurer cet effet et finalement la direction des différences trouvées dans les deux groupes (le groupe testé et celui de contrôle).

> **More specifically, latent inhibition is the detrimental effet of passive, nonreinforced preexposure of a stimulus on the subsequent ability of an organism to form new associations to that stimulus. To demonstrate latent inhibition, one must preexpose one group of subjets to the stimulus of interest, while not giving such stimulus preexposure to a control group. In the test phrase, both groups must learn to form an association between that stimulus and a new event. When the stimulus-preexposed group learns the new association to that stimulus more poorly than does the control group, we say that latent inhibition has been demonstrated. [Lubow, 1989, page 1]**

Une expérience très répandue consiste à prendre deux populations de souris que je vais appeler P et NP. La population P est pré-exposée à l'eau empoisonnée avec une substance toxique qui quoique n'étant pas mortelle cause des troubles d'ordre digestif. La population NP en revanche n'est pas pré-exposée. Une fois que le temps de pré-exposition est passé, les deux populations sont soumises à une période de privation d'eau. La phase de transfert commence lorsqu'on expose les deux populations à l'eau. La population NP va développer des comportements consistant à "boire de l'eau" bien avant la population P. On dit que la pre-exposition à l'eau empoisonnée a créé une inhibition de la conduite pertinente dans le cas d'un longue privation de l'eau.

En général, il y a une loi *ceteris paribus* qui dit :

Les mammifères lorsqu'ils ont soif et perçoivent de l'eau dans leur environnement, boivent de l'eau (ceteris paribus).

Dans le schéma suivant 7.8 on voit une réalisation de cette loi. Les termes M_p, M_b représentent les états mentaux de privation d'eau (ou d'avoir soif) et le désir de boire respectivement. Les termes qui commencent avec N correspondent à l'état physique qui les réalisent.

$$\begin{array}{ccc} M_p & \longrightarrow & M_b \\ \uparrow & & \uparrow \\ N_p & \longrightarrow & N_b \end{array} \tag{7.8}$$

Pourquoi le groupe P n'a-t-il pas vérifié cette loi?

Parce qu'ils ont été exposés à l'eau empoisonnée et malgré le fait que la propriété caractérisée comme "avoir soif et percevoir de l'eau dans l'environnement" soit instantiée chez ces individus, la propriété fonctionnelle "dégoûté de l'eau" (dans le schéma 7.9 M_e) est aussi instantiée au niveau physique. Or, l'état physique (dans le schéma 7.9 N_x) ne sera pas un antécédent de la loi stricte citée parce que la propriété "dégoûté de l'eau" est elle aussi instantiée d'où l'exception à la loi de niveau supérieur s'explique niveau neurologique.

$$\begin{array}{c} (M_e) \wedge M_p \\ \downarrow \\ N_p \wedge N_e = N_x \neq N_P \end{array} \tag{7.9}$$

L'intérêt de cet exemple se base sur le fait que lorsque les sujets P sont traités avec des amphétamines, l'inhibition latente n'est plus observée (dans le schéma 7.10 représenté par N_a).

L'explication est que l'amphétamine est un antagoniste de la dopamine qui est la cause de l'inbition latente chez les sujet du groupe P.

$$\begin{array}{ccc} (M_e) \wedge M_p & \longrightarrow & M_b \\ \uparrow & & \uparrow \\ N_p \wedge N_e \wedge N_a = N_p & \longrightarrow & N_b \end{array} \tag{7.10}$$

L'amphétamine rend l'état physique du cerveau comme étant une instantiation d'un des antécédentes de la loi fondamentale (voir figure 7.10).

Finalement j'aimerais parler des critiques que l'on pourrait adresser à cet exemple. Une des critiques possibles est de dire que j'attribue aux souris des capacités intentionnelles, ce qui n'est pas juste. Je vais rester neutre vis-à-vis de cette question. L'unique attribution que j'ai faite concerne les propriétés supérieures au niveau physique mais dans la pyramide des propriétés de L-ordre je ne suis pas obligée d'arriver au niveau intentionnel pour donner un exemple de propriétés de niveau supérieur qui ne vérifient pas les généralisations.

Résumons: Pour Fodor les lois strictes appartenant aux sciences fondamentales n'ont pas d'exception. Si une exception existe, la généralisation soumise à cette exception est d'un ordre supérieur au niveau fondamental et cette exception est explicable en vertu de leur réalisabilité multiple. Alors pour Fodor, l'explication d'une exception ne serait que l'absence des lois strictes qui font le lien causal entre l'objet instantié par l'antécédent et non une exception aux lois fondamentales.

L'implantation computationnelle des lois intentionnelles: Fodor propose que les mécanismes d'implantation des lois intentionnelles sont computationnels. Nous avons montré dans des sections précédentes les motivations de cette décision: si le mécanisme d'implantation est computationnel nous avons une chance de conserver le parallélisme syntaxique-causal, c'est-à-dire

que les différences sémantiques seront reflétées comme des différences formelles. Autrement dit, la transformation d'un symbole en un autre permet de sauvegarder la relation causale entre les symboles et le monde. [Fodor, 1994, cf. page 12]

Néanmoins, il n'est pas aussi simple de justifier métaphysiquement la réalisation des propriétés intentionnelles que celle des propriétés computationnelles lorsque l'on utilise une sémantique informationnelle.

Où est le problème? Rappelons que les cooditions que l'on doit remplir pour justifier l'existence d'un mécanisme d'implantation sont au nombre de trois (voir schema 7.7): premièrement que les Fs causent Gs, deuxièmement que F est une condition suffisante pour MF, et troisième que MG est une condition suffisante pour G.

Comme le dit Fodor:

> If the implementing mechanisms for intentional laws are computational, then we need a property theory[35] that provides for *computationally sufficient condition* for the instantiation of intentional properties and vice versa. This, however, implies a nasty dilemma. [...] you can't get computational implementations for intentional laws unless there can be both computationally sufficient conditions for the satisfaction of computational properties. If this dilemma can't be broken, it looks as though the usual constraints on property theories aren't satisfiable in the case of intentional laws. So maybe intentional properties aren't informational after all or maybe psychological laws aren't intentional after all; or maybe the implementation of psychological laws isn't computational after all. [Fodor, 1994, page 12–13]

Nous avons vu dans les sections précédentes que la réalisation des lois nécessite que cette relation de réalisation entre les propriétés intentionnelles et les propriétés computationnelles soit suffisante dans les deux sens. Cette dernière condition implique que les propriétés computationnelles et les propriétés intentionnelles entretiennent une relation de double conditionnelle, ce qui équivaut à dire qu'il doit exister une corrélation entre elles.

Cependant, cette condition de corrélation n'est pas compatible avec le postulat qui prône que la sémantique est informationnelle.

Les deux problèmes suivants réfutent la nécessité métaphysique d'une telle corrélation. Un de ces problèmes est le problème de Frege et l'autre est le problème inverse dont une des propositions est le problème des Terres Jumelles.

Je vais montrer comment Fodor a essayé de récuser les deux problèmes tantôt en soutenant une dualité des contenus, tantôt en octroyant des caractéristiques différentes aux mécanismes d'implantation.

Fodor admet finalement en 1994 que la corrélation des propriétés intentionnelles prises dans une perspective informationnelle est contingente sans être exceptionnelle. En général, il s'agissait pour Fodor de changer les modes d'individualisation des état mentaux ou des mécanismes d'implantation dans le but de trouver des conditions plus *fines* ou plus adéquates d'individualisation pour augmenter ainsi les chances d'une corrélation nécessaire.

Je vais montrer comment Fodor a essayé de réfuter les deux problèmes et ensuite que la cohérence entre les trois postulats fodoriens est possible seulement au prix de renoncer à la nécessité métaphysique entre les propriétés intentionnelles et les propriétés computationnelles. Cette corrélation requise étant contingente ne nous laisse pas affirmer que les lois psychologiques se multiréalisent computationnellement. Je pense qu'il sera plus pertinent de parler d'une *pseudo*-multiréalisation.

7.5.2 La sémantique est purement informationnelle.

Fodor défend maintenant une position ouvertement informationnelle des contenus [36] à partir de 1993 [Fodor, 1994, page 4–5] et soutient aussi l'approche de Fred Drestke [Drestke, 1981].

Mais l'externalisme de Fodor est nuancé:

> The contents of a thought depends on its *external* relation; on the way that the thought is related to the world, *not on the way that it is related to other thoughts*. [Fodor, 1994, page 4]

[35] Il faut comprendre ce terme comme faisant référence à une théorie des propriétés telle que celle de Robert Cummins citée dans le chapitre 2 (§2.3.4.) lors de la discussion du concept de réalisation

[36] Voir chapitre 4 (§4.4.2.) pour une description de la théorie informationnelle.

Pourquoi Fodor n'a-t-il pas, à ma connaissance, adopté la position informationnelle des contenus plus tôt? Le problème est qu'une position informationnelle des contenus implique la réfutation de l'existence des contenus étroits étant donné que la détermination de ces derniers se fait en vertu des leurs rôles fonctionnels et non en fonction des structures informationnelles déterminées par l'extérieur. D'autre part il semble que seuls les contenus étroits auront une chance de satisfaire à la nécessité des corrélations avec les structures computationnelles et les propriétés intentionnelles. Maintenant, je me propose d'élucider ces deux problèmes.

Le problèmes que posent les Terres Jumelles et le problème de Frege: Le problème des Terres Jumelles démontre qu'il est possible d'avoir deux occurrences d'états mentaux différents du point de vue des contenus larges qui réalisent le même ensemble de propriétés computationnelles puisqu'ils sont identiques du point de vue de leurs rôles causaux. Ainsi les propriétés computationnelles ne seront pas des conditions suffisantes pour la détermination des propriétés intentionnelles comme par exemple la référence. La corrélation entre les propriétés intentionnelles et les propriétés computationnelles ne tient pas. En effet, l'implantation computationnelle identique ne suffit pas à assurer l'identité (intentionnelle) des contenus.

Le problème de Frege, en revanche indique que deux implantations computationnelles différentes sont des réalisations du même état intentionnel vu comme contenu large. En effet, "croire que Venus est une planète" a (au moins) deux réalisations différentes, d'un côté "croire que l'étoile de matin est une planète" et de l'autre "croire que l'étoile du soir est une planète" puisque les deux réalisations auront des rôles causaux différentes.

Or, l'implantation intentionnelle des contenus larges différents ne suffit pas à assurer la différence des implantations computationnelles.

Les conférences parisiennes *Jean Nicod* ont signé l'arrêt de mort du contenu étroit (de moins pour tous les cognitivistes de ce côté de l'Atlantique), mais la théorie de Fodor a dû évoluer pour réfuter le dualisme de contenu. Je vais retracer ce parcours avant de signaler que Fodor avait une autre solution qui n'obligeait pas à récuser ce dualisme.

D'une théorie dualiste des contenus à une théorie des contenus larges J'ai déjà exposé dans le chapitre 4 (§4.4.1.) les concepts de contenu étroit et de contenu large. Nous avons vu que Fodor récuse le contenu large en arguant qu'il n'est pas pertinent dans les explications en psychologie.

Théorie dualiste des contenus avant 1989 Dans le texte [Fodor, 1980] Fodor avait commencé à relativiser le problème des Terres Jumelles en montrant que pour des raisons pratiques la psychologie ne pouvait pas répondre aussi au problème sémantique du contenu. L'unique domaine de travail était le *solipsisme méthodologique*. La psychologie aura comme théorie des contenus une théorie des contenus étroits.

Dans son livre *Psychosemantics*[37] décide de jeter sur le problème un nouvel éclairage. Pour y parvenir il établit une différence entre les concepts de *solipsisme méthodologique* et d'*individualisme méthodologique* d'un côté et des propriétés relationnelles et non relationnelles de l'autre.[Fodor, 1987, cf. 42–ss]. Cette distinction sert à démontrer que la contradiction entre les deux contenus n'est qu'apparente.

Le *solipsisme méthodologique* est une théorie selon laquelle l'individualisation des contenus est syntaxique, donc non-relationnelle, c'est-à-dire que l'individualisation se fait de manière indépendante de leur évaluation sémantique. L'*individualisme méthodologique* est un principe général de méthodologie scientifique selon laquelle l'individualisation des objets pertinents dans une science se fait en fonction des propriétés qui peuvent jouer un rôle causal.

L'individualisme méthodologique en psychologie n'interdit donc pas la possibilité d'individualiser les contenus relationnellement, les propriétés de l'individualisation peuvent parfaitement être utilisées si elles s'avèrent pertinentes du point de vue de l'explication causale.

[37] [Fodor, 1987]

Or il n'y aurait pas de contradiction, mieux il se peut que de façon contingente les deux manières d'individualisation, la solipsiste (syntaxique) et l'individualiste (sémantique) coïncident.

Fodor a besoin d'établir une relation entre les causes du comportement (son rôle fonctionnel) et son contenu. Il s'agit de trouver une façon d'individualiser les états mentaux qui soit assez fine pour éviter que deux contenus larges différents aient le même rôle causal (Terre Jumelles) ou qu'un même contenu large donne lieu à deux rôles causaux différents (problème de Frege).

Dans le but de surmonter ce problème Fodor postule dans [Fodor, 1987, page 48] que l'individualisation du contenu ne se fait qu'en relation au contexte.

Qu'est-ce que Fodor veut dire lors qu'il plaide pour une relation avec un contexte? Le contexte fixe certaines conditions qui définissent l'existence de certaines relations causales entre occurrences d'un référent et occurrences d'un type d'état mental.

Il va plus loin, il postule que la relation qui détermine l'individualisation des contenus est simplement l'inverse d'une relation causale. C'est le contexte qui donne l'antécédent des relations causales sur lesquelles se base la référence.

> But now we have an extensional identity criterion for mental contents: Two thought contents are identical only if they effect the same mapping of thoughts and contexts onto truth condition. Specifically, your thought is content-identical to mine only if in every context in which your thought has truth condition T, mine has truth condition T and vice versa. [Fodor, 1987, page 48]

Cela veut dire que, dans le cas des Terres Jumelles dans la situation suivante:

> "Si mon double moléculaire croit qu'il y de l'eau dans la gourde alors il aura le désir de boire une gorgée"

La référence du contenu "eau" sera déterminée par l'inverse de la relation causale que je viens de décrire mais étant donné que mon double moléculaire se trouve dans la Terre Jumelle le contexte veut que les occurrences du référent "eau", chaque fois qu'il est antécédent causal de la proposition que je viens de citer plus haut soit toujours XYZ.

Cependant cette solution n'est pas encore satisfaisante, comme le signale Pacherie

> Si "détermination est pris en ce sens, le contenu étroit n'a pas à inclure une définition de l'extension, ce qui fait qu'il détermine l'extension est simplement qu'il est le pivot d'une double relation. [...]"
>
> Néanmoins, si l'existence de relations de référence entre des états mentaux et des entités du monde externe doit être expliquée par l'existence de relations causales entre ces entités et les états mentaux, tout le succès de l'entreprise repose désormais sur la capacité du théoricien à énoncer les contraintes qui portent sur ces relations causales. De toute relation causale n'émerge pas une relation de référence. Est-il possible d'énoncer des contraintes qui permettent d'écarter les relations causales non pertinentes? Deuxièmement, qu'advient-il de la nécessité méthodologique autrefois proclamée par Fodor d'avoir une caractérisation des objets de pensée préalablement à toute tentative de caractérisation des relations causales entre ces objets et les pensées qui y font référence? [Pacherie, 1993, page 186]

Théorie de contenus jusqu'à 1993: En 1989 Fodor semble abandonner la position que je viens de décrire selon laquelle le référent serait déterminé par l'inverse d'une relation causale dans un contexte. Parmi d'autres raisons, cette solution ne s'applique pas au problème de Frege. En effet, si les relations causales prennent comme antécédent la référence et non le sens, ledit problème continue à se poser.

Il choisit donc une autre stratégie toujours dans le but de montrer une corrélation entre les propriétés intentionnelles et les propriétés computationelles. Il décide de rendre plus explicites les conditions d'implantation en jouant sur l'individualisation des attitudes propositionnelles.

Fodor en *Psychosemantics* concevait les attitudes propositionnelles comme une relation binaire de nature computationnelle entre le sujet et la représentation. [Fodor, 1987, cf. page 17]

En 1989, il propose une analyse différente des croyances et au lieu de les considérer comme des relations binaires, il propose une forme de relation quaternaire: le sujet, la proposition (le contenu propositionnel est le contenu large), le véhicule et le rôle fonctionnel. [Fodor, 1989b, cf. page 167]

La conception des attitudes propositionnelles comme relations binaires restreint les possibilités d'individualisation de ces attitudes. Dans ce cas, les attitudes propositionnelles auront seulement

deux *degrés de liberté* pour leur individualisation : soit elles sont différentes par leurs *modes* (les croyances sont différentes des désirs), soit elles le sont par leurs *propositions* (croire en P est différent de croire en Q). Bien entendu, elles peuvent être différentes en référence aux deux caractéristiques simultanément.

Le problème de cette individualisation lorsque l'on considère les attitudes propositionnelles binaires est qu'elle ne s'avère pas assez fine pour distinguer des croyances qui ont comme contenu les mêmes propositions du point de vue du sens.

Comme Fodor le remarque:

> The point of telling you this story is that since such cases are allowed, the proposition that *J*[ocaste] is eligible might turn out to be identical to the proposition that *O*[edipe]'s M[other] is eligible *even though* believing the one proposition is a different state from believing the other. But if these propositions may be the same then we have, so far, no reason to doubt that '*J*' and '*O*'s *M*' are synonyms. Which is to say that, at least so far as the facts about Oedipus are concerned, we have no reason to doubt that denotation is all that there is to meaning. [Fodor, 1989b, page 166]

La conception des croyances comme des relations à quatre éléments, en revanche, nous confère deux degrés de liberté supplémentaires. Les croyances d'un même agent peuvent différer en relation avec leur contenu, à leur véhicule ou à leur rôle fonctionnel. Mais qu'est-ce que le *véhicule*?

Le concept de *véhicule* est foncièrement en relation avec le langage de la pensée dans le cadre fodorien. Évidemment, il ne s'avère pas applicable à une autre théorie mais il fait référence au mécanisme d'implantation fodorien.

Le véhicule est un symbole mental qui correspond au concept de *mode de présentation* dans le cadre sémantique[38].

Comment deux véhicules se distinguent-ils l'un de l'autre? Deux véhicules concrets particuliers (*tokens*) appartiennent à des types distincts si ils sont différents selon leur syntaxe ou s'ils servent à exprimer des contenus propositionnels différents.

> A vehicle is a symbol. A symbol (token) is a spatiotemporal particular which has syntactic and semantic properties and a causal role. Vehicles, like other symbols, are individuated with respect to their syntactic and semantic properties, but *not* with respect to their causal role. In particular, two vehicle tokens can differ in their causal roles because the role that a token plays depends not just on which type it's a type of, but also on the rest of the world in which is tokening transpires. (This is true of the causal roles of symbols because it's true of the causal roles of everything. Roughly, your causal role depends on what you are, what the local laws are, and what else there is around). I assume, finally, that vehicles can be distinct but synonymous; distincts vehicles can express the same proposition. [Fodor, 1989b, page 167 ma souslignation]

Ainsi, dans la citation on voit qu'il est possible d'avoir deux véhicules concrets particuliers appartenant à un seul et même type qui auront des rôles fonctionnels distincts s'ils appartiennent à des systèmes de croyances différents. Néanmoins, le cas réciproque se vérifie également; deux véhicules concrets particuliers appartenant à des types distincts (leurs propriétés syntaxiques et sémantiques sont différentes) peuvent avoir un seul et même rôle fonctionnel. [Jacob, 1993, cf. 152–153]

Le concept de véhicule permet à Fodor de donner une réponse au problème de contexte opaque. Maintenant, bien que 'Jocaste' et 'la mère d' Œdipe' soient des synonymes, le *désir de marier Jocaste* diffère du désir *de marier la mère d'Œdipe* non en fonction de ses propriétés sémantiques mais en fonction de son rôle fonctionnel. Or le véhicule va refléter ces différences au niveau syntaxique.

L'introduction du concept de véhicule en tant que terme additionnel des attitudes propositionnelles est un effort (désespéré?) de Fodor pour ne pas renoncer à une théorie des contenus étroits.

A partir de 1993, Fodor les définit comme des relations tertiaires entre le sujet, la proposition et le mode de présentation. Je reviendrai sur les motivations de ce dernier changement.

Les problèmes que pose la notion de *véhicule* sont identiques aux difficultés d'obtenir des conditions suffisantes au niveau de l'implantation computationnelle du niveau intentionnel et vice versa tout simplement parce que le véhicule n'est qu'une partie centrale du mécanisme d'implantation.

[38] Pour une définition de *mode de présentation* voir chapitre 4

A mon avis, cette introduction peut ménager une échappatoire au problème fregéen mais ne donne aucun éclaircissement sur l'individualisation des contenus eux-mêmes. On a beau dire que le véhicule reflète des propriétés sémantiques en fonction des propriétés syntaxiques mais on ne sait toujours pas comment le véhicule y parvient. La différence existant entre deux véhicules, bien qu'il expriment des éléments co-référeotiels, relève du miracle. L'unique façon de justifier cette *révélation* serait de franchir le pas vers le monde. On sait que Fodor y a toujours été très réfractaire et qu'il s'est refusé à formuler une théorie de la perceptioo. Néanmoins, le Fodor de 1993 a abandonné le contenu étroit en disant qu'il ne le répudie pas mais simplement qu'il le considère comme superflu. [Fodor, 1994, cf. page 28].

Il propose donc une théorie des contenus larges et soutient une sémantique de type informationnel. Cependant, la construction des théories informationnelles ne semble pas être au centre de ses recherches et il tient les efforts de Drestke pour valides.

La récusation du dualisme de contenus de 1993: Dans les paragraphes précédents nous avons vu qu'il est très difficile voire impossible de trouver des conditions suffisantes de corrélation entre les propriétés intentionnelles et les propriétés computationnelles lorsque l'on prend une individualisation large des contenus. Bien entendu, si vous n'avez que des contenus étroits il n'y a pas de problème; mais j'ai déjà exposé quelques objections à cette position fondée -encore une fois- sur le problème des Terres Jumelles et sur le problème fregéen.

Ces objections montrent les difficultés auxquelles on se heurte en soutenant la réalisation computationnelle des lois intentionnelles dans le cadre d'une sémantique informationnelle.[39]

Bien que l'on ne puisse affirmer la nécessité métaphysique de la corrélation recherchée, Fodor pense qu'on peut reconnaître que ces deux situations ne se produisent qu'accidentalement.[40]

En effet, s'il en est ainsi, Fodor dira que le contenu étroit coïncide avec le contenu large dans la plupart des cas et ainsi qu'on pourra réconcilier les présomptions que les lois en psychologie sont intentionnelles, que l'implantation immédiate du niveau intentionnel est computationnelle et que la sémantique est de type informationnel.

Malheureusement, les deux problèmes cités (les Terres Jumelles et le problème fregéen) montrent que cette corrélation est *contingente*, ou qu'autrement dit, la corrélation n'est pas métaphysiquement nécessaire. L'espoir de Fodor est le suivant: si l'on arrive à prouver le caractère accidentel de ces exemples alors cette corrélation, bien que métaphysiquement contingente, s'avèrera au moins digne de foi. [Fodor, 1994, cf. page 25]

Fodor nous signale que dans la vie de tous les jours il y a beaucoup de propriétés contingentes et que malgré cela elles vont presque toujours la main dans la main. L'exemple qu'il donne concerne les propriétés suivantes: "être un dollar". et "ressembler à un dollar". Ces propriétés ne sont pas métaphysiquement co-extensionnelles, mais la structure de la société notamment le gouvernement américain et les autres gouvernements du monde en général s'occupent de les faire coïncider le plus possible par exemple en punissant les fraudes.

Fodor suggère qu'il n'existe pas de mécanisme semblable pour maintenir en corrélation les propriétés du contenu large intentionnel avec les propriétés computationnelles. [41]

---

[39] On peut assimiler les aspects les plus saillants de la sémantique informationnelle à ceux des contenus larges. J'ai discuté de cela dans le chapitre 4 lorsque j'ai exposé la théorie de Fred Dretske.

[40] Il faut comprendre *accidentel* dans le sens de *non-systématique*.

[41] Fodor rejette toutes les explications darwiniennes à ce sujet. Il dit:

> "You see," they explain, "all creatures whose computational and intentional properties weren't properly in phase died long ago" But no. If you think that's the answer, you haven't understood the question (or you haven't understood Darwin).
>
> No doubt, if there is something in situ that coordinates the intentional properties of mental states with their computational properties, then some Darwinian process must have selected it, and its having been selected explains why there is now so much of it around. But the question before us is *what the mechanism that effets this correlation is*, and evolutionary explanations *aren't of the right form* to answer that kind of question. Evolution maybe explains why there are more things around that work that there are things that don't. But it doesn't explain *how* thing work [...]. So, please, spare me; no Darwin. [Fodor, 1994, page 20]

Une des tentations est d'avoir recours aux *lois-ponts* comme on fait lorsqu'on veut construire une correspondance entre les termes d'une théorie et les termes d'une autre que l'on considère réductrice.[42] En ce faisant, ces lois pourraient établir des relations nomologiques entre les propriétés du niveau computationel et celles qui sont issues d'une sémantique informationnelle. Toutefois, pour Fodor il faut abandonner cette tentation qui n'aboutit qu'à une fuite en avant car ni les propriétés computationnelles ni les propriétés intentionnelles n'appartiennent à une science fondamentale. Or il y aura des exceptions, les corrélations ne sont que métaphysiquement contingentes et nous voilà donc revenus au point de départ.

La stratégie de Fodor est autre. Elle consiste à montrer que ce mécanisme qui assure l'existence de la corrélation entre les propriétés en question est digne de foi; bien que contingent du point de vue métaphysique il ne s'avère pas néanmoins accidentel.[Fodor, 1994, cf. page 25]

La stratégie visant à mettre ce fait en évidence change selon qu'il s'agit des Terres Jumelles ou du problème de Frege (dont Œdipe par exemple a tant souffert). Dans les deux cas néanmoins, l'argumentation qu'il nous offre n'est pas philosophique mais simplement consiste en des constats empiriques. Comme Fodor le dit:

> An impure philosopher might be curious which, if either, of these ways of answering E[ponymous] Q[uestion][43] is the right one. I am an impure philosopher by these standards; perhaps by any. Metaphilosophical scruples to the wind, I therefore propose to argue, in this lecture, that it is *plausible*– not unreasonable to believe– that the world is so organized as to prohibit the proliferation of Twin cases and Frege cases; hence that, for all we know, the laws of intentional psychology may well be broad. To argue this requires finding something –some mechanism short of a miracle–that might serve to keep broad content and computational role in harmony. [Fodor, 1994, page 28, les soulignés en italique se trouvent dans le texte original]

J'ai déjà signalé que pour Fodor les exemples du type Terres Jumelles ne sont pas de véritables objections à l'existence d'une telle corrélation, au contraire des exemples de Frege qui seraient des vraies objections si de tels cas s'avèraient systématiques. Il s'agit donc, d'établir *empiriquement* le caractère accidentel de ces derniers. Ainsi, si l'on est d'accord sur leur caractère accidentel, alors rien ne serait perdu. Il sera possible d'établir des corrélations entre des propriétés intentionnelles et des propriétés computationnelles qui admettent des exceptions seulement si ces dernières sont accidentelles.

L'orme, l'expert, Fodor et ma jumelle moléculaire : Selon Fodor tous les exemples formulés à partir des expériences de la pensée du type des Terres Jumelles ne réfutent pas le caractère accidentel de ces faits. L'objectif essentiel de la démarche fodorienne est de montrer que de façon empirique seulement le caractère accidentel voire non systématique voire non nomologique des cas où le contenu large représente les rôles fonctionnels n'est pas en corrélation avec le contenu étroit.

Mis à part le fait que l'existence de la Terre Jumelle est impossible, cette expérience de la pensée révèle seulement que la relation de survenance des contenus intentionnels larges sur le niveau computationel n'est point *conceptuellement* (voire métaphysiquement) nécessaire.

Cependant, il ne réfute point l'existence d'une survenance nomologique; cela veut dire que rien ne s'oppose dans cette argumentation à l'existence de lois empiriques servant à maintenir cette relation.

Terres jumelles :

Pour expliquer ce cas, Fodor met au point une stratégie d'argumentation en deux étapes. D'abord il essaie de montrer que les croyances eo l'eau et en l'eau-jumelle ne sont pas incompatibles avec un théorie informationnelle des conteous et ensuite il relativise l'impossibilité d'une théorie qui représente ce type de cas. En effet, étant donné que l'exemple est nomologiquement impossible et que ce type de situation est empiriquement accidentel ou non–nomologique, l'impossibilité de représentation de sa théorie ne semble pas bien grave.

[42] Voir chapitre 2.

[43] La *Eponymous Question* à laquelle Fodor fait référence est: "comment peut–on renconcilier l'idée que les lois psychologiques sont foncièrement intentionelles avec l'idée que leur implantation est foncièrement computationnelle? ".

Dans le cadre d'une sémantique informationnelle on accepte la propriété disjonctive des contenus. On se souviendra qu'au chapitre 4 nous avons vu que la théorie causale soutient que la relation de référence entre l'objet ou la propriété représentée et sa représentation se trouve être l'inverse de la relation causale. Mais pour justifier les cas de mereprésentation on accepte comme propriété représentée la disjonction du référent réel et du contenu mereprésenté. Ainsi si je méprends une vache pour un cheval le contenu de ma représentation sera "soit ceci est une vache soit ceci est un cheval".[44]

Dans le cas de Terre Jumelle, il s'agit de dire que les propriétés que ma jumelle représente sont applicables "soit à l'eau, soit à XYZ". Alors de ce point de vue il n'y a pas de problème car la faille dans la théorie n'est pas là. Cependant, l'impossibilité pour ma jumelle de faire la distinction entre H_2O et XYZ n'est qu'accidentelle puisqu'il n'y a aucune loi que le lui interdise. Le problème véritable se situe au niveau de l'application des propriétés. En effet, si les substances sont accidentellement impossibles à distinguer pour moi et pour ma jumelle, toutes les propriétés qui s'appliquent à XYZ, s'appliquent aussi à H_2O mais avec la référence XYZ. Tandis que je fais exactement le contraire, ma référence étant H_2O. C'est cela qui pose le problème en raison des réalisations computationnelles et physiques[45] qui sont les mêmes chez elle et chez moi mais du point de vue des contenus larges les états intentionnels sont différents. Elle applique les propriétés au concept XYZ et par extension de la propriété disjonctive au concept H_2O et moi je fais exactement le contraire ce qui en fait revient au même en vertu de la propriété disjonctive. Or, las réalisations computationnelles ne donnent pas des conditions suffisantes.

C'est cette généralisation que la théorie de psychologie fodorienne n'arrive pas à représenter; mais selon Fodor ce n'est pas grave parce que ce ne *sont* que des occurrences accidentelles et *non* nomologiquement vraies donc non systématiques.

L'orme et l'expert :

Dans le problème qui a été posé aussi par Putnam [Putnam, 1975a], la situation est la suivante : je ne suis pas capable de faire la distinction entre les ormes et les hêtres. Ainsi ce n'est pas moi qui détermine les conditions de vérité de ma pensée en référence à ces deux types d'arbres et il me faut demander les services d'un expert.

Or tous mes états computationnels sont les mêmes vis–à–vis des deux arbres. Fodor répond que l'on confond deux aspects différents du problème : l'aspect épistémologique et l'aspect sémantique. Pour lui le fait d'établir une différence entre les deux concepts est un problème épistémologiquement pertinent mais par contre n'a aucune importance en sémantique.

Selon Fodor cet exemple illustre que la réalisation computationnelle ne suffit pas à la détermination du contenu large mais à la différence du cas des Terres Jumelles ce type de situation n'est pas accidentel et en fait il est commun dans la vie quotidienne.

Ensuite il donne des exemples de situations semblables mais dans lesquelles, au lieu de se référer à un expert on utilise d'autres moyens, p. ex. "Je ne sais pas la date d'aujourd'hui, c'est pourquoi je consulte le calendrier". L'expert n'est qu'un instrument comme n'importe quel autre que je consulte *seulement* si j'en ai besoin. Ceci dédramatise la situation et montre que ces banales circonstances ne contredisent en rien la théorie sémantique informationnelle, puisque selon celle–ci la détermination de l'identité des concepts se base non pas sur les concepts que je puis distinguer entre eux mais sur ceux dont je *pourrais* faire la différence si je le voulais.

Fodor fait ensuite une différence (implicite) entre les concepts de contenu et les autres concepts sémantiques comme celui de valeur de vérité auxquels on a recours lorsqu'on se place dans une perspective externaliste. La référence fregéenne est une chose et la construction des contenus qui font appel à cette référence en est une autre.

Il nous rappelle que d'un point de vue externaliste, la sémantique ne fait pas partie de la psychologie. C'est justement pour cela que l'idée selon laquelle les lois en psychologie sont intentionnelles est compatible avec l'idée que les processus mentaux sont computationnels avec des contenus pris dans le sens large.

[44] Voir dans le chapitre 4 le problème de la disjonction.

[45] Il ne faut pas oublier que ma jumelle et moi nous sommes identiques molécule à molécule

> [...] it is of the essence of semantic externalism that there is *nothing that you have to believe; what inferences do you have to accept* in order to have the (deferential) concept ELM.[46] And it is of the essence of semantic externalism that there is *nothing that you have to believe, there are no inferences that you have to accept*, to have the concept ELM. According to externalism, having the concept ELM is having (or being disposed to have) thoughts that are causally (or nomologically) connected, in a certain way, to instantiated elmhood. Punkt. It is to put the point starkly, the heart of externalism that *semantics isn't a part of psychology*. The content of your thoughts (/utterances), unlike, for exemple, the syntax of your thoughts (/utterances), does not supervene on your mental processes. [...] My present brief is not, however, to reconcile you to externalism. It is rather just to convince you that psychological processes could be computational even if externalism is true and intentional laws are therefore broad. [Fodor, 1994, page 37–38, les italiques font partie du texte original.]

Selon Fodor la condition de réfutation de la corrélation entre le contenu large des états intentionnels est la suivante:

> It is nomologically possible that there are creatures for which it is nomologically impossible to distinguish between *a*s and *b*s, but of which an externalist theory of content is required to say that they have the concept *A* but don't have the concept *B*. The mental states of such creatures, though 'broadly' distinct by assumption, would be 'narrowly' identical by nomological necessity. A fortiori, a purely broad psychology couldn't articulate the intentional laws that they fall under. [Fodor, 1994, page 38,les italiques appartiennent au texte original.]

En définitive, le cas des Terres Jumelles ne vérifie pas le postulat précédent parce que l'existence de mon double moléculaire sur une Terre Jumelle telle qu'on l'a décrite est nomologiquement impossible. Idem pour l'orme parce que des êtres pourront faire la distinction s'ils le souhaitent en faisant appel à un expert. Selon Fodor on ne va pas trouver des exemples d'êtres qui vérifient la condition de réfutation. Cependant, ceci n'est qu'un constat empirique. On n'est pas en mesure d'affirmer qu'à l'avenir on ne construira pas des engins répondant à cette condition ou que sur Mars de tels êtres n'existent pas déjà.

7.5.3 La véritable objection:

Le tort immense d'Œdipe. Le problème d'Œdipe constituerait une vraie objection s'il s'avèrait être plus qu'accidentel. Rappelons qu'Œdipe veut épouser Jocaste malgré le fait que l'idée d'épouser sa mère l'horrifie; or il se trouve que Jocaste et la mère d'Œdipe sont une même personne. L'idée de Fodor est de donner les raisons qui rendent les situations de ce type accidentelles.

Le cas d'Œdipe est vraiment une exception et les exemples de ce genre ne sont pas admis au nombre des lieux communs.

Les raisons que Fodor invoque sont de deux types différents. D'abord celles qui ne font pas référence au cas particulier d' Œdipe. Dans ce premier groupe Fodor énonce le *Principe d'équilibre épistémique* (*PEI*): les sujets entretiennent un équilibre épistémique vis–à–vis des faits importants pour leurs actes. Le *PEI* se base sur le postulat (*truism*) suivant:

1. *T*1: l'agent ne peut pas choisir entre *A* et *B* à moins qu'il ne connaisse des faits qui lui font choisir l'un plutôt que l'autre.

2. *T*2: Le succès des actions est non–accidentel à moins que les croyances de l'agent ne s'avèrent vraies.

Telle est pour Fodor la base du succès des actions rationnelles. Ainsi c'est une aberration qu'Œdipe ait décidé d'épouser Jocaste. La croyance que Jocaste n'est pas sa mère est fausse mais ceci va contre le principe *T*2. La preuve que le *PEI* et les conditions qu'il implique sont vraies est que les prédictions faites sur la base de désirs et de croyances se vérifient si bien. Ce que Fodor veut dire, en gros, c'est que lorsqu'une information est pertinente au succès d'une action on la connait généralement.

[46] Pour Fodor les mot écrits en majuscules signalent les concepts auquels ces mots s'appliquent.

Le second groupe de raisons fait référence à l'exemple d'Œdipe lui-même. Fodor dit que Sophocle a eu besoin de plus de cinq cents lignes pour rendre vraisemblable l'histoire d'un fils qui épouse sa mère en ignorant totalement leur lien de parenté.

L'exemple d'Œdipe ne peut qu'être une exception à la généralisation commune à la plupart des cultures selon laquelle les gens n'épousent pas leur mère. En plus cette généralisation est acceptée comme vraie.

> Hard cases make bad laws; only a philosopher would consider taking Œdipus as a model for a normal, unproblematic relation between an action and the maxim of the act. Keep your eye firmly fixed on this: *most people do* not *marry their mothers*; that, surely, should define the norm. [Fodor, 1994, page 45, les italiques appartiennent au text original.]

Fodor ne veut pas conclure, comme il l'avait fait auparavant que les contenus sont étroits. Cependant, il lui faut montrer que les deux attitudes propositionnelles possibles d'Œdipe vis-à-vis de Jocaste sont différentes. Il propose de considérer les attitudes propositionnelles comme des relations comportant trois éléments: le sujet, les propositions et les modes de présentation. Pour que deux attitudes propositionnelles soient identiques, il est nécessaire que les trois éléments coïncident.

Le mode de présentation joue le même rôle que le concept de véhicule dans la conception des attitudes propositionnelles de 1989 que j'ai déjà exposée. Les modes de présentation ne sont que des phrases dans le langage de la pensée, différenciées non seulement en fonction de leurs contenus mais aussi en fonction de leurs syntaxes.

Supposons que l'on accepte l'individualisation des attitudes telle que Fodor la propose; encore faut-il expliquer comment les sujets arrivent à représenter les modes de présentation. Cette condition est, à première vue incompatible avec l'acceptation des contenus larges dans l'explication en psychologie.

> It seems that it's the syntactical properties of modes of presentation that are doing the work, and the attachment to an *intentional*, as opposed to computational, level of psychological explanation is merely sentimental. I don't, however, think that this sort of objection works. I think it's possible to imagine a state of affairs in which psychological laws that apply to a mental state *just in virtue of its broad content* might play an essential role in behavioral explanation, even if its assumed that the *implementation* of psychological laws is sensitive solely to modes of presentation. [Fodor, 1994, page 50, les italiques sont dans le texte original.]

En gros, Fodor veut dire que si ce n'était pas le cas, les choses ne marcheraient pas si bien. La manière selon laquelle les individus se représentent un même objet sur des modes de présentation différents ressort d'un savant dosage entre d'un côté, les similarités imposées par l'information issue de l'environnement et le fait que nous soyons des créatures d'un même type et d'autre part le caractère hétérogène de nos expériences personnelles.

Comme Fodor le dit:

> One can, I think, imagine a world where everything is delicately balanced in the following way: Content is broad, the metaphysics of content is externalist (e.g., causal/informational), and modes of presentation with similar causal histories (or nomic affilations; anyhow with similar broad contents) overlap enough in their syntax to sustain robust psychological generalisations. But not enough to make the minds that these generalizations subsume homogeneous under *syntactic* description. [Fodor, 1994, page 53]

Dans la prochaine section je ferai état d'autres solutions à la théorie fodorienne et notamment celle proposée par Pierre Jacob qui montre que Fodor pourrait bien retenir le dualisme des contenus à condition de renoncer à l'atomicité des attitudes propositionnelles pour adopter un bolisme modéré ou anatomiste.

Mais auparavant, j'aimerais émettre deux critiques aux justifications que Fodor apporte pour arguer du caractère accidentel des cas de Frege. La première concerne la rareté des situations telles que celle d'Œdipe.

D'abord il me semble qu'il existe une immense variété d'ouvrages, que ce soient des romans ou des pièces de théâtre (sérieux ou vaudevillesques) aussi bien que des films et des téléfilms où des situations semblables sont ficelées avec plus ou moins de succès; telle est la série mexicaine dont l'héroïne fut l'actrice argentine Libertad Lamarque et qui a marqué mon enfance. Dans ce

feuilleton une femme est la gouvernante d'un enfant dans une riche famille tout en recherchant son propre fils qu'on lui a ravi à sa naissance. Vous avez déjà deviné que l'enfant que l'on croyait le fils des riches se trouve être l'enfant tant recherché. Évidemment, la pauvre Libertad Lamarque a besoin d'une quarantaine d'épisodes pour s'apercevoir de ce que les spectateurs avertis avaient compris depuis le début.

On peut ignorer les cas produits par la fiction et trouver des exemples dans la chronique mondaine si l'on en croit les médias. Dans le *Paris Match* du 26 Octobre de 1995 est exposé le cas de Claudia Cardinale dont le fils a cru être le frère jusqu'à l'age de 7 ans. Dans le ton que *Paris Match* aime donner à ce type de tragédie, la situation est exposée comme suit :

> «Mon fils me fait souffrir: il s'agit presque d'une douleur physique, coincée dans ma poitrine. Je vis en permanence avec cette douleur. Et je me culpabilise; je sais que je suis grandement responsable de son état.» Si calme et si sereine, Claudia Cardinale révèle sa fêlure ... il aura fallu trente-six ans, l'âge de Patrick, son fils issu d'un viol, pour qu'elle se décide à la formuler aussi clairement. ... A l'âge de 7 ans, il [Patrick] apprit que sa sœur, de vingt ans son aînée était sa maman. [*Paris Match* 26.10.95, page 88]

Tout cela pour dire que je ne pense pas que ces cas soient purement accidentels et pour justifier ma méfiance à l'égard du *PEI*, produit d'une pétition de principe dont je ne crois pas devoir convenir. La deuxième critique vise toute la démarche de Fodor que je considère suspecte de circularité. En effet, son objectif est de naturaliser la psychologie ordinaire, ce qui revient à dire que les généralisations ont un pouvoir prédictif, voire nomologique. Aussi, naturaliser l'intentionnalité implique l'adhésion au physicalisme. Fodor postule que le niveau intentionnel se multiréalise à un niveau computationnel mais que les contenus sont larges. Pour arriver au bout de la preuve il nous demande d'admettre que les modes de présentation sont reflétés par les propriétés syntaxiques et que cela est justifié du fait que s'il en était autrement, alors:

> [...] the laws that a computational psychology implements might be *intractably and ineliminably* intentional [Fodor, 1994, page 53]

Mais hélas, *quod erat demonstrandum*.

7.5.4 Le holisme modéré des contenus comme solution acceptable

Rappelons-nous que le dualisme de propriétés permet de concilier les théories de contenu de type informationnel avec la contrainte du rôle causal des propriétés sémantiques.

Fodor abandonne le contenu étroit parce qu'il finit par le considérer superflu. Aussi abandonne-t-il aussi le dualisme des contenus. J'ai déjà montré les difficultés que présentent les individualisations des contenus étroits, qui selon Fodor se font de façon non-relationnelle car on ne tient compte que des autres états mentaux de l'individu et non de la situation de l'environnement.

Néanmoins, le résultat de l'adoption de la sémantique informationnelle n'est pas tout-à-fait satisfaisant car Fodor doit recourir à des arguments empiriques pour établir la corrélation entre le niveau computationnel et le niveau intentionnel indispensable pour soutenir la pertinence causale des propriétés intentionnelles.

En outre, Fodor propose le *PEI* qui est une pétition de principe, un espèce d'oracle auquel on doit se plier afin de donner une chance à la corrélation entre les propriétés intentionnelles et les propriétés computationnelles. Les *PEI*, à mon avis n'est qu'un pont, qui permet à Fodor d'affirmer que la cohérence entre les trois postulats de sa théorie ne dépend pas seulement des états mentaux et ne se vérifie pas de façon individuelle ou atomique mais qu'elle a besoin de s'ancrer dans l'ensemble des relations épistémiques.

Une théorie qui élève même un petit peu le caractère arbitraire du *PEI* est celle qui permet de motiver ces faits d'une autre façon. Par exemple, Fodor aurait pu considérer que l'individualisation des contenus se fait de manière holistique au lieu d'être atomique. Dans [Jacob, 1995] l'auteur montre qu'une position holistique modérée est compatible avec les lois intentionnelles en psychologie, la sémantique informationnelle et l'hypothèse computationnelle de la pensée.

Selon l'atomisme sémantique les propriétés sémantiques ou le contenu (par exemple d'une croyance) d'un individu dérivent de *dépendances nomiques* entre les états du système (par exemple

états du cerveau) et l'instantiation des propriétés de son environnement véhiculées par les systèmes perceptifs.

Le holisme sémantique, en revanche, soutient que le contenu d'une croyance dépend de ses relations *causales et/ou inférentielles* aux autres attitudes propositionnelles de l'individu. Le contenu est déterminé par les relations causales et inférentielles dans le système épistémique.

Au premier abord, le holisme sémantique permettrait de résoudre le problème des terres jumelles. Nous avons vu dans ce cas que la survenance mentale/physique ne tenait pas puisque si l'on détermine les contenus étroits de façon atomique (en fonction des leurs rôles computationnels[47]) alors ma jumelle et moi avons les mêmes états physiques internes malgré le fait que pour moi l'eau soit H_2O et XYZ pour elle. En revanche, si l'on détermine les contenus étroits de façon holistique, les deux contenus étroits seront différents comme le sont aussi les contenus larges en vertu du fait que j'entretiens la croyance que l'eau est H_2O alors que ma jumelle la conçoit comme XYZ. On peut donc démontrer que nos réseaux épistémiques respectifs ne sont pas identiques.

On voit que le holisme sémantique est une conséquence immédiate de la sémantique des rôles fonctionnels. La question à se poser est : Pourquoi Fodor réfute-t-il le holisme sémantique? Le problème est que le holisme menace la psychologie scientifique qui croit les lois intentionnelles.

Si l'on accepte le holisme sémantique, cela implique que si le contenu d'une croyance dépend des relations causales ou inférentielles de l'ensemble ou d'une partie des autres attitudes propositionnelles de l'individu, alors il est possible qu'aucune loi psychologique intentionnelle ne soit instantiée par deux individus. En d'autres termes, il devient très peu probable qu'une même croyance soit instantiée par différentes personnes, ce qui empêchera dès lors de soumettre différents individus aux mêmes lois psychologiques intentionnelles. Ces dernières ne s'appliqueront guère non plus à un même individu dans des tranches différentes de temps.

Cependant le concept de holisme est multiple. Fodor a écrit avec Ernest Lepore un ouvrage qui analyse toutes les formes de holisme dans la philosophie de l'esprit et des sciences dans le but de donner une série de raisons pour lesquelles le holisme pourrait être réfuté. [48]

La stratégie consiste à démontrer qu'il n'existe que deux positions possibles pour l'individualisation des contenus, le holistique et l'atomique, mais que seule cette dernière est compatible avec la psychologie scientifique. Pierre Jacob montre qu'il existe une autre possibilité, soit une position modérément holistique où l'individualisation des contenus se fait de façon anatomiste ou molécullariste sémantique.

Je vais donner différentes définitions du holisme sémantique.

DEFINITION 2 (HOLISME SÉMANTIQUE, VERSION UNIVERSELLE OU STRICTE)
Le holisme sémantique est la thèse que le contenu (ou propriété sémantique) d'une croyance C d'un individu est déterminé par les relations entre C et toutes *les autres croyances de l'individu (par la totalité de ce que Fodor nomme des* liaisons épistémiques*)* [49] *de C.* [50]

DEFINITION 3 (PROPRIÉTÉ ANATOMIQUE)
Une propriété P est anatomique *dans le cas où si une entité possède P, alors au moins une autre entité la possède aussi. (Par exemple, "être marié")*[51]

DEFINITION 4 (PROPRIÉTÉ ATOMIQUE)
Une propriété se dit atomique *si elle n'est pas* anatomique.

DEFINITION 5 (PROPRIÉTÉ HOLISTIQUE)
Une propriété P est dite holistique *au cas où si une entité possède P, alors beaucoup d'entités possèdent P.*

[47] Le rôle computationnel correspond au niveau instantié ou computationnel (le niveau inférieur au niveau intentionnel) tandis que le rôle inférentiel ou conceptuel ou fonctionnel appartient au niveau supérieur, c'est-à-dire intentionnel.

[48] [Fodor and LePore, 1992]

[49] Les relations épistémiques sont les relations inférentielles et/ou causales avec les autres croyances.

[50] cfr. [Fodor, 1987, page 56–57] et [Jacob, 1995, page 12]

[51] [Fodor and LePore, 1992, cf. Glossary]

Autrement dit, une propriété est holistique si elle est très anatomique.

DEFINITION 6 (HOLISME SÉMANTIQUE VERSION NON-UNIVERSELLE OU NON-STRICTE)
La version non-universelle du holisme sémantique est la thèse métaphysique qu'une propriété sémantique comme avoir un contenu est holistique en ce sens qu'une expression appartenant à un langage ne peut posséder cette propriété que si beaucoup d'autres expressions (non-synonymes) appartenant à ce langage la possèdent aussi.[52]

La position que Jacob revendique comme une alternative possible entre l'atomisme et le holisme sémantique en la qualifiant d'anatomisme sémantique affirme qu'il est possible de prendre une partie des éléments du réseau épistémique de deux individus qui partagent une même croyance pour déterminer le contenu sémantique de cette croyance sans réfuter ainsi l'existence de lois psychologiques intentionnelles.

> On ne pourra en conclure à l'impossibilité de lois psychologiques intentionnelles subsumant différents individus qu'à condition d'adopter l'interprétation stricte de la notion de propriété holistique. On ne conclura que deux individus ne peuvent partager une seule croyance et être subsumés sous des lois psychologiques intentionnelles que si on suppose que deux individus ne peuvent partager une seule croyance si ils ne partagent pas *toutes* leurs croyances. Si on suppose que deux individus ne peuvent partager une seule croyance s'ils ne partagent pas *beaucoup* des croyances, il est parfaitement possible que deux individus partagent des croyances et soient subsumés par des lois psychologiques intentionnelles. Il n'est pas absurde de penser que des individus distincts partagent beaucoup de croyances: comme les croyances que l'eau à une température supérieure à 0^0 C est un liquide, qu'elle bout à 100^0 C, que l'or est un métal, que deux est un nombre pair, que lundi précède mardi, que l'herbe est verte, que la neige est blanche, que les roses sont des fleurs, que la plupart des oiseaux sont couverts de plumes, et ainsi de suite. [Jacob, 1995, page 16]

Cependant, Jacob signale aussi qu'il est possible de soutenir même le holisme sémantique dans son interprétation universelle à condition d'ajouter l'hypothèse supplémentaire que les contenus à individualiser dépendent des croyances *effectives*. Cependant, les croyances effectives d'un individu dépendront de sa biographie, et ceci pose un problème, qui n'est qu'apparent [53], des croyances qui seront *a priori* en relation avec ce contenu.

Jacob affirme:

> Mais on pourrait faire valoir que les croyances d'individu qui sont pertinentes pour la formulation de lois psychologiques intentionnelles propres à l'espèce humaine ne sont pas toutes les croyances effectives d'un individu mais les croyances que l'individu formerait contrefactuellement s'il était placé dans telles ou telles circonstances. Cet amendement devrait convenir à un partisan de l'innéisme comme Fodor qui suppose que beaucoup de concepts humains (sinon tous) sont innés. [Jacob, 1995, page 17]

Par la suite, Jacob signale qu'une position anatomiste n'empêche pas de préconiser une sémantique informationnelle. Bien au contraire, une sémantique informationnelle n'est pas du tout contrainte à admettre l'atomisme sémantique des pensées primitives dépourvues de structure syntaxique. Les arguments de Jacob sont de deux types. Ceux du premier type se basent sur l'architecture cognitive d'une créature dotée de propriétés sémantiques informationnelles plus propres à favoriser l'anatomisme sémantique et à invalider l'atomisme sémantique que le contraire. Sa seconde ligne d'argumentation part d'un point de vue évolutionniste selon quoi, également en vertu de contraintes architecturales, il se révèle plus facile selon un processus évolutif gouverné par la sélection naturelle, d'engendrer une créature douée d'un système visuel capable de détecter plusieurs couleurs qu'une ne pouvant détecter qu'une seule couleur. Il en est de même pour l'architecture cognitive et les propriétés sémantiques. [Jacob, 1995, page 30]

La solution de Jacob me semble plus satisfaisante que celle de Fodor car elle intègre mieux les termes de la trilogie composant le credo fodorien de l'esprit : les lois psychologique intentionnelles, la sémantique informationnelle et la pensée computationnelle.

[52] cf. [Fodor and LePore, 1992, page 258], [Jacob, 1995, page 12]

[53] Je ne vais pas exposer les raisons citées par Fodor et Lepore pour prouver que la seule alternative possible pour la détermination des propriétés sémantiques se situe entre le holisme strict et l'atomisme, ce dernier étant seul compatible avec les lois intentionnelles en psychologie. A la base de la discussion on découvre que le rejet de la distinction analytique/synthétique ne permet pas l'individualisation du concept d' *effectivité*.

En particulier l'invocation de l'anatomisme sémantique rend plus plausible la corrélation entre les propriétés computationnelles de niveau inférieur et les propriétés intentionnelles de niveau supérieur puisque les croyances pertinentes n'ont pas un rôle explicite et font partie de la raison epistémique de l'individu. Toutefois, le holisme modéré prôné par Jacob limite à mon avis les possibilités de la multiréalisation. En effet, la définition des croyances *effectives* qui sont les composantes du noyau *kernel* d'une croyance se montre très peu propre à une comparaison entre espèces ou avec des instruments. Ceci toutefois n'est pas une limitation de la solution de Jacob mais une limite inhérente à la théorie fonctionnaliste computationnelle de la pensée.

7.6 Les critiques au concept de multiréalisation

Dans cette section j'émets deux critiques différentes à l'égard concept de multiréalisation. La première est d'ordre formel et concerne la possibilité de considérer que les énoncés qui ont comme antécédents des disjonctions sont nomologiques. La seconde met en doute non la possibilité conceptuelle de l'hypothèse de la multiréalisation de la cognition mais les prémisses théoriques qui la rendent empiriquement possible.

7.6.1 Le problème des antécédentes disjonctives

Rappelons que la multiréalisation des propriétés et des lois est une alternative à la réduction, une alternative physicaliste non-réductionniste. La définition de multiréalisation d'une propriété P_L du niveau L au niveau $L-1$ est obtenue lorsqu'il existe une disjonction des propriétés du niveau $L-1$ dont l'instantiation d'un des termes est suffisante à l'instantiation de P_L et lorsque l'instantiation de P_L est suffisante à l'instantiation de la disjonction mais non pas à l'instantiation d'un de ces termes.

J'ai aussi montré que dans le cadre fodorien la multiréalisation des lois sert également à expliquer les exceptions dans les sciences spéciales. Le schema 7.11 suivant montre la multiréalisation d'une loi d'une science spéciale, p. ex. la psychologie.

$$\begin{array}{ccc} M_F & \longrightarrow & M_G \\ \uparrow & & \uparrow \\ NF_1 \vee NF_2 & \longrightarrow & NG_1 \vee NG_2 \end{array} \tag{7.11}$$

Néanmoins, pour que réalisation il y ait, il est nécessaire que la relation du niveau réalisateur $NF_1 \vee NF_2 \longrightarrow NG_1 \vee NG_2$ puisse faire office de loi. C'est justement cela qui pose des problèmes.

Traditionnellement, les prédicats qui sont des antécédents des lois doivent se vérifier comme étant des espèces scientifiques (*kinds*); autrement dit, les antécédents doivent représenter ou être caractérisés par une seule propriété. L'identité extensionnelle entre M_F et le prédicat disjonctif suffit à établir la multiréalisation, à la différence de la réduction nagelienne qui exige de plus l'identité des propriétés entre les deux entités en corrélation. Néanmoins, dans la multiréalisation le problème se présente lorsque l'on s'interroge sur le caractère nomologique de l'énoncé réalisateur.

En effet, il s'agit de voir si la disjonction des espèces scientifiques différentes ou hétérogènes est acceptable comme antécédente d'une loi stricte ou non.

Dans [Kim, 1992b] l'auteur rejette la possibilité de considérer comme nomologiques des énoncés qui ont des disjonctions comme antécédents. Il prend l'exemple du minéral *jade*. On supposait que le jade était un minéral mais on s'est aperçu plus tard qu'il est formé de deux structures moléculaire différentes: la *jadéite* et la *néphrite*.

Alors, on déduit la loi suivante:

(L) Le jade est vert.

n'est qu'une conjonction des lois qui suivent:

(L_1) La jadéite est verte.

(L_2) La néphrite est verte.

Maintenant supposons que nous nous demandions si l'énoncé L continue à être un énoncé nomologique. Une des caractéristiques des lois est qu'elles admettent des contrefactuelles mais de ce point de vue l'énoncé L semble robuste.

L'autre caractéristique est qu'une loi doit être confirmée par l'observation d'instances positives. Les généralisations que l'on soupçonne nomologiques sont rendues fiables par l'observation d'instances positives. Kim se demande si l'énoncé L passera ce test avec succès. Il nous propose de faire la supposition suivante : imaginez–vous que tous les échantillons de jade que l'on a examinés jusqu'à maintenant et qui sont des instances positives de L se trouvent être tous des exemples de jadéite et aucun de néphrite.

Peut–on continuer à tenir L pour confirmé? D'un point de vue intuitif la réponse serait négative car on a une très forte présomption de la validité de L_1 mais aucune pour L_2. Néanmoins toutes les instances de jadéite sont aussi des instances positives pour L parce qu'elles confirment l'antécédente et la conséquente. Ce qui cause notre résistance à accepter la confirmation de L est qu'elle n'est pas donnée sous la forme que nous attendions. Kim attribue cette situation au fait que le *jade* n'est pas une espèce nomique mais plutôt la disjonction de deux espèces nomiques différentes : la *jadéite* et la *néphrite*.

Kim en conclut que les antécédentes disjonctives risquent de rendre fausse ou peu robuste la confirmation par l'observation d'instances positives.

> [...] disjunctive antecedants would sanction a cheap, and illegitimate, confirmation procedure. [Kim, 1992b, page 12]

Si l'on se place dans une perspective logique, il est évident qu'un énoncé qui affirme que "Tous les Fs ou Hs sont des Gs" implique logiquement "Tous les Hs sont Gs". Or, tout énoncé impliqué par un autre bien confirmé doit aussi être confirmé. On pourrait arriver à la fâcheuse conclusion que le constat "Tous les Hs sont Gs" est confirmé par le fait qu' on a confirmé par l'observation que "Tous les Fs sont Gs", ce qui o'est pas une situation très soubaitable.

Le problème se présente donc, lorsque l'on a des énoncés nomologiques où les antécédentes sont des disjonctions, des prédicats hétérogènes ou largement disjonctifs. Au centre du problème se trouve le fait qu'une disjonction des propriétés n'est pas forcément une propriété. Kim le dit de la manière suivante :

> More generally, the phenomenon is related to the point often made about disjunctive properties: disjunctives properties, unlike conjunctive properties, do not guaranter similarity for instance falling under them. And similarity, it is said, is the core idea of a property.... Properties of course can be conjunctions, or disjunctions, *of* other properties. The point about disjunctive properties is best put as a closure condition on properties: class of properties is not closed under disjonction (presumably, nor under negation). Thus, there may well be properties P and Q such that P *or* Q is also a property, but its being so doesn't follow from the mere fact that P and Q are properties. [Kim, 1992b, page 13]

Si maintenant on revient au problème corps–esprit, je pense que l'on peut convenir que la situation est semblable à celle du *jade*. De même que le jade est une espèce monique parce qu'il est disjonctif étant donnée la structure moléculaire différente de la jadéite et la néphrite, l'état physique N sera la disjonction de deux espèces scientifiques différentes (disons comme dans le schema 7.11 NF_1 et NF_2). En effet, supposer qu'il s'agit d'espèces scientifiques différentes du point de vue de la microstructure correspond à la *raison d'être* de la multiréalisation. Rappelons que selon l'hypothèse de base la multiréalisation peut s'avérer réalisée sur différents substrats. Cependant, si l'on tient compte des critiques de Kim, les énoncés physiques, voire neurologiques pour lesquelles ces disjonctions jouent le rôle d' antécédentes méritent difficilement la qualification de nomologique, ce qui met en échec la prétention des multiréalisations de lois.

J'aimerais rendre clair le tableau; le concept de réalisation est différent de celui de réduction, le premier étant explicatif tandis que le second s'avère être une corrélation extensionnelle (bien que celui là n'écarte pas explicitement le deuxième). Les propriétés mentales humaines sont réalisées sur

le cerveau humain en vertu des propriétés de leur microstructure. Le caractère non-réductionniste de la démarche est assuré par l'hypothèse de la multiréalisation de la cognition. Néanmoins, ce caractère de multiréalisé ne permet pas d'affirmer que les relations obtenues au niveau neurologique telles que la disjonction des différents types de réalisation méritent la qualification de lois.

Les fonctionnalistes computationnels peuvent nous répliquer que la disjonction a, pour ainsi dire, un dénominateur commun qui transformerait la disjonction en une espèce scientifique. Ce dénominateur commun serait l'appartenance à une même espèce scientifique caractérisée par son rôle causal et ceci est l'unique chose qui compte dans l'individualisation des objets pertinents à la théorie.[54]

Cependant, on doit à mon avis repouser cette tentation qui n'offre qu'une fuite en avant. Si l'on postule comme trait commun à toutes les réalisations d'un état mental qu'elles ont le même pouvoir causal alors ceci entraîne forcément l'identification de ce pouvoir causal des réalisations avec le pouvoir de l'état mental qu'elles réalisent ou instantient. En effet, c'est lui le dénominateur commun. Le concept de pouvoir causal doit être compris dans le sens strict du terme et non par exemple dans le sens de responsabilité causale. [55] Ceci accordé il est évident que les pouvoirs causaux de chaque réalisation reposent sur des propriétés projetées des lois intentionnelles mais que ces propriétés projetées[56] sont physiques, particulières à chaque réalisation et probablement très dépendantes du substrat en question. Par conséquent, cette manière de postuler une espèce en la caractérisant par son pouvoir causal doit être récusée.

Quelles sont les conséquences d'un tel constat? Il faudrait d'abord renoncer à la prétention de bâtir une psychologie unifiée; il y aurait une psychologie pour les humains, une psychologie pour les martiens, une psychologie des mollusques, une psychologie des reptiles, une psychologie pour les cerveaux simulés par le peuple chinois où chaque chinois représente un neurone relié au commandement supérieur qui lui indique les actions à suivre par radio satellite[57], une autre psychologie pour robots réels, une autre psychologie pour robots virtuels. Aussi une psychologie pour cerveaux simulés par des souris, des pigeons et des fromages suisses. Il est même possible que Brentano ait vu juste en proposant l'existence d'un psychologie pour les hommes et d'une autre, différente pour les femmes.

Cette liste qui énumère des exemples pris au sérieux dans les textes de la philosophie de l'esprit, a aussi pour but de montrer comment le concept de multiréalisation a été pris à la légère. En lisant toute la bibliographie de cette époque (les années soixante-dix et quatre vingt) j'ai l'impression que l'on a souvent discuté de la multiréalisation comme d'un fait accompli et non comme d'une possibilité logique dans un monde possible dont d'ailleurs personne n'a jamais pu prouver qu'il est le monde réel.

Si on a pris ces exemples comme valables, ce qui en philosophie veut dire logiquement possibles, est-ce à cause du caractère très libéral de la définition de rôle fonctionnel dans le fonctionnalisme computationel?

Néanmoins il faudrait se poser la question des évidences dans le champ des neurosciences qui vont dans le sens de la multiréalisation et j'aborderai ce point dans la section suivante.

---

[54] J'ai déjà expliqué cette position de Fodor dans le chapitre 3 §3.4.1 au paragraphe intitulé: **Le dualisme de contenus selon Fodor en 1987** en référence à l'exemple de la pièce magique qui change les molécules. Mais pour ajouter encore un éclaircissement supplémentaire à cet exemple, il y a un principe largement accepté, que Kim appelle le Principe d'individualisation causale de l'espèce et que je citerai ici:

> [Principle of Causal Individualisation of Kinds] Kinds in science are individuated in the basis of causal powers; that is, objects and events fall under a kind, or share in a property, insofar as they have similar causal powers. [Kim, 1992b, page 17]

[55] c.f. §7.5.1. paragraphe **Les généralisations en psychologie.**

[56] J'ai défini le concept de propriété projetée dans la note de pie de page §29 du chapitre 7 §. 7.5.1.

[57] cf. [Block, 1978]

7.6.2 Les limites de la non-pertinence physique de l'implantation fonctionnelle

Joëlle Proust étudie ces limites dans un intéressant article. Dans [Proust, 1993] elle reprend les arguments empiriques que Fodor et Block évoquent dans [Block, 1980b].

Le premier des éléments que ces auteurs mettent à contribution est la doctrine de l'équipotentialité neurologique développée par Lashley. Dans son ouvrage *Brain Mechanisms and Intelligence* paru en 1929, Lashley doute de l'importance de la localisation de l'activité neuronale et donne la primauté aux relations dynamiques entre les diverses parties du système nerveux. La caractérisation d'équipotentielle doit être comprise comme la capacité fonctionnelle de contrôler un comportement qui peut acquérir une zone voire une aire du tissu cérébral. La capacité d'équipotentialité est tributaire d'une autre : la plasticité. La plasticité d'une zone est la capacité qu'elle a de prendre en charge des fonctions qui appartenaient à un autre groupe de neurones après que ces derniers eussent été endommagés par exemple.

Néanmoins, les développements postérieurs en neurosciences, par exemple les travaux de D. Hubel et T. Wiesel[58] montrent qu'il existe bien des cellules spécifiques du cortex visuel qui répondent à des informations spécifiques de couleur et d'orientation.

Ces travaux montrent que l'on ne peut pas écarter complètement la possibilité d'une pertinence de la localisation des neurones du moins pour certaines fonctions. On ne saurait donc affirmer que les interactions neuronales se font selon des critères purement fonctionnels. Joëlle Proust affirme que le clivage *structure* et *fonction* a été dépassé; il serait plus judicieux de prendre en compte l'existence des *niveaux d'organisation*. Bien que la plasticité et l'équipotentialité neurologique soit l'un des arguments de Fodor et Block, ils semblent maintenir l'ancien clivage. Il persistent à considérer que la détermination des états physiques ne passe que par deux alternatives possibles : structure ou fonction.

Peut-être la caractérisation structurelle qu'ils critiquent est-elle dans l'esprit des fonctionnalistes de types mais le clivage entre structure et fonction n'est pes l'unique caractérisation possible des types physiques.

La recherche de régularités neurophysiologiques de base pour les états mentaux peut suivre une voie intermédiaire entre ces deux approches. Il existe encore la possibilité de prendre en considération le niveau d'organisation. Un des éléments de la thèse que je défends est la proposition d'une solution physicaliste non-réductionniste qui permette de tenir compte de la dynamique du substrat.

L'autre argument de Fodor et de Block en faveur de la multiréalisation consiste à préconiser l'existence d'une possible convergence entre l'évolution phylogénétique de la morphologie et le comportement. J'ai déjà adressé mes critiques à des arguments qui se basent sur des théories de l'évolution. En effet, il est difficile de voir par quel mystère la sélection naturelle ou d'autres facteurs évolutifs ont pris une influence quelconque sur les machines.

La troisième critique touche l'argument selon lequel il y aurait des ressemblances psychologiques entre espèces sans qu'il existe nécessairement de ressemblances neurologiques. Ainsi, on trouve pertinent de parler de la douleur que peuvent subir un mollusque, un robot, un humain et un martien. Proust se demande comment comprendre ce que veut dire "l'organisme éprouve de la douleur" lorsque l'on abandonne tout référence neurophysiologique. Dans quelles conditions pourrons nous dire qu'un organisme phylogénétiquement éloigné de l'espèce humaine éprouve de la douleur? La réponse de Proust distingue deux cas; elle appelle le premier le cas *homologique*, c'est-à-dire quand les espèces à comparer ont des structures et des fonctions très semblables (par exemple des groupes d'individus dans la classe des mammifères). Le second est le cas *analogique* où les structures et les fonctions ont des ressemblances bien plus éloignées que dans le cas précédent et où l'analogie se base sur l'existence d'états fonctionnels semblables dans les deux espèces à comparer.

> L'axe de l'homologie pourra suggérer qu'à systèmes nerveux voisins, expériences voisines, et même articulation entre croyances, apprentissages et planification des actions. Mais sur l'axe de l'analogie,

[58] Voir [Hubel and Wiesel, 1990].

> les choses sont loin d'être claires. Si l'on suppose que les comportements décrits fonctionnellement permettent de mettre en correspondance l'état hypothétique de douleur dans l'organisme humain et dans un type de sujet n'appartenant pas au même phylum, comme par exemple un gastéropode ou un insecte, on perd progressivement tout repère au fur et à mesure que divergent l'une de l'autre davantage non seulement l'implémentation, mais la caractérisation fonctionnelle de l'état considéré. [Proust, 1993]

Proust développe subséquemment des arguments tendant à démontrer l'impossibilité de soutenir la pertinence de l'attribution des états qualitatifs ou qualia, comme celui de la douleur notamment, dans l'axe analogique. Ces arguments constituent une critique à l'égard de l'amalgame que l'on fait entre la fonction mathématique ou computationelle et la fonction biologique. Je ne développe ces critiques ici parce que dans la même ligne j'exprimerai mes réserves sur la pertinence de cet amalgame dans le chapitre suivant lorsque je débattrai des limites de la vie artificielle.

Finalement Proust fixe un cadre philosophico–méthodologique de la multiréalisation en concluant :

> Il me semble justifié de concevoir les systèmes psychofonctionnels comme strictement relatifs à une espèce, en ne conservant la multiréalisabilité que pour les variations interindividuelles de réalisation à certains besoins théoriques, à un niveau fonctionnel "général" ; mais il n'est plus nécessaire d'attribuer à ce niveau autre chose qu'un usage heuristique. L'avantage de ce retour à un fonctionnalisme physicaliste est de restituer sa place dans la recherche interdisciplinaire sur la structure du mental, et d'éclairer la question des rapports entre le corps et l'esprit. [Proust, 1993]

Je suis d'accord avec les critiques que Proust émet à l'encontre de la multiréalisation et je suis d'accord avec sa postulation méthodologique qu'il faut rendre une place privilégiée aux neurosciences dans la recherche en sciences cognitives. On s'est trompé en les reléguant au deuxième plan au bénéfice des modèles artificiels.

Cependant, si la multiréalisation est mise à mal par toutes ces critiques c'est à cause du caractère ambigu du concept d'une part et de l'aboutissement à des lois de type disjonctif de l'autre.

La dernière des difficultés que je viens d'évoquer serait insurmontable parce qu'il semble acquis qu'il existe de très fortes présomptions contre l'impossibilité de trouver des propriétés des microstructures réalisatrices capables de caractériser les différents substrats remplissant le même rôle fonctionnel. En définitive, il semblerait qu'il n'existe pas de caractérisation de ces états physiques qui soit à la fois neutre et non ambiguë; dans la conclusion de ce travail je plaide pour une possible solution. Ainsi, la multiréalisation trouvera un cadre plus strict pour sa définition et on disposera en principe d'un outil pour sa validation.

En outre, une des raisons qui plaide contre une telle possibilité est la forte présomption que les microstructures sont conçues comme des entités statiques, c'est–à–dire comme des entités caractérisées par leurs relations spatio–structurelles. Dans la conclusion de ce travail je vais soutenir que si l'on arrive à se détacher de ce présupposé tout en tenant compte de la dynamique du substrat nerveux, une classification des états physiques sera possible suite à l'application des modèles de morphodynamique de Thom.

La tâche majeur des neurosciences est de découvrir cette dynamique.

7.7 Conclusion

J'ai présenté dans ce chapitre les positions fonctionnalistes computationnelles qui réfutent les théories de l'identité des types. D'abord j'ai présenté le fonctionnalisme computationnel turingien qui prend comme moyen d'individualisation des rôles computationnels la formalisation de la Machine de Turing. Nous avons vu que Putnam a opté pour cet amalgame entre le concept de fonction mathématique et le concept de fonction téléologique car il était conscient de la difficulté que comporte la définition d'un isomorphisme fonctionnel entre systèmes: c'est–à–dire la possibilité d'établir une correspondance entre deux systèmes qui sont dans le même état mental sans faire aucune allusion au domaine physique. Nous avons repéré quelques difficultés posées par cette démarche.

Ensuite, j'ai présenté la théorie fodorienne de l'esprit et l'hypothèse du langage de la pensée. Nous avons vu que l'hypothèse de la multiréalisation des propriétés intentionnelles est fondamentale, selon Fodor pour justifier l'existence des lois psychologiques intentionnelles. Néanmoins, nous avons remarqué que Fodor se heurte aussi à un problème semblable voire identique à celui de l'isomorphisme fonctionnel tel que Putnam le définit au début de ce chapitre. En effet, les arguments de Fodor pour justifier un corrélation entre propriétés computationnelles et propriétés intentionnelles sont des argumentations non pas philosophiques mais bien empiriques.

Dans la conclusion je vais souligner les limitations du concept de multiréalisation et montrer que, si ces limitations se vérifient, elles restreignent l'intelligence artificielle dans le cadre des sciences cognitives à la simple simulation. La possibilité de créer des instruments intentionnels dans le sens propre du mot s'avère improbable étant donné que les propriétés intentionnelles ne sont pas totalement indépendantes du substrat, contrairement à ce que les fonctionnalistes computationnels veulent croire. Les difficultés que Fodor a rencontrées tout au long de son programme ne font qu'illustrer cette improbabilité. Il me semble que le caractère non-réductible du mental ne peut pas être basé sur la définition des rôles fonctionnels et je pense que la tentative de contourner les propriétés du substrat, c'est-à-dire la réalisation au niveau neuronal, a échoué. Néanmoins, le physicalisme réducteur n'est pas l'unique alternative et le but de ma thèse est d'en proposer une autre.

Partie III

Le modèle émergentiste

Chapitre 8

Vous avez dit "émergence"?

The higher–quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that lower level, but constitutes its possessor a new order of existent with its special laws of behavior. The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. Its admits no explanation.
[Alexander, 1974, page 46]

8.1 Introduction

Le terme d'émergence est souvent utilisé dans les sciences cognitives pour évoquer différents types de caractéristiques. Dernièrement certaines branches de l'informatique comme la robotique ou la vie artificielle ont utilisé le terme émergence pour décrire un comportement en tant que synonyme de nouveau ou d'imprévisible mais en général sans préciser à quelles caractéristiques on se réfère.

Dans la philosophie des sciences de l'après–guerre, le concept d'émergence est donné comme étant l'opposé d'une réduction, ce qui implique que les entités ou propriétés émergentes seront des faits ou des phénomènes réfractaires à l'analyse scientifique constituant donc des entraves à l'hypothèse de l'unité des sciences.

L'usage abusif ou vague que l'on fait souvent du terme éveille des soupçons. On se demande en vertu de quoi une entité ou une propriété mériterait d'être considérée comme émergente. De plus on se demande s'il existe des entités, des événements, des propriétés, des lois ou des théories proprement émergentes. Si l'on accepte l'existence des propriétés émergentes, par exemple, quelle est la relation entre ces propriétés et les propriétés survenantes[1] ? Ces propriétés auront–elles un rôle causal ou s'agira–t–il de simples épiphénomènes?

En tout cas, les difficultés pour définir ce terme semblent considérables.

Selon Francisco Varela[2] le problème est que le concept ne repose pas sur une théorie unifiée. Il constitue une façon nouvelle de parler du phénomène jadis appelé en cybernétique *l'auto–*

[1] Pour une définition de survenance voir chapitre 2 §2.3.3
[2] [Varela et al., 1993, cf. page 88]

organisation. Il représente le passage de la cohérence locale à la cohérence globale d'un système. Le fait qu'il n'y ait pas de théorie unifiée ressort de ce que les phénomènes émergents sont orthogonaux à plusieurs domaines, par exemple les oscillations chimiques, les réseaux génétiques, les populations génétiques et les réseaux immunologiques entre autres. Mais Varela et al. ajoute :

> What all these diverse phenomena have in common is that in each case a network gives rise to new properties, which researchers try to understand in all generality. One of the most useful ways of capturing the emergent properties that these various systems have in common is through the notion of an *attractor* in dynamical system theory. [Varela et al., 1993, page 88–89]

A première vue, le concept d'émergence s'avère être une notion polyvalente et polymorphe puisqu'elle peut s'appliquer à une foule de phénomènes sans qu'il existe une définition précise des propriétés qu'on leur octroie. Le plan de ce chapitre est le suivant. D'abord je présente le concept d'émergence du courant britannique du début de siècle pour mettre en évidence que le caractère d'imprévisibilité ou de non-réductibilité ou d'inexplicabilité que l'on a prêté aux phénomènes dits émergents trouve son origine historique dans l'impossibilité pour l'émergentisme britannique de s'adapter aux nouvelles données émanant des découvertes scientifiques de son temps. Ensuite, je vais montrer que ces caractéristiques doivent être rejetées parce qu'elles privent de toute pertinence scientifique le concept d'émergence ainsi construit. Dans §8.4. je vais donner des arguments qui militent contre une possible conciliation de l'émergence avec l'irréductibilité. La troisième des stratégies que je présenterai vise à prouver que l'émergence est le contraire du concept de réalisation et je vais récuser cette argumentation.

Ensuite, je vais justifier (§8.5) la nécessité de trouver une caractérisation de l'émergence qui ne s'avère pas triviale. Je vais donc discuter le concept de propriété émergente en relation au concept de propriété relationnelle et par la même occasion je vais expliquer l'usage que l'on fait de ce terme dans le cadre de la vie artificielle.

Finalement, je présenterai la définition d'émergence de Mario Bunge (§8.6.) comme une possibilité de donner à ce terme la place qu'il mérite dans la caractérisation scientifique des phénomènes.

Le but de ce chapitre est d'appuyer la thèse que je défends selon laquelle une conception rationnelle de l'émergence est l'inverse du concept de réalisation et non son contraire.

Mais qui a dit "émergence" pour la première fois?

8.2 L'émergentisme britannique

L'émergentisme britannique est apparu comme une des positions contraires au *vitalisme substantiel* vers la moitié du XIX siècle.

Le *vitalisme substantiel* est la théorie de l'existence d'une entité particulière *entéléchie* en tant que facteur nécessaire à l'existence des comportements dans des corps vivants. Cette entité est absente des objets inorganiques ou des corps non-vivants.

L'émergentisme, les *théories mécanistes* et les *théories des composantes* [3], tous adversaires du vitalisme substantiel, partagent la conviction que les caractéristiques distinctives d'êtres vivants peuvent être expliquées en termes de leur structure, de leurs composantes ou des deux à la fois et non en postulant une entité particulière comme le veut le vitalisme.

L'émergentisme et le mécanicisme sont néanmoins d'accord sur les critiques que l'on peut émettre sur les théories des composantes.

En effet, ces critiques signalent que les théories des composantes peuvent être assimilées aux théories vitalistes substantielles au sens où elles prônent l'existence d'une composante particulière dans la structure des choses vivantes qui est responsable de la vie en elles. Les théories vitalistes et celles des composantes sont considérées non-scientifiques pour diverses raisons.[4] D'abord, on n'a jamais pu isoler ni des entéléchies ni des composantes jouant un rôle semblable. Secondo ces entités sont supposées avoir des caractéristiques *ad hoc* différentes des autres entités matérielles. Il n'est pas clair qu'elles aient une étendue dans l'espace et dans le cas des théories des composantes on ne sait pas si elles ont en partie la structure matérielle du corps vivant ou si l'on doit

[3] Ma traduction de *component theories*

[4] Pour la défense des positions vitalistes voir [Driesch, 1926]

comprendre au contraire que les corps vivants sont composés de structures matérielles et d'une entéléchie.[Beckermann et al., 1992, cf. page 101]

Les théories mécanistes soutiennent que le comportement d'un ensemble est totalement déterminé par la nature et par l'arrangement de ses composantes et aussi que l'on peut déduire le comportement de la totalité si l'on a une connaissance suffisante de la façon dont ces composantes se comportent isolément ou lorsqu'elles sont placées dans des ensembles plus simples.

Les théories émergentistes, en revanche, soutiennent que le comportement d'une totalité ne peut, même pas en théorie, être déduit de la connaissance aussi complète soit-elle, du comportement des composantes prises séparément ou placées dans des autres combinaisons, ni de leurs proportions, ni de leurs arrangements dans la totalité en question [Broad, 1925, cf. page 59].

8.2.1 La genèse du terme *émergence*

La tradition émergentiste trouve sa source avec John Stuart Mill dans son *System of logic* paru en 1843. Elle se retrouve chez d'autres auteurs dont Alexander Bain, George Henry Lewes, Samuel Alexander, Lloyd Morgan et C. D. Broad qui a publié en 1923 le dernier grand travail de cette première époque de l'émergentisme. Son livre *The Mind and Its Place in Nature* est un ouvrage de référence dans toutes les discussions sérieuses sur ce sujet. Sur l'origine du terme *émergence* on peut dire qu'il est dû à G. H. Lewes qui l'oppose au terme *résultante*. Ces deux termes se rapportent respectivement aux notions d'*hétéropathie* et d'*homopathie* proposées par John Stuart Mill.

Mill introduit ces notions dans son texte de 1843 [Mill, 1843] alors qu'il discute les effets de la composition des agents causaux conjoints. Selon Mill, il y a deux modes différents d'action des causes conjointes : le mode mécanique et le mode chimique.

Dans le mode mécanique, l'action de causes conjointes prend comme paradigme la composition des forces en physique qui peut être déduite de la sommation (par exemple en utilisant la loi du parallélogramme) des vecteurs de ces forces. Mill appelle l'effet produit dans le mode mécanique un effet *homopathique*. De même il appelle *lois homopathiques* celles qui expriment des relations causales entre des causes et des effets homopathiques.

Dans le mode chimique en revanche, l'action conjointe des causes ne cause pas un effet consistant en la sommation des effets de chaque type de causes prise séparément. L'un des exemples typiques du mode chimique est l'émergence des propriétés de transparence et de liquidité de l'eau; ces propriétés n'étant pas des propriétés des composantes de l'eau : l'hydrogène et l'oxygène. Le mode chimique est tout simplement l'absence de mode mécanique. L'effet produit par l'action conjointe des causes dans le mode chimique est appelée *hétéropathique*. De même Mill appelle les lois qui expriment des relations causales entre les causes et les effets hétéropathiques, *lois hétéropathiques*.

Les lois hétéropathiques sont, selon Mill des failles ou des ruptures du principe de composition des causes. L'existence de ces lois justifie l'existence des sciences spéciales puisque leurs effets produisent des changements dans les propriétés de l'objet qui s'avèrent différentes de celles de l'état précédent.

Mill pense que les agents causaux dont les sciences spéciales s'occupent ont la possibilité d'*agir sur un certain nombre de choses*.

Les agents causaux hétéropathiques agissent comme des forces qui changent l'accélération et par conséquent le mouvement des éléments appartenant à un niveau donné. Les processus chimiques, par exemple entraînent la réorganisation des éléments, de même que les processus biologiques peuvent réorganiser non seulement les éléments mais aussi les parties composées. Néanmoins, l'action de ces agents causaux ne peut pas être déduite ni prédite à partir des principes de composition des causes.

Nous allons voir par la suite que la notion d'agents causaux des lois hétéropathiques est reprise plus tard par Broad sous l'appellation de *forces dispositionnelles*.

Penchons-nous sur l'origine du mot *émergence*. J'ai déjà souligné que Lewes [Lewes, 1875] utilise *émergence* comme étant l'opposé de *résultante*. Une chose est émergente si elle a un effet hétéropathique au sens donné par Mill.

Les lois hétéropathiques de Mill seront des lois émergentes pour Lewes.

Voilà pour les lois émergentes. Le concept d'émergence a été aussi appliqué aux propriétés. C'est Samuel Alexander qui parla le premier de *qualités émergentes.*

> The emergence of a new quality from any level of existence means that at that level there comes into being a certain constellation or collocation of the motions belonging to that level, and this collocation possesses a new quality distinctive of the higher–complex. [Alexander, 1920]

L'action des agents causaux hétéropathiques dans la réorganisation des éléments composés ou des structures d'un niveau donne naissance à des caractéristiques ou des propriétés nouvelles et inexplicables. Elles sont nouvelles dans le sens qu'elles ne sont pas présentes aux niveaux inférieurs et elle sont inexplicables parce que ce sont les effets des agents causaux hétéropathiques et à ce titre sont donc imprévisibles. D'où il ressort qu'une qualité émergente n'admet pas d'explication et selon Alexander doit être acceptée au nom de la piété naturelle (*natural piety*).

Ayant décrit la genèse du terme *émergence*, je vais maintenant schématiser les postulats de l'émergentisme britannique.

8.2.2 Le *credo* des émergentistes britanniques

On pourrait résumer le *credo*, même au risque d'affaiblir quelques différences entre les courants divers, comme suit :

- Tout est composé de matière.
- La matière est granulée et discontinue. Au niveau le plus bas nous trouvons des particules élémentaires qui sont d'un type unique et néanmoins composent tous les types d'objets matériels.
- Rien n'a lieu sans le mouvement de quelques particules élémentaires mais néanmoins tous ces mouvements répondent aux lois de la mécanique. Au contraire, dans les structures des niveaux supérieurs, il y a des mouvements que ne répondent pas aux lois de la mécanique parce que leurs accélérations sont affectées par des agents causaux hétéropathiques.
- Il y a une hiérarchie des niveaux d'organisation selon la complexité dans l'organisation de ces particules matérielles. Les niveaux sont les suivants (du bas à haut) :
  1. strictement physique
  2. chimique
  3. biologique
  4. psychologique [5]

Chaque niveau possède une substance matérielle spécifique, ce qui n'empêche pas que le niveau soit totalement composé par la substance du niveau immédiatement inférieur. Le plus bas niveau est composé de particules élémentaires. Il existe des propriétés spéciales de la matière qui sont spécifiques à chaque type de substance appartenant à un niveau donné malgré le chevauchement que l'on vient de décrire .

Les propriétés spécifiques de chaque niveau constituent l'objet des différentes sciences spéciales.

La physique est une science commune à tous les niveaux puisque toute substance est composée de particules élémentaires. Ainsi, les propriétés physiques (propriétés d'inertie, gravitationnelle, des charges électriques, d'attraction, de répulsion) concernent tous les niveaux de complexité.

Les propriétés spéciales, par contre, ont leur origine dans les relations d'organisation de chaque niveau. Les sciences spéciales ont pour tâche la formulation des lois qui gouvernent ces relations particulières dues à l'organisation d'un niveau donné. Les émergentistes soutiennent que chaque type de structure interne correspondant à un niveau donné a des pouvoirs causaux à ce niveau.

[5] cf. [McLaughlin, 1992]

Selon Broad, certains organismes biologiques ont le pouvoir de respirer, de digérer et de se reproduire en vertu de leur organisation ou structure interne particulière. De même certains organismes, en vertu de l'organisation interne spécifique de leur système nerveux ont des pouvoirs cognitifs, de mémoire et d'association entre autres.

Les pouvoirs causaux ont deux caractéristiques particulières. Premièrement, les lois émergentes qui les gouvernent ne sont pas *réductibles à* des lois ou *dérivables de*[6]lois qui gouvernent les niveaux inférieurs de complexité. Cette caractéristique que l'on peut baptiser désormais l'*irréductibilité des lois* (mais aussi des propriétés) émergentes est revendiquée par certains auteurs qui prônent l'émergence pour caractériser certain phénomènes, surtout comme une solution alternative du problème corps–esprit. Donc, l'émergence se voit souvent caractérisée comme étant l'opposé de la réduction. Secondo, les pouvoirs causaux inhérents à l'organisation interne des structures de chaque niveau sont responsables de la production des mouvements de divers types à ce niveau.

La façon dont ces pouvoirs causaux agissent sur les mouvements des particules appartenant à un type donné de structure, soit-elle chimique, biologique ou psychologique est imprévisible.

Je soutiendrai plus loin que les courants émergentistes qui professent l'imprévisibilité et l'irréductibilité[7] des propriétés émergentes méritent la qualification d' *irrationnels*.[8] Quelle que soit la position prise quant à l'imprévisibilité de l'émergence - et McLaughlin nous met en garde contre les tentatives d'amalgame, - pour l'émergentiste du début du siècle il ne s'agissait pas d'affirmer qu'une chose s'avère être émergente parce qu'elle est imprévisible, mais plutôt de dire qu'elle peut être imprévisible parce qu'émergente. [McLaughlin, 1992, footnote page 73]

Jusqu'à maintenant j'ai présenté un panorama général de l'émergentisme britannique et dans les pages suivantes je traiterai de deux concepts centraux: le concept de *forces configurationnelles* et le problème de la *causalité descendante* (*downward causation*).

8.2.3 Les forces configurationnelles

Les pouvoirs causaux agissant sur les mouvements de particules typiques d'une structure à un niveau donné le font par l'intérmédiaire des *forces configurationnelles* aussi appelées *forces fondamentales*.

Les forces configurationnelles contrastent avec les forces qui se vérifient entre une paire de particules élémentaires. Ces dernières sont le résultat des pouvoirs d'attraction ou de répulsion dans une paire de ces particules. Dans le cas de la mécanique classique, les forces des paires de particules sont la force gravitationnelle et la force électromagnétique.

L'idée de l'existence de forces configurationnelles n'est pas totalement absurde à la lumière du développement scientifique de nos jours. On connaît, en effet des forces comme les forces de Van der Wass ou les forces Dipôle–Dipôle que l'on ne peut pas expliquer en termes de paires de particules. Néanmoins elles ne s'avèrent pas pour autant être des exemples de forces configurationnelles puisqu'elles sont des forces électromagnétiques.

L'irréductibilité des forces fondamentales à la physique met les émergentistes face à un dilemme. Si les forces fondamentales ne peuvent point être dérivées de la physique et si l'on considère qu'elles répondent aux lois qui gouvernent le mouvement, étant donné que la mécanique est la partie de la physique qui étudie les lois du mouvement, il en découle que seule une des deux propositions est acceptable: soit la mécanique ne doit être pas considérée comme une partie de la physique, soit les lois émergentes ne sont ni irréductibles ni imprévisibles.

En référence au paradoxe que j'ai présenté plus haut, la défense de Broad consiste à dire que les lois de la mécanique classique (la loi de l'inertie, de l'accélération et de la conservation de la masse) tiennent compte seulement des conditions générales du mouvement sans se soucier du type de substance dont il s'agit tandis que les lois des sciences spéciales prennent en considération ce dernier aspect, mise à part la masse inertielle.

[6]Le concept de réduction des lois doit être compris dans le sens du modèle de Nagel que j'ai exposé dans le chapitre 1 §1.4. Tandis que le concept de déduction signale la possibilité d'obtenir la loi du niveau supérieur comme conclusion d'un raisonnement qui a pour prémisses des lois du niveau inférieur.

[7]Je vais définir ces deux notions dans le cadre de l'émergentisme britannique dans le courant de ce chapitre.

[8]Je reprends ici la terminologie de Mario Bunge à cet égard. [Bunge, 1977, page 503]

A mon avis, la postulation des forces fondamentales est explicable dans le cadre du développement scientifique de l'époque. D'un côté les émergentistes défendaient, en général, le déterminisme causal. [9] Les mouvements des particules seront déterminés causalement par des forces de configuration. Ainsi, les propriétés ou arrangements dans l'espace de tous les objets existants seront le résultat de telles forces conjointement avec les forces entre paires de particules.

De l'autre, les émergentistes expliquaient ainsi les liens entre les éléments chimiques en fonction des forces configurationnelles spécifiques au niveau chimique. Ils niaient qu'elles pussent être microréductibles en termes de particules sub–atomiques. De façon analogue, les émergentistes suggéraient que dans les processus biologiques, les mouvements des particules à ce niveau étaient affectés par des forces vitales fondamentales. L'action de ces forces vitales consistait à accélérer les particules de ce niveau de façon telle qu'elles ne pouvaient pas être dérivées des influences décrites par des forces entre paires d'éléments ni par des forces jouant au niveau immédiatement inférieur qui est celui de la chimie.

Pour illustrer la situation de l'époque, je me propose de dresser une petite chronologie des faits scientifiques pertinents à notre discussion. Je laisserai de côté les événements du $XIX^{ème}$ siècle pour éviter de trop m'étendre.

1905 Théorie de la relativité restreinte. (Einstein).

1916 Théorie de la relativité générale. (Einstein).

1922 Série de conférences de Niels Bohr à l'Université de Göttingen sur la théorie quantique et la structure atomique : les nouvelles applications de son modèle d'atome comme un système solaire.

1923 Broad donne des conférences au *Trinity College* (Cambridge) sur les idées à paraître dans son livre *The mind and its place in Nature.*

1925 Publication du livre de Broad *The mind and its place in Nature.*

1926–1928 Heisenberg, Schrödinger et Dirac arrivent de manière indépendante aux théories de la mécanique quantique.

La théorie émergentiste de Broad se situe au cœur d'une formidable poussée scientifique qui a mené l'émergentisme britannique à sa chute.

Néanmoins, l'existence de forces fondamentales, bien que peu plausible, n'est pas contradictoire avec la mécanique quantique non–relativiste. [10] En effet, on pourrait faire la lecture suivante de l'équation de Schrödinger qui constitue la loi fondamentale de la mécanique quantique non–relativiste et qui gouverne l'évolution des systèmes dans le temps :

$$H\Phi = ih\frac{\partial \Phi}{\partial t}$$

[9] Considérons par exemple le paragraphe suivant de Mill :

> ... the state of the whole universe at any instant ... [is] the consequence of its state at the present instant; inasmuch that one who knew all the agents which exist at the present moment, their collocation in space, and all their properties, in other words, the laws of the agency, could predict the whole subsequent history of the universe [Mill, 1843, page 247]

La différence entre predictibilité et déterminisme causal tient à ce que ce sont deux thèses différentes parce qu'elles appartiennent à deux types de domaine différents. La première est une thèse *épistémologique* tandis que la deuxième est une thèse ontologique. Si l'on dit d'un événement qu'il est imprévisible, cela veut dire que dans l'état actuel des sciences on ne sera pas capable d'affirmer à l'avance si le dit événement aura lieu ou pas. En revanche, si l'on dit d'un événement qu'il n'est pas soumis au déterminisme causal c'est dire qu'il n'existe pas d'états des choses nomologiquement nécessaires pour qu'il advienne. Selon [McLaughlin, 1992, note bas de page 73] les émergentistes n'étaient pas toujours clairs à cet égard et utilisaient les deux termes *prévisibilité* et *déterminisme causal* comme s'ils fussent interchangeables. Selon [Achim, 1992, cf. page 29] pour Mill et Lewes les effets émergents ne seront pas prévisibles avant leur première occurrence

[10] [McLaughlin, 1992, cf. page 54]

où H est l'opérateur hamiltonien et h est la constante de Planck divisée par 2π. L'utilisation de l'équation nécessite l'hamiltonien, cette connaissance est une condition préalable. L'hamiltonien concerne justement l'énergie. Si l'on admet que l'on peut amalgamer l'énergie avec le concept de force, dans une perpective émergentiste on pourrait dire que le type d'énergie concerné par l'équation est spécifique au type de structure auquel elle se réfère. Or, la compatibilité entre les deux théories ne semble point impossible au premier abord.

Nonobstant cette possibilité de conciliation entre le point de vue émergentiste et la mécanique quantique non relativiste, l'existence de forces fondamentales semble devoir être désavouée à la lumière du progrès scientifique récent.

> Schrödinger's equation could be the fondamental equation governing motion in a world with energies that are specific to types of structures of particles that compose certain chemical, biological, and psychological kinds. But, as will become apparent, quantum mechanical explanations of chemical bonding in terms of electro-magneticism, and various advances this made possible in molecular biology and genetics-for exemple, the discovery of the molecular structure of DNA-make the main doctrine of British emergentism, so far as the chemical and biological are concerned at least, seem enormously implausible. Given the achievements of quantum mechanics and these other scientific theories, there seems to be not a scintilla of evidence that there are emergent causal powers or laws in the sense in question; there seems not a scintilla of evidence that there are configurational forces; and there seems not a scintilla of evidence that there is downward causation from the psychological, biological, or chemical levels. [McLaughlin, 1992, page 54]

Un des piliers du concept d'émergence selon cette école s'avère donc être faux. C'est une des causes, comme je l'ai déjà expliqué, de la chute de l'émergentisme.

Irréductibilité, émergence et forces fondamentales

De tout ce que l'on vient d'exposer on peut retenir que l'irréductibilité des propriétés émergentes a été prônée dans le but de réconcilier deux thèses qui semblaient irréconciliables. D'un côté la défense du déterminisme causal, de l'autre l'impossibilité d'expliquer certains phénomènes (comme par exemple les liens chimiques) en fonction de leur microstructure. L'inertie de près d'un siècle dans les travaux sur l'émergence n'a pas permis aux derniers émergentistes, dont Broad, de mettre à jour la théorie compte tenu des dernières découvertes. Ils ont gardé comme hypothèse l'existence de forces fondamentales pour pouvoir concilier le déterminisme causal avec l'imprévisibilité épistémologique.

La théorie de l'irréductibilité des propriétés émergentes au microniveau qui est née d'un déphasage avec la connaissance scientifique survit de nos jours. Je reviendrai sur le problème de l'irréductibilité comme trait distinctif des théories de l'émergence de la deuxième moitié de ce siècle.

Néanmoins les émergentistes ont attiré l'attention sur un fait qui sera pris en compte dans l'évolution scientifique postérieure : l'importance dans toute organisation des parties d'un tout et des relations qu'elles entretiennent soit entre elles-mêmes, soit avec la totalité devient cardinale pour le développement de la systémique; d'où l'on comprend mieux que le concept d'*auto-organisation* soit amalgamé avec celui d'émergence comme Francisco Varela et al. l'affirment dans la citation antérieure.

8.2.4 Causalité descendante

La *causalité descendante*[11] est dictée par deux facteurs différents. L'un d'eux est une certaine parcimonie ontologique propre à la tradition émergentiste : une des caractéristiques de l'existence d'une entité est son pouvoir causal. Or, les émergentistes ne peuvent pas accepter que les entités émergentes s'avèrent finalement être des épiphénomènes. [12]

[11] Ma traduction de *Downward causation*

[12] Samuel Alexander par exemple signale que la condition indispensable pour considérer une entité comme réelle est son pouvoir causal :

> ... it supposes something to exist in nature which has nothing to do, no purpose to serve, a species of *noblesse* which depends on the work of this inferiors, but is kept for show and might as well, as undoubtedly would in time, be abolished. [Alexander, 1920, page 1920-Vol. 2]

Le second facteur est le rôle que l'émergence joue dans l'explication de l'évolution.

Dans le premier cas, le fait que les entités aient un rôle causal pouvait être expliqué de deux façons différentes : soit les entités émergentes étaient des causes de propriétés appartenant au même niveau, soit elles jouissaient d'un certain rôle causal au niveau immédiatement inférieur.

Le rôle causal pour le niveau immédiatement supérieur n'entre pas en ligne de compte dans la discussion car les entités émergentes peuvent appartenir au dernier niveau et en conséquence ne trouvent là aucun niveau plus élevé pour y agir. La démarche consiste à démontrer la nécessité de la pertinence causale des entités émergentes et par conséquent le rôle qu'elles pourraient jouer au niveau immédiatement supérieur n'est pas intéressant dans ce sens.

La première des hypothèses, postulant que les pouvoirs causaux des propriétés émergentes se vérifient au même niveau d'appartenance, implique forcément la seconde hypothèse qui est la causalité descendante.

En effet, les entités émergentes bien que non–réductibles, bssent leur existence sur le niveau ou le degré de complexité du niveau inférieur. Ainsi, pour qu'une entité émergente puisse être recrutée comme étant la cause d'une autre du même niveau, il est indispensable que des conditions physiques de complexité soient présentes au niveau inférieur. Autrement dit, il faut que la propriété du niveau, disons n ait des pouvoirs causaux sur le niveau $n-1$ pour être la cause d'une entité émergente au niveau n.

Kim nous met en garde au cas où cette démarche nous semblerait bizarre. Il nous dit que finalement elle est tout à fait analogue à celle du matérialisme non–réductionniste pour le problème corps–esprit. [Kim, 1992b, cf. page 121] Dans ce dernier cas, il s'agit de récuser l'épiphénoménalisme des entités mentales mais non pas en leur attribuant des pouvoirs, des rôles causaux pour d'autres entités mentales. On vise, par contre, à signaler des pouvoirs causaux ou du moins des corrélations entre les entités physiques et les entités mentales de façon à expliquer les dernières à partir des premières et à pouvoir de ce fait récuser la qualification d'épiphénoménalisme à leur égard. Rappelons ici les corrélations entre les propriétés intentionnelles et les propriétés computationnelles indispensables à la démarche fodorienne exposée dans le chapitre 7.

Le rôle de la causalité descendante dans l'évolutionnisme émergentiste est expliqué par C. Lloyd Morgan qui publie en 1923 le livre *Emergent Evolution*. Morgan fait un mariage entre une thèse darwinienne de l'évolution et l'idée d'émergence. Selon lui, tout au long du processus d'évolution il y a émergence des phénomènes complexes qui sont nouveaux et non prévisibles. [13] De même que les autres émergentistes, il croit à l'existence de divers niveaux d'organisation selon la complexité de leur substance et il souscrit à la conception des sciences spéciales que j'ai décrite plus haut.

A l'opposé des autres émergentistes dont Broad, il ne parle pas des forces de configuration. McLaughlin signale que pour Morgan le mot *force* semblait ambigu; il craignait, en outre d'être accusé de postuler une intermédiation spéciale qui puisse être interprétée comme un succédané de l'*entéléchie*. Il préférait parler des facteurs causaux qui influenceraient l'accélération des particules. [McLaughlin, 1992, cf. page 68]

Comment la causalité descendante doit–elle être comprise dans le cadre de l'évolution émergentiste? Je pense qu'une façon de l'interpréter est reflétée par l'exemple suivant. Dans le cours de l'évolution pour certaines espèces, par exemple des oiseaux, certaines capacités perceptives et instinctives ont émergé. Ces capacités font que les oiseaux peuvent développer des conduites pour satisfaire leurs besoins. Supposons par exemple l'état d'avoir faim; l'oiseau va voler d'un endroit à l'autre et avec l'aide de son système perceptif, il cherchera de la nourriture de même qu'à l'aide de son système moteur, il va la capturer pour satisfaire ce besoin. Les capacités que je viens de citer seront considérées comme émergentes dans l'évolution et appartenant au domaine vital (donc de haut niveau) dans le cadre émergentiste.

Maintenant, toutes ces activités perceptives et motrices sont réalisées dans le cerveau et sont les résultantes de certains états du cerveau. Ces états n'existeraient pas si l'oiseau en question n'était

Selon Alexander, lorsqu'une entité n'a pas de pouvoirs causaux son existence ne fait absolument aucune différence pour le reste des autres choses existantes et elle ne devra donc point être considérée réelle.[Kim, 1992b, cf. page 134]

[13] Pour un résumé du rôle du concept d'émergence dans la tradition de l'évolutionnisme néo–darwinien voir [Blitz, 1990]

pas arrivé à cette étape de développement phylogénétique; ceci justifie, aux yeux des émergentistes que les propriétés émergentes du niveau supérieur, dans notre cas, la condition vitale, aient un rôle causal sur le niveau inférieur, le niveau physico–chimique.

Achim signale que l'expression la plus ancienne du concept de causalité descendante se trouve dans la citation suivante d'Alexander qu'il commente de la façon suivante :

> When some new kind of relatedness is supervenient (say at the level of life), the way in which the physical events which are involved run their course is different in virtue of its presence – different from what it would have been if life had been absent [Alexander, 1920, page 16]

> This passage can be understood as affirming that the course of the physico–chemical processes would be different if they were involved in vital processes in comparison to their absence that is to say, vital processes influence the lower processes. This looks like an early version of downward causation. [Achim, 1992, page 42].

Alexander fait dans le texte précédent une identification entre les processus physico–chimiques et le processus mental. Morgan, en revanche, ne conclut point à une identité car il est un dualiste des propriétés. Or, selon lui, les processus mentaux ne sont pas des épiphénomènes des processus neuraux parce que la caractéristique d'être mental s'ajoute à celle d'être neuronal de façon telle que le processus en question serait différent s'il n'avait pas la caractéristique mentale. [Morgan, 1923, cf. page 8-9]

Achim entre autres, nous propose d'interpréter la causalité descendante comme suit :

> Lorsqu'un processus B est déterminé par un processus A, alors le processus A s'avère différent en l'absence du processus B. Autrement dit, le processus B ne sera plus survenant du processus A. [Achim, 1992, cf. page 42]

Si l'on veut que la proposition ci–dessus ait véritablement un sens, il faudrait savoir dans quel sens les situations causées par le processus A sur lequel survient le processus B (appelons cette situation une situation–b) seront différentes de celles où ce dernier processus est absent (désormais situation–$\neg b$).

Le fait d'arguer la présence du processus B ne réfute en rien la thèse de l'épiphénoménalisme de B puisque selon la profession de foi des émergentistes, exister signifie avoir des pouvoirs causaux et supposer d'emblée l'existence de B est tout simplement une affirmation de principe. Ainsi, l'unique différence qui compte est l'existence des lois émergentes dans la situation–b (et l'absence de ces lois dans la situation $-\neg b$). Les lois émergentes sont celles qu'expriment les règles gouvernant les pouvoirs causaux de B et qui ne sont point *réductibles* à des propriétés des niveaux inférieurs.

Il faudra donc montrer la *nécessité causale* du processus B en relation avec des événements appartenant au niveau inférieur. Le procédé de démonstration de cette nécessité causale consiste à déterminer qu'en l'absence de B le niveau inférieur serait différent. Je pense que l'on peut donner à cette expression deux interprétations possibles. La première constitue une critique difficile à surmonter. La seconde interprétation est plus indulgente car elle place cette expression dans le cadre d'une causalité téléologique, qui est caractéristique d'un cadre évolutionniste comme celui–ci.

L'analyse de l'expression proposée plus haut par [Achim, 1992, cf. page 42] montre qu'elle constitue tout simplement un truisme.

Lorsqu'un processus B est déterminé par un processus A, alors le fait que le processus A soit différent en l'absence du processus B est trivialement vrai. Si A cause nécessairement B il est évident que s'il ne causait pas B la relation causale serait inexistante (puisque sa conséquence n'existe pas) et aussi que A serait différent (puisqu'il n'aurait pas les mêmes rôles causaux). Si ceci peut être utilisé pour démontrer l'existence d'un rôle causal de B sur A, il s'agit d'une erreur consistant à prendre les conséquences pour les causes. L'expression d'Achim d'après cette première interprétation ne garantit en rien le rôle causal de B.

La seconde interprétation que je puis donner est, comme je l'ai remarqué, plus charitable. Il s'agit de faire une interprétation dans un cadre causal téléologique. Cette interprétation est dictée par le critère d'inspiration aristotélicienne selon lequel *les effets se transforment en buts et les*

causes en significations. [14]

Prenons le cas de l'apparition des yeux dans les espèces animales comme un effet. Dans la tradition téléologique nous pouvons faire la lecture qui suit. Il était une fois un animal, le nautile dont l'œil était un trou qui permettait le passage de la lumière. D'autres animaux ont développé des lentilles de façon à accroître cette capacité de passage. Cependant, ces lentilles ont introduit une complication additionnelle car elles ne permettaient de voir clairement que les objets situés à une certaine distance. Alors on (la nature? le hasard?) a fait le nécessaire et les animaux ont créé des muscles permettant de mouvoir la lentille plus ou moins loin du récepteur. Ainsi, les animaux ont pu accomplir une série de fonctions vitales telles que se nourrir ou échapper aux prédateurs de façon plus performante.

L'application du critère aristotélicien cité plus haut aura pour résultat que l'effet, en l'occurrence l'œil n'eût pu exister si une série de causes et de conditions conjointes n'avait permis son apparition. [Espinoza, 1990, cf. page 184–185]

Revenons maintenant à l'exemple de notre oiseau. Supposons que nous disions que l'état (ou la suite des états) physique(s) de son cerveau, la cause effective de ce comportement (ici et maintenant) lui permettant de voler pour rechercher de la nourriture à l'aide de son système perceptif (constitué des capacités émergentes relevant du niveau vital) ne serait point tel qu'il est en l'absence de ces mêmes capacités.

Cet énoncé est un truisme si on l'analyse dans une perspective ontogénétique mais il pourrait avoir un sens à partir d'une perspective phylogénétique.

L'interprétation selon laquelle les capacités dont les oiseaux font preuve maintenant sont basées sur les capacités gagnées au cours du développement de l'espèce (i.e. une certaine maturité du système perceptif, de la capacité motrice) est plus acceptable. Ainsi, tout le long du développement de l'espèce il y a eu des transformations de la structure physico–chimique dues en partie, au moins, aux capacités émergentes de la période précédente comme j'ai essayé de l'illustrer par le cas de l'œil.

Je pense que c'est seulement à partir de cette perspective qu'on peut donner aux propriétés émergentes un rôle sinon causal (dans le sens de cause effective) du moins explicatif *post–facto* ou tout simplement de description téléologique. Selon [Hempel and Oppenheim, 1948, cf. page 144] les considérations de type téléologique font office d'approximations heuristiques donnant lieu, par la suite le plus souvent à des résultats exprimés en termes non–téléologiques. Ces résultats serviront à accroître la connaissance scientifique des connexions causales, mais les énoncés téléologiques ne pourront, en aucun cas, être considérés comme des lois. Il me semble que c'est seulement à partir de cette perspective phylogénétique que l'on peut trouver un sens à la *causalité descendante*. Néanmoins, celle–ci méritera à peine la caractérisation causale.

Jusqu'ici j'ai dépeint les points saillants les plus pertinents de l'émergentisme britannique en préambule à mon propre exposé. Par la suite, je vais décrire l'évolution du concept d'émergence durant deux époques différentes. La première période comprend la deuxième guerre mondiale et les années de l'immédiat après–guerre et la seconde concerne les concepts d'émergence durant les années soixante–dix, notamment la théorie sur l'émergence de Mario Bunge.

8.3 L'émergence de l'après guerre

Selon la recherche bibliographique que j'ai pu faire, il me semble que la polémique sur le concept d'émergence soit revenue sur la scène philosophique pendant les années 40–50. Une des raisons de ce retour était, à mon avis, la mise à jour du concept d'émergence par rapport aux concepts de prévisibilité, de réduction et de microréduction proposés par Thomas Nagel, Carl G. Hempel, Paul Oppenheim et Hilary Putnam.

[14] J'ai emprunté l'énoncé de ce critère à Miguel Espinoza. Il l'appelle *le critère d'Aristote–Bousset–Janet* et cite ce dernier [Espinoza, 1990, cf. page 185]:

> [...] if there are proportions well taken, proper to certain effects, then there are goals.[Janet, 1901, page 16]

8.3.1 L'émergence selon Carl Hempel et Paul Oppenheim

Hempel et Oppenheim dans [Hempel and Oppenheim, 1948] dédient la deuxième partie de ce texte au concept d'émergence[15].

Ces auteurs caractérisent un phénomène comme émergent s'il résiste à toute tentative d'explication en termes des éléments du microniveau. Par ailleurs ce texte introduit clairement le fait que les explications qui utilisent le concept d'émergence impliquent la détermination de différents niveaux dans la nature. Néanmoins, les explications basées sur le microniveau sont les seules que les auteurs reconnaissent valables pour la connaissance scientifique.

Dans ce même texte ces auteurs critiquent les explications qui utilisent les concepts émergents à cette époque. Ces critiques sont les suivantes :

Première critique: Le caractère ambigu de la définition de la totalité et de ses parties. Bien que l'on soit d'accord avec le postulat de la *Gestaltthéorie* qui dit que le tout est plus que la somme de ses parties, il s'avère que très souvent la relation "partie de" n'est pas définie clairement d'un contexte à l'autre.

Deuxième critique: Le caractère relatif d'une propriété émergente attribuée à un tout. En effet, elle dépend des parties définies et de l'ensemble des caractéristiques de ces dernières. L'un des exemples typiques donné par les émergentistes britanniques est l'émergence des propriétés telles que la transparence et la liquidité de l'eau qui n'étaient pas des propriétés de ses parties : l'hydrogène et l'oxygène. Cependant on a vu par la suite que l'on pourrait inférer les propriétés de l'eau susmentionnées à partir de certaines propriétés de ses composés, ce qui sert d'exemple pour éclairer la signification de la deuxième critique faite par les auteurs.

Je pense que la fixation d'un contexte de référence formé par la définition des composantes et la relation "fait partie de" conjointement avec la classe des propriétés des composantes qui sont prises en compte sera critique à l'élucidation du caractère parfois ambigu ou même trivial de certains faits considérés comme émergents. Le caractère relatif de l'émergence devrait être une des premières choses à prendre en compte chaque fois que l'on utilise le terme. Dans les sections suivantes je vais rapporter une tentative de Paul Teller visant à établir des critères propres à éviter les explications émergentes triviales des caractéristiques qui ne le sont pas.

Revenons maintenant à la troisième critique de Hempel et d'Oppenheim qui vise le caractère imprévisible des choses émergentes.

Troisième critique: Le caractère d'imprévisibilité donné à un objet sur la base de l'information de ses parties dépend des lois ou des théories disponibles au moment de cette caractérisation. Ils maintiennent explicitement que

> [...] if a theory does not include this micro–macro law, then the phenomenon is emergent with respect to that theory. [...] Emergence of a characteristic is not an ontological trait inherent in some phenomena; rather it is indicative of the scope of our knowledge at a given time; thus it has no absolute, but a relative character; and what is emergent with respect to the theories available today may lose its emergent status tomorrow. [Hempel and Oppenheim, 1948, page 149–150]

Cette critique est également importante parce qu'elle souligne le risque de transformer le concept d'émergence en un concept épistémologique puisque relatif aux connaissances du moment. Ainsi, une propriété sera émergente jusqu'à preuve du contraire. Ceci met certains émergentistes de l'époque dans le dilemme suivant : soit ils acceptent que le concept d'émergence soit un concept dépendant foncièrement de l'état de la connaissance scientifique, soit ils admettent la possibilité que certaines propriétés soient imprévisibles de manière absolue, c'est à dire quelque soit l'état de la connaissance scientifique en obstruant le chemin qui mène à l'unité de la science.

Des auteurs comme par Arthur Pap se sont trouvés aux prises avec cette situation. [Pap, 1951] Ce dernier essaie de s'en sortir en proposant une analyse sémantique du concept d'émergence qu'il

[15] Ils signalent que cette deuxième partie résulte de la correspondance et des discussions avec le Professeur Kurt Grelling et ils en font un hommage à ce dernier et à sa femme. Tous deux furent victimes des nazis pendant la guerre.[Hempel and Oppenheim, 1948, page 135 note bas de la page 1]

appelle émergence absoute. Je ne vais pas discuter la théorie sémantique de l'émergence absoute ici car, entre autres raisons, je considère que la condition d'imprévisibilité n'est qu'un cliché hérité des anciens émergentistes. Néanmoins, selon Pap une loi qui établit la corrélation entre une qualité Q et ses conditions causales est à priori imprévisible si le prédicat qui désigne Q est définissable seulement de façon ostensible [Pap, 1951, cf. page 304]. Par exemple l'odeur d'un gaz ne sera pas déductible de sa structure moléculaire.

Finalement, Hempel et Oppenheim donnent la définition suivante de l'émergence:

> **The occurrence of a characteristic W in an objet w is emergent relatively to a theory T, a part relation Pt, and a class G of attributes if that occurrence cannot be deduced by means of T from the characterization of Pt − −*parts* of w with respect to all the attributes in G.[Hempel and Oppenheim, 1948, page 151]**

La définition ci-dessus de l'émergence admet parmi ses traits fondamentaux l'impossibilité de déduction des caractéristiques pour une occurrence donnée de l'entité émergeote et elle précise aussi qu'il faut fixer la relation Pt et les attributs des parties. Cette conception de l'émergence est qualifiée par ses auteurs de conception *rationnelle*. Par la suite, je présenterai l'intéressante théorie de l'émergence de Mario Bunge et je maintiendrai avec lui que les théories émergentes rationnelles sont celles qui ne basent pas ce concept sur son irréductibilité à partir des caractéristiques des parties mais plutôt sur une bonne définition de la relation Pt qui ne doit pas être considérée comme l'équivalent de la relation inclusive.

Dans la même ligne de pensée Bilary Putnam et Paul Oppenheim publient le fameux texte sur l'hypothèse de l'unité de la science [Oppenheim and Putnam, 1958] dans lequel les auteurs nous mettent en garde quant à leur conception de la relation Pt.

Il signalent que dans un sens très large, cette hypothèse est assimilée à la relation d'inclusion. Cependant, ils veulent se placer dans un perspective plus étroite et considérer la totalité comme douée d'une structure d'organisation. La façon de procéder comprendra deux étapes. Premièrement, la construction d'un système de calcul où la relation Pt prise dans sa conception étroite sera une des notions primitives. Secondo, la définition d'une relation particulière Pt satisfera les axiomes de ce système de calcul [Oppenheim and Putnam, 1958, cf. page 11].

Cette conception sera reprise, plus tard, par Mario Bunge pour bâtir sa théorie de l'émergence. Mais Oppenheim et Putnam se tiennent à une conception épistémologique de l'émergence qui a comme caractéristique principale l'irréductibilité aux caractéristiques de niveaux inférieurs [Oppenheim and Putnam, 1958, cf. page 15]. Pour eux, réduction veut dire réduction ontologique, ils sont installés dans la dynamique du "nothing but". Dans cette optique les phénomènes émergents ne seront qu'un cumul des faits en l'attente des jours meilleurs de la science.

Dans la section suivante je vais illustrer les efforts de certains auteurs qui cherchent une conception de l'irréductibilité des propriétés émergentes qui soit compatible avec l'hypothèse de l'unité des sciences de façon à vider le concept d'émergence de tout soupçon d'irrationalité. Néanmoins cet objectif n'est pas atteint parce que ces efforts vont aussi dans le sens d'un renforcement de la dichotomie entre émergence et réduction.

8.4 L'émergence et l'irréductibilité

8.4.1 Les stratégies de conciliation des deux concepts

Certaios auteurs contemporaines de tendance émergentistes ont entrepris des démarches conciliatrices dans le but de préserver le caractère irréductible du concept d'émergence.

Je vais citer trois stratégies différentes qui ont en commun une exégèse du texte [Broad, 1925]. Le choix de Broad comme référence est évident; en effet Braod a été le dernier des émergentistes britaniques a produire un *corpus* plus ou moins complet sur le sujet.

Chacune de ces stratégies adopte une théorie différente de la réduction postérieure à Broad et elle essaie de rendre compatible le concept de réduction qui lui est propre avec celui de Broad. Il s'agit donc d'un aggiornamento de la position de cet auteur à la lumière de la philosophie des sciences actuelles.

Si ces démarches aboutissent, on pourrait conclure que le concept d'émergence a une place dans la théorie des sciences modernes comme l'opposé de la réduction ou de la réalisation. L'objectif de cette partie est de démontrer qu'aucune de ces démarches n'est réussie; cette critique est capitale pour la conclusion que j'avance, selon laquelle l'émergence est l'inverse de la réalisation et non son contraire.

Le concept d'irréductibilité de Broad est confronté à la théorie de Thomas Nagel; plus précisément la clé de la démarche se trouverait dans l'existence des lois-ponts. La deuxième tentative met en jeu la théorie sur la microréduction de Robert Causey. Enfin, la troisième fait appel au concept de réalisation dû à Robert Cummins.

D'abord il faut préciser dans quel sens les propriétés ou entités émergentes seront irréductibles selon la conception de Broad.

8.4.2 L'irréductibilité selon Broad

Selon Brian P. McLaughlin[16] les points pertinents dont il faut tenir compte pour comprendre le concept d'irréductibilité et d'émergence de Broad sont les suivants: premièrement, la différence entre lois intra-ordinales et lois trans-ordinales; deuxièmement, le caractère sémantique de la déduction et troisièmement le cadre spécifique où le problème de changement de vocabulaire entre le micro-niveau et le macro-niveau se pose.

Pour Broad il existe deux types de lois. D'abord, il y a des lois *intra-ordinales* qui traitent des relations entre les propriétés des éléments d'un même niveau. Les lois intra-ordinales sont l'objet des sciences spéciales, car comme je l'ai signalé plus haut, les sciences spéciales traitent des relations entre les éléments composant un même niveau.

Ensuite, les lois *trans-ordinales* gouvernent les relations entre les propriétés des composants de deux niveaux consécutifs. Broad donne la définition suivante de ces lois:

> *A* and *B* would be adjacent, and in ascending order, if every aggregate of order *B* is composed of aggregates of order *A*, and if it has certain properties which no aggregate of order *A* possesses and which cannot be deduced from *A*-properties and the structure of the *B*-complex by any law of composition which is manifested itself at lower-levels. ... A transordinal law would be a statement of the irreductible fact that an aggregate composed of aggregates of the next lower order in such and such proportions and arrangements has such and such characteristic and non-deducible properties [Broad, 1925, page 77-78]

Néanmoins, Broad ne considérait pas que les lois trans-ordinales fussent nécessairement émergentes. La condition d'émergence pour une loi trans-ordinale est une question empirique; elle ne peut pas être établie *a priori*. Cependant, si une loi trans-ordinale s'avère émergente Broad nous dit qu'il faut la considérer comme un *fait nomologique brut* car on ne peut pas l'expliquer. On peut les interpréter comme des lois fondamentales au sens où elles ne sont pas dérivables des autres lois ou des autres propriétés. Le fait qu'elles ne soient pas dérivables ou déductibles des lois du niveau inférieur, des conditions et des principes de composition[17] ne signifie pas, pour autant qu'il s'agisse d'un simple problème de changement de vocabulaire entre des termes du niveau supérieur (p. ex. *le pouvoir de reproduction*) et du niveau inférieur (le niveau de la chimie). McLaughlin rejette cette démarche car, selon lui, elle n'est qu'une trivialisation de la position de Broad.[McLaughlin, 1992, cf. page 51]

Le deuxième point que j'ai signalé comme pertinent porte sur la notion du concept de déduction chez Broad. Les éléments qui méritent d'être relevés sont les suivants:

- La notion de déduction est sémantique et non-formelle. Or, pour Broad *B* est déductible d'*A* si lorsqu'*A* est vrai, alors *B* est vrai. Ceci exclut que *B* soit déductible de *A* en fonction de leur forme logique.

- Dans la déduction des lois trans-ordinales il acceptera l'utilisation des identités entre des éléments. Ces propositions d'identité seront des propositions *a posteriori*. Voilà pourquoi il

[16][McLaughlin, 1992]

[17]Les principes ou lois de composition d'un niveau inférieur gouvernent les relations entre les propriétés de ces substances des niveaux inférieurs et les propriétés des composantes du niveau dont ces lois relèvent.

passe du vocabulaire du macro–niveau à celui du micro–niveau sans que ceci le gêne outre mesure. (p. ex. il utilise de façon indistincte le terme *eau* et H_2O).

- Les lois pour Broad ne sont pas des entités linguistiques car il les considère comme des relations singulières entre des universaux. Ceci veut dire qu'il accepterait l'égalité suivante: La loi selon laquelle $NaCl$ se dissout en H_2O= la loi selon laquelle le sel se dissout dans l'eau.

- Le problème du changement de vocabulaire entre le micro–niveau et le macro–niveau n'est pas pertinent dans le cadre de lois trans–ordinales. La raison est que les lois trans–ordinales gouvernent les relations entre les éléments de niveau plus bas et les composantes du niveau immédiatement supérieur mais tous les niveaux sont composés à la base par le même type de particules élémentaires. La différence entre les composantes, rappelons–le, était basée sur les caractéristiques spatio–temporelles de chaque niveau et celles–ci sont du domaine des lois intra–ordinales. Les lois trans–ordinales peuvent changer d'un vocabulaire à un autre parce que ces différences ne sont pas pertinentes au type de relations que ces dernières lois gouvernent.

Bien que le problème de changement de vocabulaire du micro–niveau au macro–niveau ne se pose pas pour les lois, il s'avère en revanche pertinent pour les énoncés des propriétés issues de ces lois.

La raison est la suivante [McLaughlin, 1992, cf. page 82–83]: d'abord Broad ne pense pas que deux prédicats aient besoin d'être synonymes pour exprimer la même propriété. Le fait que ces prédicats expriment une même propriété est établi *a posteriori*; or la réductibilité d'une propriété d'un certain type de composante appartenant à un certain niveau à une autre propriété appartenant à une composante du niveau inférieur est aussi déterminée *a posteriori*. Seulement alors pourra–t–on dire s'il s'agit d'une propriété réductible ou émergente.

Maintenant on est en mesure de définir le concept de loi trans–ordinale non–émergente.

DEFINITION 7 (LOI TRANS–ORDINALE NON–ÉMERGENTE)
Une loi trans–ordinale est non–émergente si l'on peut la réduire, en principe, à des lois de niveau plus bas, à des principes de composition et à des énoncés d'identité des propriétés a posteriori.

Résumons: les lois trans–ordinales établissent les relations entre les propriétés de deux niveaux consécutifs. Le fait qu'elles soient réductibles ou émergentes dans le sens décrit par la définition citée ci–dessus définit les propriétés du niveau supérieur et qui sont issues d'elles comme étant respectivement non–émergentes ou émergentes. L'identité entre propriétés est établie a posteriori.

Si une propriété d'un niveau supérieur est réductible alors on peut utiliser la propriété du niveau inférieur à sa place. [18] Maintenant la question que l'on se pose est dans quelle mesure l'on peut assimiler les lois trans–ordinales aux lois prises dans le sens de Ernest Naguel.

8.4.3 Première tentative: Lois trans–ordinales et lois–ponts nageliennes

Selon McLaughlin, Broad n'admet pas que les lois–ponts soient suffisantes pour la réduction. J'en exposerai les raisons plus bas, mais il faut d'abord donner quelques précisions.

Premièrement, Broad n'admet pas comme *loi* un énoncé qui affirmerait que si une propriété Q est non–émergente alors la propriété P qui est co–extensive à Q est elle aussi non–émergente. La raison est la suivante; cette loi qui affirme la non–émergence de P à partir de celle de Q en raison de leur co–extension est elle–même une loi émergente. Elle est émergente parce qu'elle ne vérifie pas la définition de loi trans–ordinale émergente donnée ci–dessus.[19] En second lieu, souvenons–nous que d'après la stratégie réductive de Nagel[20], la théorie réductrice est composée de ses propres lois

[18] Je ne fais ici aucune hypothèse sur le type de réduction dont il s'agit, qu'elle soit ontologique ou épistémologique.
[19] En effet, une telle loi n'est pas réductible à des lois de niveau plus bas, à des principes de composition ni à des propriétés *a posteriori* étant donné que l'unique donnée dont nous disposions est leur coextension.
[20] Voir chapitre 1 §1.4

plus un ensemble d'énoncés biconditionnels universels qui mettent en relation le vocabulaire de la théorie réduite avec celui de la théorie réductrice. Ces énoncés sont les lois-ponts.

On peut considérer que Broad souscrit à la notion nagelienne de réduction selon laquelle les lois de la théorie réduite peuvent être déduites de celles de la théorie réductrice.

Pour que les lois trans-ordinales à la Broad puissent être considérées équivalentes aux lois-ponts nageliennes, on doit les considérer comme étant de type formel et non matériel. Cependant, cette caractérisation ne convient pas à l'interprétation broadienne car, comme nous l'avons déjà vu précédemment, celle-ci est sémantique.

En effet, lorsque l'on part d'une perspective émergentiste, les lois-ponts s'avèrent insuffisantes pour formuler une théorie de la réduction. La raison est que ces lois-ponts dans le cadre broadien empruntent la forme de lois trans-ordinales sauf qu'elles ne sont pas expliquées car elles n'entrent pas dans la théorie nagelienne. Les lois en question seront émergentes dans le sens de Broad, et donc irréductibles par définition. Cela veut dire que les propriétés qui en découlent seront aussi émergentes dans le sens de Broad.

En conclusion, le concept de réduction de Broad peut difficilement être mis en corrélation avec celui de Nagel car si l'on tient compte de leurs caractéristiques importantes les deux théories demeurent séparées. En particulier, je viens de signaler, étant donnée la conception sémantique des lois de Broad, que la simple coextension entre deux propriétés ne suffit pas pour que la réductibilité de l'une implique la réductibilité de celle qui est sa contrepartie selon une loi-pont. Je pense donc que l'on peut considérer cette tentative comme un échec.

8.4.4 La deuxième tentative : la théorie de Causey

Dans les années soixante-dix il y eut un mouvement critique à l'encontre du concept de réduction de Nagel. Un des auteurs de ces critiques est Robert Causey. Il attaque notamment le recours que Nagel eut aux lois-ponts comme éléments de la réduction.

Causey s'oppose à l'idée de Kripke selon laquelle l'identité invoquée par les lois-ponts est métaphysiquement nécessaire. Pour lui, une théorie de la réduction comprendra en tous cas l'énonciation des identités mais ces identités doivent être justifiées [Causey, 1977, cf. Chapitre 4]. Ces justifications ne seront considérées qu'*a posteriori*. Ce dernier trait du concept de réduction de Causey permet d'espérer une conciliation et le rend plus compatible que la pensée de Nagel avec le concept de réduction de Broad.[21]

Il nous faut d'abord énoncer quelques précisions sur la stratégie de Causey. Pour cet auteur, une théorie T est constituée de trois types différents d'énoncés: Les énoncés F qui sont l'ensemble des lois fondamentales de la théorie de T; l'ensemble I des énoncés vrais et des énoncés des propriétés d'identités; l'ensemble D qui contient l'ensemble des lois dérivées de T.

Si la théorie T se réfère à une totalité, le domaine d'application de T sera constitué aussi bien des éléments composés que des éléments de base. Les éléments de base de T répondent à des énoncés ouverts de T qui expriment seulement des caractéristiques de ces éléments de base. Une situation analogue se vérifie pour les parties composées.

Les éléments de base sont aussi appelés par Causey éléments libres quand ils ne sont pas combinés avec d'autres éléments de base pour former un élément composé. Les éléments composés sont aussi appelés éléments liés (*bound*).

Mais comment opère-t-on la réduction des éléments composés aux éléments de base? La relation entre les éléments de deux niveaux se fait à l'aide des prédicats quantifiés et d'un ensemble de terminologies théoriques selon lequel les prédicats qui expriment les attributs des éléments composés seront définissables en termes des prédicats exprimant des attributs des éléments de base.

La différence entre lois fondamentales et lois dérivées est la suivante: Les lois fondamentales traitent des attributs des éléments de base ou éléments libres dans une situation donnée de l'environnement. Il y a aussi dans ce groupe des lois qui traitent des relations structurelles entre les éléments libres ou éléments de base. Les lois dérivées sont les lois qui traitent des attributs des

[21] [McLaughlin, 1992, cf. 84]

composés et on justifie ce choix en fonction du concept même de microréduction. En effet, le but de la réduction est d'expliquer le comportement de la totalité en termes des comportements des parties. Or, il est nécessaire d'être en mesure de déduire les lois de T qui relèvent des composés à partir de celles que s'appliquent aux éléments de base. Voilà pourquoi les lois des composés doivent être des lois dérivées.

Le problème qui se pose est celui de la relation entre deux théories : quand une théorie T sera-t-elle micro-réductible dans le sens de Causey à une théorie $\bar{T}$? La micro-réduction sera possible seulement si les lois de T sont déductibles des *lois fondamentales* de $\bar{T}$. Néanmoins, on acceptera aussi les énoncés suivants :

1. Les énoncés d'identité des propriétés.

2. Les énoncés exprimant les conditions aux limites (*boundary condition*).

3. Tout énoncé analytique.

Selon McLaughlin ces conditions s'avéraient trop strictes pour Broad. Ce dernier aurait accepté en plus comme éléments de la micro-réduction les principes de composition comme par exemple le principe d'addition de masse qui est contingent selon Broad [22]. N'oublions pas que pour Broad les lois qui expriment des propriétés des éléments composés (dans l'acception donnée par Causey) peuvent être déduites des lois des éléments libres (toujours selon la définition de Causey) seulement si l'on fait appel aux principes de composition (au sens qu'en donne Broad).

A la question : les lois trans-ordinales émergentes de Broad s'avèrent-elles microréductibles au sens de Causey selon Broad? La réponse de McLaughlin est négative et je suis de son avis. Une loi trans-ordinale ne peut être un énoncé analytique pour Broad. Elle devra octroyer une certaine propriété de niveau inférieur à un élément composé du niveau immédiatement supérieur. Si l'on tient compte de la définition de la loi trans-ordinale émergente de Broad, cette loi ne sera pas déductible à partir des lois fondamentales (dans le sens de Causey) du niveau inférieur, aux propriétés d'identité (dans le sens de Causey qui coïncide avec le sens de Broad), ni à des principes de composition du niveau plus bas (dans le sens de Broad) ni à d'autres principes analytiques. Ainsi les lois trans-ordinales émergentes de Broad ne seront pas micro-réductibles au sens où l'entend Causey.

D'après tout ce que l'on vient de voir, la deuxième tentative de rendre compatible le concept de réduction de Broad avec celui de Causey semble plus adéquate que la précédente mais toutefois pas tout à fait satisfaisante. En effet, bien que les lois trans-ordinales émergentes dans le sens de Broad continuent d'être non réductibles au sens de Causey, les règles établies par ce dernier pour cette réduction se révèlent trop strictes pour Broad. En particulier, il semble que certaines lois considérées non-réductibles par Causey ne soient pas émergentes selon Broad. Il ne serait donc pas acceptable à mon avis de faire l' amalgame entre les deux théories qui justifient la dichotomie émergence-réduction que certains préconisent.

8.4.5 La troisième tentative : les propriétés déductives à la Cummins

Le choix du modèle de Cummins[23] donne à l'analyse un tournant très intéressant car il ne base pas l'analyse du concept de réduction sur la différence entre synthétique et analytique comme dans les cas précédents mais il met plutôt l'accent sur l'organisation des structures de microniveau.

Dans son livre [Cummins, 1983] cet auteur établit une différence entre deux types de théories scientifiques. D'un côté on trouve les théories qui visent à expliquer les événements, voire les changements et de l'autre, les théories destinées aux explications de propriétés.[24]

Selon [Cummins, 1983, cf. page 1] les premières sont des théories de transition qui ont comme but fondamental la réponse à la question *quand*.

[22] Il est possible que ce principe soit considéré analytique par Causey et alors qu'il soit, après tout, un des éléments de la réduction.

[23] cf. [Berckermann, 1992] J'ai déjà exposé cette théorie dans le chapitre 2 §. 2.3.4..

[24] Voir chapitre 2 §2.3.4.

Les autres théories qui tendent à expliquer des propriétés ne doivent pas être comprises comme répondant à la question "Pourquoi le système S a–t–il acquis la propriété P? " mais plutôt "En vertu de quoi S possède–t–il maintenant la propriété P? "[25]. Selon Cummins seul le deuxième type de question nous conduira à une réponse qui tienne compte du microniveau.

Si l'on est capable de répondre à la deuxième question, c'est à dire "En vertu de quoi le système S possède–t–il maintenant la propriété P? " alors la propriété P sera réductible dans le sens de Broad. Toutes les fois que la réponse à la question donnée répond à la description ci–dessous, elle prend en compte l'organisation du micro–niveau. Cette réponse est la suivante :

> [...] when we come to see that something having the kinds of components specified, organized in the way specified, is bound to have the target property. [Cummins, 1983, page 17]

Je pense que cette réponse est dans l'esprit de la théorie de Broad. La déductibilité d'une propriété se trouve dans ce dernier cas en rapport avec les propriétés de ses parties et de leurs modes d'organisation.

Selon Cummins on sait qu'un objet muni d'une certaine microstructure doit nécessairement posséder la propriété F si l'on sait que l'énoncé suivant est une loi :

> Si x a comme composés $C_1, \ldots, C_n$ organisés de la façon R, voire si x a la microstructure $[C_1, \ldots, C_n; R]$, alors x possède la propriété F.

Pour Beckermann cette loi ne satisfait pas encore la notion de réduction de Broad car ce dernier convient même pour les propriétés émergentes de l'existence de ce type de loi si elles s'avèrent "uniques et ultimes". Broad ferait un pas supplémentaire vers la réduction et il se poserait aussitôt la question de *pourquoi* certains types de structures déterminent certaines propriétés. Beckermann signale, a juste titre que ce qu'il faut pour répondre à cette question est une théorie générale des composantes (T_c) qui puisse, une fois fixée la propriété F, dire quels systèmes caractérisés comme $[C_1, \ldots, C_n; R]$ seront destinés à avoir cette propriété F. La façon de procéder consistera à prendre l'ensemble des lois dont F est issu et à regarder leurs images, c'est à dire les ensembles de paires $[C_1, \ldots, C_n; R]$.

Cependant, si l'on regarde de plus près, on voit que cette dernière démarche n'est pas si simple. Le problème est que la propriété F peut être réalisée par différents types de microstructure. Cela veut dire qu'il existera une théorie générale des composantes que j'appellerai $T_{\tilde{c}}$ qui vérifiera que s'il existe une microstructure $[\tilde{C}_1, \ldots, \tilde{C}_n; \tilde{R}]$ alors elle aura aussi la propriété F.

Le problème qui se pose est donc le suivant; le prédicat "avoir la propriété F" n'est pas co–extensif avec le prédicat "avoir la microstructure $[C_1, \ldots, C_n; R]$". Néanmoins on pourra soutenir qu'il existe un corrélation nomologique entre eux à l'intérieur du système de microstruture C. Or, la micro–réduction est une propriété relative au système.

Je crois que c'est le moment de donner les définitions de *micro–réduction* ou *d'explication* ou de *réalisation* d'une macropropriété F par la structure $[C_1, \ldots, C_n; R]$ dans le système S en citant Beckerman :

> Dans un système S donné, la propriété F sera microréductible à la microstructure $[C_1, \ldots, C_n; R]$ si et seulement si il existe une théorie générale T_c, à partir de laquelle on peut déduire une image relative à ce système pour chacune des lois caractéristiques de F. [Berckermann, 1992, page 115 ma traduction]

Je conviens que cette définition de la micro–réduction est plus proche de ce que Broad entendait lorsqu'il parlait de réduction parce que l'on tient compte des propriétes des composantes et de leurs modes d'organisation.

Avant d'exprimer quelques observations finales sur le sujet, il convient de présenter la définition de propriété émergente dans ce cadre.

> Let S be a system having the microstructure $[C_1, \ldots, C_n; R]$, then F is an emergent property of S if (a) there is a law to the effect that all systems with this microstructure have F, but (b) F is not microreducible to $[C_1, \ldots, C_n; R]$. [Berckermann, 1992, page 115]

[25] Un intéressant exemple de cette situation est discuté par [Berckermann, 1992, cf. page 110]

A partir de cette définition on voit que le concept d'émergence est le complément exact du concept de réduction [Berckermann, 1992, page 115]. Cette observation est importante parce qu'elle exprime l'objectif de la stratégie et semble, en principe, satisfaisante.

Néanmoins, je voudrais émettre quelques critiques et proposer un pas supplémentaire dans le raisonnement permettant d'y répondre.

Lorsqu'on regarde la seconde condition de la définition de propriété émergente, F est émergente parce qu'elle n'est pas microréductible. Le fait de ne pas être microréductible, toujours selon Beckermann, signifie qu'il n'existe aucune théorie générale des composantes T_c à partir de laquelle on puisse déduire une image relative à ce système pour chacune des lois qui ont F comme conséquente. Supposons maintenant qu'on trouve cette théorie; dès lors la propriété F ne sera plus émergente. Soit le concept d'émergence est défini sur des bases épistémologiques ou relatives à l'état de la science, soit il constitue une entrave au progrès scientifique et en définitive ce concept est là pour nous signaler les limites de la recherche scientifique.

Plus tard je présenterai le concept d'émergence de Mario Bunge qui prétend que ce concept se définit de manière relative avec une théorie. Des propriétés peuvent s'avérer émergentes en relation à une théorie et non émergentes vis-à-vis d'une autre selon Bunge. Il n'y a rien d'irrationnel, me semble-t-il à affirmer qu'une propriété est relative à une théorie. En revanche, Beckermann affirme qu'une propriété est émergente s'il n'existe *aucune* théorie pour expliquer la microréduction et cette condition qui exige l'absence de toute théorie est récusable pour les raisons que j'ai déjà exposées plus haut.

Je pense qu'il faut rejeter cette tentative d'opposer l'émergence à la réduction pour toutes les raisons que je viens d'exposer. En outre, le concept de microréduction de Beckermann mérite aussi une autre critique. En effet, la méthode de Beckermann pour obtenir la microréduction d'une propriété est de caractère extensif. Selon lui, le fait d'avoir une théorie générale T_c nous permettra de donner un compte rendu de la microréduction parce qu'elle nous permet de faire une liste des tous les antécédents des lois dont la propriété F est conséquente. Beckermann semble dire que pour rendre une propriété F microréductible il suffit de trouver une loi qui ait comme antécédente la disjonction des toutes ces microstructures et comme conséquente la propriété F en question.

Cette loi extensionnelle a deux défauts: le premier que je viens de signaler est sa condition disjonctive (voire chapitre 7), l'autre est le caractère douteux de sa capacité explicative.

Il me semble que l'on pourrait faire de la solution de Beckermann la même critique que Beckermann lui-même a adressé à Cummins. En effet, Beckermann ne répond plus à la question "en vertu de quelles propriétés ce substrat est-il destiné à posséder la propriété F? " car le caractère extensionnel de sa réponse semble plutôt répondre à la question "Quels sont les microstructures de ce substrat qui sont destinées à posséder la propriété F? .

En définitive, Beckermann a raison de signaler que Broad n'aurait pas accepté une réponse à la Cummins comme caractérisant une propriété non-émergente, mais il a tort de penser que sa solution soit acceptable aux yeux de Broad.

Pour trouver une solution acceptable à Broad, il nous faudrait définir un type d'abstraction qui se fonde uniquement sur les caractéristiques de l'organisation sans faire référence aux types de substrat ou aux différents éléments du même substrat. La question est de savoir si l'on pourra donner une caractérisation susceptible de décrire les propriétés de toutes les structures du type $[C_1, C_2, \ldots, C_n; R]$ qui soient indépendantes du substrat, c'est à dire des composantes C_i. Si l'on trouve une telle caractérisation, alors on aura une définition de réalisation transversale aux différents systèmes. Dans la conclusion je vais donner des pistes qui nous permettront d'espérer qu'une démarche comme celle que je viens de proposer aboutisse.

Si l'on récuse l'irréductibilité comme un trait du concept d'émergence, alors la dichotomie émergence-réduction, c'est-à-dire en fait l'opposition entre les concepts d'émergence et de réalisation disparaît. En fait dans ce cadre il s'avère que le concept d'émergence sera l'inverse[26] du concept de réalisation. Je vais soutenir cela dans la conclusion de la thèse.

[26] Je prends comme traduction du mot anglais *converse* le terme français *inverse* à ne pas confondre avec *contraire* qui veut dire *opposé*.

8.5 L'émergentisme et les propriétés relationnelles

Dans la section précédente, j'ai discuté le caractère d'irréductibilité que certains auteurs veulent octroyer au concept d'émergence en le récusant. Je vais m'attaquer dans cette section à une autre particularité que l'on veut garder de l'émergentisme britannique, le caractère non-déductif des propriétés émergentes à partir des autres caractéristiques du système. Je vais également réfuter ces démarches en montrant qu'elles aboutissent à une conception naïve de l'émergence. A partir de ces résultats, je démontrerai que certains usages que les chercheurs de la vie artificielle font du terme émergence peuvent aussi être considérés comme naïfs.

Le texte de Paul Teller[27] est une source intéressante de réflexion à ce sujet.

Les observations de Teller sont très utiles puisqu'elles ouvrent des pistes pour déceler les cas où le concept d'émergence se révèle trivial. Je vais m'appuyer sur ces observations pour prouver que certaines définitions de l'émergence données dans le cadre de la vie artificielle ne résistent pas à une telle analyse. En effet, on verra qu'il existe une certaine conception de l'émergence qui s'avère être une vague métaphore pour mesurer l'autonomie d'un agent artificiel.

Existe-il des propriétés non relationnelles?

Paul Teller fait une intéressante dissertation sur les types des propriétés émergentes d'une totalité en partant d'une notion intuitive du concept d'émergence [Teller, 1992] et de déduction.

Le concept de déduction est pris ici en un sens large ou intuitif, ce qui donne à cette discussion un cadre moins strict que celui de la section précédente. Il va de même pour la définition d'émergence.

DEFINITION 8 (ÉMERGENCE : ÉNONCÉ 1)
Il n'est pas possible de prédire, voire de réduire ou de définir explicitement une propriété émergente d'une totalité à partir des propriétés de ses parties.

Teller discute de la possibilité de définir des propriétés émergentes en fonction des propriétés relationnelles/non relationnelles [28] de leurs parties. Cette démarche a le mérite d'expliciter la définition de propriété émergente de façon telle qu'elle élimine les cas triviaux d'émergence. En ce faisant, il trouve un moyen de classer les propriétés des parties qui échappent à la trivialité. Pour y parvenir, Teller reformule la caractérisation de propriété émergente d'une manière qui semble à première vue équivalente à l'énoncé 1 :

DEFINITION 9 (ÉMERGENCE : ÉNONCÉ 2)
Une propriété d'une totalité est émergente si elle n'est pas réductible aux propriétés non relationnelles *des ses parties.*

Le terme *réduction* a ici le sens d'*explicitement définissable.*

Cette nouvelle énonciation du présupposé émergentiste me semble très utile pour analyser la possible trivialité d'une propriété relationnelle émergente.

L'exemple suivant proposé par Teller est très éloquent.

EXEMPLE 3
Soit une boîte à crayons; considérons la propriété "être le crayon le plus long de la boîte". [Teller, 1992, cf. page 141]

[27] [Teller, 1992]

[28] Rappelons-nous que les propriétés relationnelles sont celles qui peuvent être définies comme des relations d'ordre égal ou supérieur à 2; des exemples de propriétés relationnelles sont "être marié", "être le gagnant d'une course", "être un satellite". Les propriétés non relationnelles sont celles que peuvent être décrites comme des propositions d'ordre égal à 1, par exemple "être rouge", "être un électron". Dans les chapitres précédents j'ai aussi fait référence à ces dernières en tant que propriétés *intrinsèques*.

Selon l'énoncé 1 cette propriété relationnelle est émergente puisque le fait de vérifier qu'il satisfait à la propriété d'être le plus long ne peut pas être déduit à partir des seules propriétés du crayon pris de façon isolée.

Cependant, elle ne vérifie pas la condition d'émergence selon l'énoncé 2. En effet, le fait d'être le crayon le plus long de la boîte est définissable explicitement en fonction (ou réductible à) des propriétés non relationnelles des éléments, en l'occurrence la longueur des crayons. [29]

Teller, par la suite, cite un autre type de propriétés pour étudier la possibilité de les considérer comme émergentes. Il s'agit des propriétés de la totalité survenantes des propriétés non relationnelles des parties. Ces propriétés non relationnelles de la totalité sont telles que bien que non-réductibles aux propriétés non relationnelles des parties, elles s'avèrent survenantes de ces dernières. Teller donne comme exemple la propriété non-relationnelle (en principe) d'être une calculatrice ou une clé. [30]

Dans cette catégorie, j'ai cité la propriété d'être une calculatrice. Teller met en évidence que cette propriété est considérée d'emblée comme non-relationnelle de la totalité. Le fait qu'elle ne soit pas réductible aux propriétés non relationnelles de ses parties est évidente parce que cette machine peut être réalisée avec divers matériaux et même sous différentes formes, telles celle de l'abaque par exemple. Cependant, lorsque l'on copie une machine à calculer atome par atome le résultat est aussi une machine à calculer. [31] Or, la propriété d'être une calculatrice est survenante des propriétés non-relationnelles de ses parties.

Teller par la suite traite de la caractérisation non-relationnelle que l'on donne à la dite propriété. Selon lui, l'ensemble des possibilités de machines méritant la définition de calculatrice est si vaste, par exemple tout objet ayant le même nombre de degrés de liberté qu'un abaque comme un ensemble de couteaux et de fourchettes, les feuilles de différents arbres, etc. selon le contexte social où l'on se trouve. Ainsi la propriété d'être une calculatrice pourrait être considérée, à juste titre comme relationnelle à l'environnement social dans lequel on se place. En réalité, maintes propriétés que l'on considère non relationnelles sont en fait des propriétés relationnelles implicites, d'où le fait qu'elles s'avèrent non-réductibles mais seulement survenantes des propriétés non relationnelles des parties, dans ce groupe.

Pour préciser cette idée prenons une propriété relationnelle comme "être le frére de". Nous pouvons la représenter comme une fonction propositionnelle F à deux variables x et y (c'est-à-dire $F(x,y)$). La propriété relationnelle n'est pas survenante des propriétés relationnelles des parties parce que les copies, molécule par molécule, des deux frères ne seront pas liées entre elles par une relation de fraternité. En effet, les deux copies n'ont pas les mêmes parents et en fait n'ont pas de parent de tout. C'est seulement à ce moment que l'on se rend compte que l'on a simplifié l'expression de la relation "être le frère de" puisqu'en fait, il s'agit non d'une fonction propositionnelle avec deux variable libres mais bien d'une fonction propositionnelle à quatre variables dont deux sont bornées c'est-à-dire, $\exists p, m : F(x, y, p, m)$ où m et p représentent le père et la mère respectivement de x et y.

Or, certaines propriétés que l'on considère comme non relationnelles ne sont que des propriétés relationnelles que l'on a simplifiées. Il faudrait tenir compte de cette considération puisque je vais revenir sur cette constatation lors de la discussion du concept d'émergence dans la vie artificielle.

Teller inclut dans ce groupe toutes les propriétés fonctionnelles. Par exemple, être une calculatrice ou être une clé sont des propriétés fonctionnelles.

Je suis en accord avec l'analyse de Teller jusqu'ici. Cependant, je ne puis accepter que les considérations suivantes servent à établir une division nette entre les propriétés émergentes et celles que ne le sont pas. Les raisons qui me poussent à récuser l'énoncé qui suit ne sont pas différentes de celles de Teller qui avoue ne pas en être convaincu lui non plus.

[29] En effet, il suffit d'établir une relation d'ordre d'une propriété non relationnelle des crayons, en l'occurrence leur longueur, pour opérer la réduction.

[30] Teller mentionne un troisième type de propriétés: les propriétés émergentes non-relationnelles de la totalité qui ne sont même pas survenantes des propriétés non relationnelles des parties mais je n'en parlerai pas parce que je ne les considère pas pertinentes à ma discussion.

[31] Notez que toutes les propriétés ne surviennent pas des propriétés non relationnelles des éléments de la relation. Par exemple, les copies molécule pour molécule de deux frères ne sont pas pour autant "frères".

DEFINITION 10 (ÉMERGENCE : ÉNONCÉ 3)
Une propriété relationnelle est une propriété émergente seulement si elle est une propriété fonctionnelle ou si la relation en question ne peut pas être définie en termes des propriétés non relationnelles de leurs relata.[Teller, 1992, cf. page 146]

Une des raisons qui pousse Teller à récuser cet énoncé est que le concept de fonction n'a pas encore le poids analytique nécessaire pour être utilisé comme facteur d'individualisation des propriétés fonctionnelles. En effet, il soutient :

> [...] for ruling out relations which can be defined in terms of non-relational properties of the relata, I don't know that the relevant relations underlying being an adder and be a key can't, in principle, be so defined. This problem is that these relations are hopelessly complex, involving understanding of social practice which we have not yet begun to master. [Teller, 1992, page 146]

Les raisons qui me poussent à rejeter l'énoncé 3 sont d'un autre ordre.

A mon avis le problème fondamental tient à la compression tacite que l'on fait du concept "totalité-partie". En effet, premièrement la relation entre une totalité et ses parties est considérée comme une relation d'inclusion et deuxièmement la définition de l'ensemble des parties n'a pas de critère fixe car dans le meilleur cas il s'agit des unités fonctionnelles; en troisième lieu, le postulat émergentiste de l'existence des niveaux est totalement ignoré. Face à ces points qui restent flous parce qu'ils sont considérés comme implicitement définis, la caractérisation d'une propriété comme émergente se trouve être non une classification cahotante des propriétés mais simplement un problème de degré. Teller même reconnaissait ce dernier fait :

> For the moment, at least, I suggest that we recognize that relational properties lie on a continuum of complexity, that the correlative notion of emergence is a matter of degree, and that we count as interestingly emergent only those relational properties which are both non-trivial and are involved in subject-matters which we find important. [Teller, 1992, page 146]

Tel que Teller présente le concept d'émergence, on peut se poser la question s'il vaut la peine d'en tenir compte dans la caractérisation. Je pense que la réponse est négative. D'un côté, pour les propriétés fonctionnelles il s'agit d'un problème de degré, de l'autre ce concept dans une théorie risque, sinon de devenir totalement subjectif du moins de faire office d'un sac où l'on déverse toutes les propriétés fonctionnelles que l'on ne peut pas définir.

Je pense que l'utilité du concept d'émergence dans une théorie peut encore être préservée mais il faut être moins libéral dans sa définition. Autrement dit, il faut renoncer à la caractérisation d'une propriété émergente sur la simple base des postulats intuitifs de l'émergentisme comme Teller s'en est rendu coupable.

Je vais présenter plus bas une notion d'émergence qui me semble la plus intéressante et précise, celle proposée par Mario Bunge. La démarche de Teller a le mérite d'attirer notre attention sur le fait que les propriétés non relationnelles et fonctionoelles sont souvent le résultat d'une simplification des propositions qui les expriment par l'omission des variables liées que l'on considère implicites.

Pour des propriétés relationnelles comme celle d'*être frère de*, les variables liées qui sont omises ne sont que deux et sont donc faciles à expliciter. Dans le cas de la calculatrice, en revanche l'explicitation devient pratiquement impossible, tout simplement pacce que la complexité de la situation et le nombre de variables sont énormes compte tenu de notre capacité de compréhension.

Teller s'interroge sur la possibilité de trouver d'autres propriétés que l'on puisse considérer émergentes et survenantes des propriétés non relationnelles des parties dans les sciences de la vie, par exemple la propriété d'être vivant ou d'autres propriétés plus spécifiques.

Il cite la propriété de *fitness* en affirmant qu'elle pourrait être considérée comme une contrepartie de la propriété d'être une calculatrice dans le domaine des science de la vie.[32] La propriété de *fitness* est, comme chacun sait, la capacité d'un organisme à s'adapter à son environnement. La faculté de s'accommoder à un environnement est le fait d'un système complexe spécialisé dans les relations avec le milieu. Par exemple, pour continuer à être vivant un organisme doit se nourrir mais en mangeant de la nourriture qui lui soit appropriée et ceci illustre la complexité de la relation.

[32][Teller, 1992, cf. page 148]

Au lieu de dire que la propriété de *fitness* est une propriété non-relationnelle qui survient des propriétés des parties de l'organisme, ne faudrait-il pas plutôt conclure que c'est une propriété relationnelle dont on ne donne qu'une partie des variables liées exprimant des conditions environnementales dites implicites parce que l'on n'est pas en mesure de les expliciter étant donné leur complexité?

Je pense que c'est peut-être une intuition de ce type qui poussait les premiers émergentistes à invoquer l'émergence dans le contexte des tentatives d'expliquer la vie. Il est aussi possible que ces même intuitions incitent de nos jours la communauté scientifique de la vie artificielle à parler d'émergence.

8.5.1 L'émergence dans la vie artificielle

La vie artificielle (*VA*) veut se différencier foncièrement des disciplines de l'IA. Un des éléments de cette différentiation est le rôle que l'on fait jouer à l'environnement.

L'*IA* traditionnelle privilégie les relations entre les différents éléments fonctionnels du système en figeant les relations avec l'environnement par des descriptions préalables des changements dans les structures des données suivant des régularités fixées au départ. Les chercheurs de la vie artificielle récusent cette démarche car des caractéristiques très importantes de la relation avec l'environnement y sont laissées de côté.

La richesse de l'environnement est tellement grande (même lorsque l'environnement peut être considéré comme simple) que l'on ne peut pas en faire l'inventaire à l'avance. Il faudrait donc doter le système de la faculté de réagir à des conditions environnementales diverses. Pour y parvenir la stratégie des chercheurs de la vie artificielle dote le système qui évolue dans un environnement donné de comportements simples et de la capacité d'être coordonnés entre eux. En ce faisant, le comportement observé du système se révèle doté de capacités d'adaptation d'une souplesse remarquable.

Un autre trait qui différencie l'*IA* traditionnelle de la vie artificielle est le domaine d'expérimentation. Dans le premier cas, on travaille dans des environnements virtuels avec des agents virtuels, alors que pour la vie artificielle on utilise des agents réels comme des robots nantis de capteurs et évoluant dans un environnement réel bien que dans la plupart des cas il s'agisse d'environnements simplifiés dans le sens qu'ils sont plus ou moins statiques.

Selon les chercheurs de la *VA* ces deux modèles constituent des théories des phénomènes naturels, mais les modèles de l'*IA* sont des *simulations* tandis que ceux de la VA sont des *réalisations*.[33]

Or, la conception de la *VA* en tant que réalisation a été contestée par Elliot Sober dans [Sober, 1991]. Je reviendrai plus tard sur ces critiques que je partage.

D'abord quelques observations à la caractérisation que je viens d'exposer entre *IA* et *VA*. Il ne me semble pas justifié de dire que les modèles d'*IA* sont des simulations tandis que ceux de la *VA* sont des réalisations. A l'origine l'*IA* tentait de modéliser des processus dits intelligents, en particulier la solution générale de problèmes; le domaine d'application était la psychologie et il y

[33] Selon la caractérisation de H. Patee cf. [Pattee, 1989]
En effet, Luc Steels affirme

> Computational models and artificial models, or what [Pattee, 1988] calls simulations and realisations, must be clearly distinguished. For exemple, it is possible to build a computational model of how a bird flies, which amounts to a simulation of the environment around the bird, a simulation of the aerodynamics of the body and the wings, a simulation of the pressure differences caused by mouvement of the wings, etc. Such a model is highly valuable but would, however, not able fly. It is forever locked in the data structures and algorithms implemented on the computer. It flies only in a virtual world. In contrast, one can make an artifact in terms of physical components (a physical body, wings, etc.). Such an artifact would only be viewed as satifactory if it is able to perform real flying. This is a much stronger requirement. Very often, results from simulation only partially carry over to artificial systems. When constructing a simulation, one selects certain aspects of the real world that are carried over into the virtual world. But this selection may ignore or overlook essential characteristics that play a role unknown to the researcher. An artificial system cannot escape the confrontation with the full and infinite complexity of the real world and is, therefore, much more difficult to construct. [Steels, 1995, page 76]

avait alors une forte tendance à caractériser ces processus comme purement symboliques. La *VA*, en revanche a un autre domaine d'application : la biologie et la définition des processus intelligents en est fortement influencée; ainsi on peut trouver chez Steels la définition suivante d'un système intelligent :

> **The behavior of a system is intelligent to the extent that it maximizes the chances for self-preservation of that system in a particular environment. [Steels, 1995, page 77]**

Il s'agit tout simplement d'un changement de l'objet auquel les modèles de chacune de ces théories se réfèrent. Dans le camp de l'*IA*, spécialement en ce qui concerne la théorie corps-esprit comme fonctionnalisme représentationnel non-réductionniste on est moniste de substance et dualiste de propriétés. Les propriétés mentales, bien que réalisées dans un substrat physique, vont au delà de ce substrat, ce qui leur octroie une indépendance ontologique due à des capacités du moins descriptives qui réfutent le réductionnisme extrême.

Mutandis mutandis telle est aussi la démarche qu'a suivie la *VA* mais en relation à la propriété de *fitness*; cette propriété n'est pas une propriété physique car elle transcende le domaine physique et mérite un statut ontologique à part entière. En plus cette propriété dont tous les êtres vivants sont dotés se manifeste ou se réalise de différentes façons pour les divers types d'organismes. Sober signale justement qu'il est très difficile de déterminer ce qu'ont en commun un cafard et un zèbre pourtant tous deux sont bien adaptés à leur milieu. Il sera en effet difficile voire impossible de trouver des propriétés physiques communes. Il en va néanmoins de même pour les propriétes fonctionnelles car en effet, qu'y-a-t'il de commun entre une trappe à souris et un chat? En conclusion, il me semble que la supposition que l'*IA* est inférieure dans sa modélisation, parce que c'est une simulation, à la *VA* qui est une réalisation n'est pas justifiée. C'est seulement le résultat d'un amagalme des deux domaines distincts : le domaine cognitif et le domaine biologique. Une autre raison de cette appréciation tient à l'intuition que l'*IA* traditionnelle a finalement échoué dans sa tentative de parvenir à une vraie réalisation des activités intelligentes et que la *VA* qui vient de naître a encore un long chemin à parcourir avant de pouvoir répondre aux espoirs qu'elle a éveillés.

Néanmoins, il ne peut être exclu qu'à l'instar des modélisations en *IA* qui n'ont pu expliquer l'intentionnalité en termes de représentations internes et de relations entre les états fonctionnels, la *VA* fasse fausse route dans sa démarche de modéliser les propriétés biologiques en termes du couplage environnement-agent.

La VA est-elle véritablement une réalisation des propriétés biologiques?

L'idée maîtresse selon laquelle la *VA* est une réalisation des propriétés biologiques se fonde sur le fait que les agents évoluent en général dans un environnement réel. La question est de savoir si cette caractéristique suffit pour affirmer qu'il s'agit bel et bien d'un réalisation.

Je pense que le terme réalisation est excessif. Les simplifications dans la *VA* s'opèrent à deux niveaux différents et le premier effet de la simplification ressort de la complexité généralement très limitée de l'environnement où les agents artificiels évoluent. En général, ces environnements sont les locaux d'un laboratoire d'informatique. Les autres facteurs de simplification sont les comportements modélisés, schémas sommaires parfois à peine comparables aux processus biologiques même les plus simples.

On demande au robot de suivre un mur ou de se tenir au parcours d'une trajectoire définie par une caméra, de poursuivre une cible, ou bien, s'il s'agit d'un programme plus sophistiqué de trouver l'accumulateur le plus proche pour recharger ses batteries. Je reconnais que ces comportements sont de bonnes modélisations ou simulations de processus biologiques comme par exemple l'alimentation de l'agent pour ce qui est de la recherche des batteries; en revanche il me semble excessif de leur attribuer le caractère d'une réalisation. Les traits saillants des relations entre les agents naturels et l'environnement ne sont pas identiques à ceux qui sont propres aux rapports entre les agents artificiels et l'environnement.

Le fait que le résultat extérieur soit le même ne veut pas dire qu'il s'agisse d'une vraie réalisation, car c'est tout simplement un effet de la méprise qu'a illustrée John Searle par l'exemple de

la chambre chinoise mais transposé au cas de la *VA*.

> Many biological properties and processes involve relationships between an organism (or a part of an organism) and something outside itself. An organism reproduces when it makes a baby. A plant photosynthesizes when it is related to a light source in an appropriate way. A predator eats other organisms. Although a computer might replicate aspects of such processes that occur inside the system of interest, computers will not *actually* reproduce or photosynthesize or eat unless they are related to things outside themselves *in the right way*.[Sober, 1991, page 759, souligné par moi]

Comment peut–on savoir que les relations avec l'environnement telles qu'on les a modélisées reproduisent les traits fondamentaux des comportements en question de façon à nous autoriser à dire qu'elles constituent une réalisation? L'unique vérification possible consiste en l'observation de la performance mais ceci ne me semble pas suffisant pour affirmer que l'agent fait preuve du comportement de se nourrir lorsqu'il se dirige vers le lieu de recharge des batteries.

Elliot Sober signale que le danger est plus grand lorsqu'on dispose d'une structure matbématique pour le processus biologique en question.

A ce moment–là, on a la fâcheuse tendance à confondre le domaine empirique d'application de cette structure mathématique avec la structure elle–même. Il donne comme exemple la loi des populations génétiques de Hardy–Weinberg. Cette loi dit que

> [...] what frequencies the diploid [34] genotypes at a locus will exhibit, when there is a random mating, equal numbers of males and females, and no selection or mutation. It is, so to speak, a "zero–force law"– it describes what happens in a population if no evolutionary forces are at work. [Sober, 1991, page 759]

Si p est la fréquence de gènes A et q de a (où $p + q = 1$) alors, dans les circonstances citées ci–dessus, les fréquence des paires AA, Aa et aa seront respectivement p^2, $2pq$ et q^2.

A ce point Sober nous propose de considérer deux illustrations différentes de ces lois. La première concerne une usine de chaussures qui produit des chaussures marrons et noires. Il y a eu un problème dans la chaîne d'assemblage qui a conduit à la dissociation des paires de chaussures. Le résultat est qu'il y a deux piles, l'une composée de chaussures noires et marrons mais toutes du pied gauche et la seconde des chaussures des deux teintes du pied droit.

Le patron de la chaîne envisage de trier les paires en utilisant une machine alétoire. Si la fréquence des chaussures noires est p et celle des autres est q, la fréquence que l'on atteindra pour les trois paires de chaussures résultantes possibles sera p^2, pq et q^2. Ainsi la machine obéira à la même loi qu'une population génétique de Hardy–Weinberg.

En outre, une population de *Drosophila* qui est la mouche d'un fruit, donc un objet biologique et vivant se reproduit selon la loi de Hardy–Weinberg. Cela veut dire que cette population a la même structure mathématique que la "population" des chaussures.

Néanmoins, il y a une différence fondamentale, c'est que la machine à trier les chaussures n'est pas vivante et n'est donc pas un objet biologique même si elle se conforme à une loi applicable aux objets biologiques. Sober introduit ce qu'il appelle la méprise chaussure–mouche consistant en l'argumentation suivante: Les mouches sont des êtres vivants.

La loi L décrit les mouches.
La loi L décrit les chaussures.
Alors, les chaussures sont des êtres vivants.

En dernière analyse, je suis totalement d'accord avec la conclusion proposée par Sober.

> Functionalist theories abstract away from physical details. They go too far– confusing mathematical form with biological (or psychological) subjet matter – when they commit the Shoe–Fly Fallacy. The result is an overly simple?
>
> The idea of the Shoe–Fly Fallacy is a useful corrective against claims that a particular artificial system is alive or exhibits some range of biological characteristics. If one is tempted to make such claims, one should try to describe a system that has the relevant formal characteristics but is clearly not alive [Sober, 1991, page 760].

[34] La définition est la suivante: *A diploid is "an organism or a cell having double the basic or haploid number of chromosomes.* Selon le Webster's Encyclopaedic Unabridged Dictionary (1989)

En conclusion, l'idée que la vie est un processus physique mais infiniment complexe pour notre actuelle connaissance des choses semble une des pensées-maîtresses de la *VA*. Si tel est le cas, alors il se pourrait qu'un processus computationel se révèle vivant à l'avenir. [35]

Le problème est qu'il est très difficile de trouver des critères objectifs (comme par exemple le test de Turing pour l'intelligence) pour déterminer jusqu'à quel point un système mérite la qualification de vivant. C'est pourquoi on a songé au concept d'émergence qui pourrait permettre de qualifier les relations écologiques entre l'agent et l'environnement. Je vais donc traiter des différentes conceptions de l'émergence dans la VA.

8.5.2 L'émergence et la VA

L'émergence dans la vie artificielle sert de cadre à deux intuitions très fortes concernant la relation agent–environnement.

Une de ces intuitions est purement théorique; elle se réfère à la possibilité de bâtir un système que puisse surpasser les spécifications selon lesquelles il a été construit.

La seconde intuition a surtout des conséquences que je qualifierais de techniques car elle joue un rôle fondamental dans l'architecture des systèmes. J'en ai déjà parlé plus haut lorsque j'ai mentionné la conception des relations entre agent et environnement. Ces relations s'avèrent tellement étroites qu'il est impossible de définir des comportements complexes qui prennent en compte tous les changements possibles dans le milieu contrairement à ce qu'assume, à tort, l'approche cognitiviste. Pour surmonter cette difficulté on a doté les systèmes artificiels de comportements simples mais on leur a donné la possibilité de les combiner entre eux. De plus, pour pouvoir définir des comportements de plus en plus complexes on a muni ces systèmes de capacités d'apprentissage. [36]

Néanmoins, pour parvenir à ce résultat on a été obligé de faire encore un autre changement. On a du prendre, pour ainsi dire, le point de vue du robot. Cela veut dire que le concepteur a décidé de catégoriser différents signaux physiques des capteurs et a aussi décidé lesquelles de ces partitions sont à l'origine des changements des comportements.

Ainsi, cette caractérisation des stimuli permet d'en user comme éléments dans les boucles que l'on appelle des comportements de base. Ces comportements sont dits "innés" dans le jargon de la vie artificielle par un abus de langage très fréquent dans cette communauté.[37]. C'est l'effet de cette opération de choix technique qui est désigné par des chercheurs de la VA comme "le point de vue du robot" ou, ce qui est aussi un abus de langage évident, comme *méthodologie non-objectiviste*. Finalement, j'aimerais citer en exemple de cette approche la réponse que donne Miguel Rodriguez à la question "Sur quoi la connaissance de l'agent cognitif porte-t-elle? "

> L'agent évolue dans un environnement extérieur à lui. Il n'y a aucun accès, si ce n'est à travers ses *sens* et ses *actions*.
>
> Les sens lui fournissent continuellement une image de son environnement. C'est à partir de là qu'il construit sa réalité subjective. L'action lui permet d'intervenir dans l'environnement et de modifier ainsi l'image reçue. C'est par l'action et la perception que l'agent *dialogue* avec l'environnement. Ce dialogue forme l'interaction milieu-agent ....
>
> Encore une fois, au lieu de vouloir représenter de façon "adéquate" un "monde" extérieur, objectif et prédéterminé, notre approche est plus empirique; l'agent prend comme référent, non pas son partenaire de dialogue, mais le dialogue lui-même, l'interaction. Elle est, à notre sens, l'ultime frontière que l'agent puisse traiter, la dernière qu'il puisse maîtriser. C'est donc sur l'*expression des régularités de l'interaction* que l'agent va fonder sa connaissance de base autour de laquelle s'articuleront les processus structurateurs (processus cognitifs).[Rodriguez, 1994, page 117, les italiques appartiennent au texte original]

Si l'on vide la citation précédente de tout vocabulaire anthropomorphique et qu'on traduit cette métaphore utile pour éviter de longues périphrases, on traduira la phrase : "Les sens lui fournissent

[35] Pour une discussion des postulats de la vie artificielle maintenant que le processus de la vie est purement physique voir [Rasmussen, 1991]

[36] Je prend le terme *cognitiviste* dans le strict sens technique comme il est usuel pour certains chercheurs de la *VA* dont Miguel Rodriguez [Rodriguez, 1994, cf. page 88]. Ce terme, selon cette école définit l'opposé du terme *constructiviste* que Rodriguez explique par la métaphore suivante: "l'agent devient l'auteur de ... [la] connaissance [des relations avec l'environnement]".

[37] Par exemple dans une modélisation donnée voir [Rodriguez, 1994, page 108]

continuellement une image de son environnement" par "on a prévu dans notre système une lecture aussi continue que possible des signaux des capteurs et on a prévu, à partir des interprétations que nous, les concepteurs, avons fourni comme référence de base un changement dans le comportement aussi rapide que possible". Dans cet exemple on voit d'un côté l'utilité d'un langage métaphorique, de l'autre ce que l'on voulait illustrer en tant qu'approche non-cognitiviste. Cependant, il ne faudrait pas prendre la métaphore pour la réalité et c'est le risque que l'on court lorsqu'on tient certains discours émergentistes en *VA*.

Les discours émergentistes évoqués dans la *VA*

L'émergence se prête à trois conceptions différentes dans le cadre de la vie artificielle.

L'émergence computationnelle: L'idée fondamentale de ce groupe de théories est que la complexité des formes globales peut être causée par des interactions computationnelles locales. Il s'agit d'une démarche *bottom-up* qui est compatible avec les modèles connectionnistes. Paul Smolensky a formulé une intéressante théorie sur l'émergence des structures universelles du langage à partir des modèles connectionnistes à l'aide du concept mathématique des tenseurs.[38]

Dans le domaine de la *VA* proprement dite ces modèles sont utilisés par Langton et Toffoli[39]. Ces programmes de recherche ont pour but la réalisation de comportements émergents et la simulation des processus évolutifs par l'emploi d'automates cellulaires.

Cette caractérisation de l'émergence a été critiquée par le fait que les comportements émergents que l'on obtient au macro-niveau sont déterminés par le micro-niveau. Par exemple Peter Cariani dit :

> In assuming rule-governed, bottom-up organization rather than semi-autonomous levels of organization, computational emergence tacitly incorporates the older reductionist assumption that micro-orders determine macro-orders but not vice versa. [Cariani, 1991, page 776]

On voit bien la difficulté qu'ont certains chercheurs à faire la différence entre la réduction épistémologique et la réduction ontologique. Connaître les lois qui gouvernent le phénomène émergeant ne change en rien la qualité d'émergence de ce dernier. L'idée ancienne et fausse de l'émergence n'est point le concept de réduction épistémologique mais bien le concept d'imprévisibilité. Nous avons dit qu'il n'est ni utile ni pertinent de fonder un concept sur des bases épistémologiques.

En définitive, la théorie computationnelle de l'émergence a la vertu de proposer une conception rationnelle de l'émergence bien qu'elle ne réussisse pas à donner une définition nette des processus émergents les distinguant de ceux que ne le sont pas. En outre, elle repose sur de vrais modèles scientifiques qui démontrent des processus émergents dont on peut comprendre les causes grâce à la modélisation qu'ils simulent. De ce point de vue ce sont de bons outils de recherche.

L'émergence basée sur les modèles morphodynamiques: Ces théories forment une partie d'un groupe plus vaste dit des théories thermodynamiques de l'émergence. Le but de ces théories est de mettre en évidence que la stabilité et la complexité des structures peuvent survenir même quand on est loin d'une situation d'équilibre. Elles essaient d'expliquer comment de nouvelles structures émergent à la faveur des fluctuations du système.

Selon ce type d'explication le concept d'*attracteur*[40] joue une rôle central. Cette abstraction mathématique permet d'expliquer les macro-structures qui ont un caractère discret à partir du continuum des micro-processus.

Les théoriciens dans ce domaine sont I. Prigogine et J. Petitot entre autres.[41]

L'application de ces théories aux phénomènes de la vie requiert d'importants efforts. Selon Jean Petitot le recours aux modèles morphodynamiques pour expliquer ce type de phénomènes est une affaire de générations qui ne disposaient pas encore d'un corpus cohérent et compact.

[38] Pour un description détaillée voir [Smolensky, 1994].
[39] cf. [Langton, 1986] et [Toffoli, 1982]
[40] Je définis le concept d'*attracteur* dans le cadre de la théorie morphodynamique dans le prochain chapitre
[41] cf. [Prigogine, 1980], [Petitot, 1992], [Nicolis and Prigogine, 1985].

L'émergence relative à un modèle: On peut trouver dans la littérature technique différentes définitions de comportement émergent. Dans tous les cas les chercheurs se proposent de définir deux points de vue différents: le point de vue du système et le point de vue de l'observateur extérieur.

Dans la plupart des textes que j'ai suivis, l'observateur apparaît comme extérieur à la programmation de l'agent et le comportement est défini émergent selon la théorie ou le modèle que ledit observateur se forge de l'agent en question.

Cet observateur fera une liste des comportements en étudiant les régularités de ces derniers. Il s'agit d'une démarche sémiotique–interprétative [Cariani, 1991, voir page 784] qui vise à mesurer le pouvoir d'adaptation du système mais qui n'est pas émergeote au sens strict.

D'autres approches du même type mais basées par exemple sur le vocabulaire nécessaire à la description du système sont proposées dans [Steels, 1995]. L'observateur dresse une liste de catégories de comportements. Ainsi, un comportement sera émergent si l'observateur a besoin d'un nouveau terme pour le définir [Steels, 1995, page 80].

Le concept d'émergence est présenté comme un propriété graduelle par certains auteurs comme par exemple [Assad and Packard, 1991]. Cela veut dire que les critères d'individualisation des entités émergentes n'existent pas mais que les auteurs les définissent pour signaler qu'il y a des entités plus émergentes que d'autres.

Néanmoins, le concept d'émergence défini dans ce cadre est davantage un outil pour mesurer l'ampleur d'une métaphore qu'un réceptacle empli d'un contenu philosophique ou instrumental quelconque. Même si l'on veut assimiler le concept d'émergence à une intuition ou à un concept très libéral, la définition n'y gagne aucune pertinence métaphysique ou empirique. Il existe toujours une interprétation naïve de la définition selon laquelle les comportements s'avèrent émergents.

En effet, si l'on veut appliquer la définition de Teller basée elle aussi sur un postulat très libéral d'émergence, le comportement ne sera pas émergent puisqu'il pourra être réduit à des propriétés des sous–systèmes par son concepteur. Peter Cariani qui est l'auteur de la théorie de l'émergence relative à un modèle le reconnaît lui–même.

> The emergence–relative–to–a–model view has deep implications for the interpretation of artificial life simulations. All computer simulations can be described in terms of finite–state automata, as networks of computational state transitions, as formal symbol manipulation systems. As observer–programmers we can always find a frame which will make our simulation appear non–emergent. If we choose our observables to coincide with the stable computational states of the finite state automaton being implemented by the simulation, then we will always see it as a nonemergent state–determined system. Here the state variables of the simulation can take on and the state transition will correspond to all the simulation rules which govern the values of the states variables. Every time the simulation is run with the same initial conditions, the simulations will transit through the same trajectory of variable values. The computer simulation will be completely replicable; there will be no deviation of the simulation's behavior from the model of possible trajectories built up by the observer. Thus, from this perspective many of the breakout strategies that have been proposed to make artificial life simulation "open–ended" and "emergent" will simply not improve the situation because they do not change the formal, completely replicable nature of the process. Increasing the size of the simulation, adding new layers of simulation rules, simulating random or chaotic processes, or representing genotypes and phenotypes will not in any way change the replicability of the simulation; hence these changes will be ineffective at transforming a previously nonemergent simulation into an emergent one. ... *The interesting emergent events that involve artificial life simulations reside not in the simulations themselves, but in the ways that they change the way we think and interact with the world.* Rather than emergent devices in their own right, these computer simulation are catalysts for emergent processes in our minds; they help us create news ways of seeing the world. [Cariani, 1991, page 790]

Il n'est donc pas étonnant que la définition du comportement en VA soit une définition complétement behavioriste car il s'agit d'une définition qui prend en compte le point de vue de l'observateur extérieur au système. Ainsi, on arrive au paradoxe suivant, on accepte la VA comme une possible modélisation des processus biologiques, on construit un système qui remplit des (ou quelques?) conditions d'adaptation à l'environnement requis et ensuite on considère ce même système comme s'il s'agissait d'une boîte noire dont on ne sait rien en se disant qu'il s'agit d'un modèle de vivant.

La valeur de la VA comme simulation du processus biologique

Le fait qu'un modèle a une valeur, une pertinence quelconque à l'étude des phénomènes biologiques est directement en rapport avec sa genèse. En effet, si le modèle a été créé dans un cadre interdisciplinaire selon le protocole strict donné par le spécialiste en biologie, alors ce système aura valeur d'expérimentation. Le changement de point de vue du concepteur à l'observateur a une valeur, peut–être pour la recherche en biologie.

Si le système est construit en vue de la performance finale, même s'il est inspiré par des concepts plus ou moins proches empruntés à la biologie ou à la sociologie, il est difficile de croire que ce système puisse avoir des retombées quelconques sur les conclusions des sciences naturelles ou sociales.

Le concept que l'émergence est relative à un modèle devient purement technique, une mesure acceptée par la communauté de la VA pour établir le niveau de fitness ou d'autonomie d'un système. La prétention de découvrir dans ces observations des implications plus ou moins directes pour les sciences de la vie est irrecevable. En plus, cette prétention est malheureuse car la VA veut être considérée comme un science nouvelle et pour obtenir ses lettres de noblesse dans la communauté scientifique en général il est important qu'elle ne confonde pas la valeur technique indéniable de sa démarche avec sa capacité plus discutée à servir d'outil pour l'expérimentation en science de la nature. Les soupçons que suscite cette dernière considération sont différents. Le premier de ces soupçons tient au doute que peut éveiller le postulat que la vie puisse se réduire à des phénomènes purement physiques. La seconde appréhension est qu'une conception behavioriste des comportements dit de *fitness*, aussi performante soit–elle pour un système artificiel, ne suffit pas à assurer que ce système puisse être qualifié de vivant.

8.6 La théorie de l'émergence de Mario Bunge

Le mérite de la théorie de Bunge est d'éclaircir certains concepts qui sont centraux pour la définition de l'émergence: le concept de niveaux et les relations entre totalité et partie.

L'intérêt de Bunge pour le concept d'émergence dans le cadre de son travail philosophique est double. D'un côté ce concept est à la base de sa théorie sur le problème corps–esprit, de l'autre il aide à expliquer sa théorie de l'évolution. Il écrit d'ailleurs lorsqu'il se réfère aux concepts d'émergence:

> In fact the whole thing is a sort of generalization of the theory of evolution. [Bunge, 1977, page 504]

Pour l'exprimer de manière abrégée, au risque de faire une simplification qui je pense s'estompera lors de l'exposé qui suit, pour Bunge une propriété est émergente si elle appartient à la totalité sans appartenir aux composantes [Bunge, 1977, cf. page 501].

Une propriété[42] émergente appartient à une totalité complexe. Il faut comprendre ici le terme totalité en tant que système comme un ensemble de composantes qui sont reliées ou couplées les unes avec les autres [Bunge, 1977, cf. page 502].[43]

[42] Bunge donne une définition de la propriété basée sur des concepts mathématiques. Il affirme qu'on la représente comme une fonction mathématique en principe à valeurs réelles. Les n fonctions représentant des propriétés d'une chose donnée peuvent être réunies dans une fonction unique comme suit:

DEFINITION 11
Let each of the n properties of a concrete thing be represented by real value fonction F_i of time, with $1 \leq i \leq n$. Then

1. $\mathbf{F} = \langle F_1, F_2, \ldots, F_n \rangle : T \to \mathbf{F}^n$ *is called the* state function *of the given thing;*
2. *the value $s = \mathbf{F}(t)$ of $\mathbf{F}$ at the time t is called the* state *of the given thing at t;*
3. *the ordered pair $\langle s, s' \rangle$ of values of $\mathbf{F}$ at times t and t' respectivement is called an* event *occurring in the thing concerned between t and t';*
4. *the sequence of states joining two states s and s' of a given thing is called the* process *leading from s to s', or the* history *of the thing between t and t'.*

[Bunge, 1977, page 501–502]

[43] Pour un exposé plus précise du concept de système chez Bunge voir [Bunge, 1979, chapitre 6]

Plus précisément la définition du système de Bunge est la suivante:

DEFINITION 12 (SYSTÈME)
Un système est un triplet ordonné $\langle C(x), S(x), E(x)\rangle$, où $C(x)$ est la composition du système; les parties qui le composent doivent être au moins deux pour qu'on puisse qualifier l'entité en question de système, $S(x)$ est la structure et $E(x)$ est l'environnement du système (p. ex. les parties du monde extérieur avec lesquelles une partie du système entretient des rapports.[Bunge, 1979, page 5]

On appelle les propriétés qui appartiennent à une chose complexe *propriétés globales*. Les propriétés globales peuvent être *résultantes* ou *émergentes*.

DEFINITION 13 (PROPRIÉTÉ GLOBALE ÉMERGENTE OU RÉSULTANTE)
Soit P une propriété d'une chose complexe x différente de la propriété "être une composante de x"[44]. *Alors,*

1. *P est une propriété* résultante *ou* héréditaire *si P est une propriété appartenant à l'une des composantes du système.*

2. *Autrement dit, si aucune composante de x ne possède P, alors P est* émergente, collective, systémique ou gestaltiste.

Bunge cite l'énergie de quelque chose comme étant une propriété résultante parce qu'elle appartient à toutes les composantes de la totalité. En revanche, la propriété d'être vivant est une propriété émergente, selon Bunge parce que cette propriété n'est pas héritée des composantes de l'objet vivant. [Bunge, 1979, cf.page 502]

De cette définition découle le *caractère relatif* de la propriété d'émergence au système qui est considérée comme une totalité. Ainsi, la capacité de penser est une propriété *émergente* du cerveau du primate lorsqu'on la considère comme relative à ses composantes, les neurones. Cette même propriété se révèle, en revanche, *résultante* pour le primate puisqu'elle appartient à une de ses composantes, le cerveau.

L'idée de rendre compatible les propriétés résultantes avec les propriétés émergentes dans un même système place la démarche de Bunge entre les deux alternative extrêmes, l'atomisme et le holisme. Rappelons que pour la première alternative les propriétés d'une totalité sont hérités des propriétés des parties. Pour la seconde, du moins dans ses versions les plus radicales, la totalité transcende ses parties et les propriétés de la totalité sont indépendantes de celles des parties. Comme l'exprime David Blitz :

> The emergentist view takes a *via media* between these two extremes: some system properties are hereditary, others are emergent; consideration of the properties of the parts are necessary, but not sufficient, condition of understanding the system, and must be supplemented with an examination of the properties of the wholes. [Blitz, 1990, page 159]

La caractérisation comme rationnelle d'une théorie de l'émergence est plus stricte pour Bunge que pour Bempel et Oppenheim. Ces derniers ont utilisé le terme *irrationnel* pour caractériser les théories émergentes qui professent l'imprévisibilité *absolue* pour les entités ou les propriétés émergentes. [Hempel and Oppenheim, 1948]. Bunge va plus loin, il qualifie d' *irrationnelles* toutes les théories qui affirment l'irréductibilité ou l'impossibilité d'explication de ces propriétés. Mario Bunge énonce deux postulats pour caractériser l'émergence de manière *rationnelle*. Le premier de ces postulat est d'ordre ontologique:

POSTULAT 1
Tout système possède quelques propriétés qui sont émergentes. [Bunge, 1979, cf. page 503]

Le second postulat est d'ordre rationnel ou épistémologique.

[44] La propriété "être un composant de" n'entre pas en ligne de compte parce que dans la conception de Bunge celle-ci est une propriété universelle. [Bunge, 1979, cf.page 245]

POSTULAT 2
On peut expliquer toute propriété émergente d'un système à partir des propriétés de ses composantes et des couplages entre ces dernières. [Bunge, 1979, cf. page 503]

Bunge semble faire la part des choses entre la réduction ontologique et la réduction épistémologique. Le fait que l'on puisse expliquer les propriétés émergentes en fonction des propriétés des parties et de leurs relations ne signifie pas que l'on ne puisse pas soutenir, par la suite que ces propriétés émergentes ne sont rien d'autre que les propriétés qui sont à la base de l'explication ainsi que le voudrait le réductionnisme ontologique. Cette dernière position radicale est appelée par Bunge réduction simple. L'explication ou réduction d'une propriété émergente ne signifie pas qu'elle n'ait aucune rôle explicatif (voire causal) dans les explications des autres phénomènes. Il s'agit pour Bunge d'expliquer sans pour autant éliminer. [45]

Voici un exemple qui démontre la dualité entre la réduction épistémologique et la réduction ontologique appliquées à un propriété physique.

EXEMPLE 4
For example, refraction is not a bulk [ou global] property of transparent bodies: it is an emergent property relative to the atomic (or molecular) components of such bodies, for none of those components possesses the property of refrangibility. Yet this emergent property of the whole is explained by electrodynamics in terms of the electrical properties of atoms (or molecules) and light. However, this explanation is not reductive in a simple sense, as it does not consist in attributing refrangibility to individual atoms: it is reductive in consisting in the deduction of the formula for refractive power from premises concerning the interaction between electromagnetic waves and atomic lattices.[Bunge, 1977, page 503]

Cette réconciliation entre la réduction et l'émergence est, à mon avis une des grandes contributions de la théorie de l'émergence de Bunge. Elle donne une autre envergure au concept d'émergence car bien qu'une propriété émergente soit relative au système–cadre choisi, elle n'est plus *la salle d'attente* des concepts pour lesquels nous ne disposons pas d'un système d'explication. De plus Bunge donne une théorie–cadre pour caractériser les propriétés émergentes qui s'avère être un véritable critère d'individualisation.

La structuration en niveau chez Bunge: Bunge critique, à mon avis à juste titre, l'ambiguïté du concept de niveaux des systèmes.[46] Je vais donner sa définition de niveaux pour la commenter ensuite.

DEFINITION 14
Soit L une famille des ensemble L_i non–vides des choses, où $1 \leq i \leq n$. Si L_i et L_j sont des éléments de L, alors L_i précède L_j si et seulement si chaque élément de L_j est composé exclusivement de choses appartenant à L_i. En symboles:

$$L_i < L_j =_{df} \forall(x \in L_j \Rightarrow \mathbb{C} \subset L_i)$$

où $\mathbb{C}$ est la fonction composition. ($\mathbb{C}$ va de l'ensemble de parties des éléments à l'ensemble des éléments, ainsi si x est une chose alors $\mathbb{C}$ = l'ensemble de parties de x.) En bref, $(L, <)$ est une relation d'ordre partiel.[Bunge, 1977, ma traduction, cf. page 504]

[45] David Blitz fait une analyse semblable à la mienne, il écrit:

> While it has been traditionallly held that emergence and reduction are incompatible, Bunge has combined them in an innovative way. The incompatibility holds, but only between ontological emergence and ontological reduction, and not between ontological emergence and epistemological reduction. Epistemological reduction is a theoretical operation which does not alter the basic ontology: "In other words, reduction does not imply levelling: it relates levels instead of denying that they exist. Reduction, then is a theoretical question that does not alter the level of the structure of the world"[Bunge, 1979, page 79] [Blitz, 1990, page 159]

[46] Pour une critique voir [Bunge, 1976]

Bunge attire notre attention sur les deux faits suivants. En premier lieu, un niveau n'est pas un objet mais un *concept* et il ajoute un concept utile. La caractérisation des niveaux comme concepts et non comme objets fait que Bunge ait récusé l'interaction des niveaux entre eux. En particulier, il récuse la notion de causalité descendante (*downward causation*) entre niveaux, il le fait explicitement, en particulier :

> [...] levels cannot act upon one another. In particular the higher levels cannot command or even obey the lower ones.[Bunge, 1977, page 504]

En second lieu, il met en avant le caractère spécial de la relation entre niveaux. Ce n'est pas une relation méréologique totalité–partie et ce n'est pas non plus une relation d'inclusion. Bunge la qualifie de relation *suis generis* définissable en termes de composition des fonctions. Seule cette fonction composée sera définissable en termes d'une relation totalité– partie.

Je pense que la définition de cette relation *suis generis* rend l'émergence selon Bunge non–triviale. L'utilisation de ce concept à l'intérieur d'une théorie, en effet nécessite cette définition comme condition préalable. C'est pour cette raison que dans certains domaines, comme par exemple concernant le problème corps–esprit, l'approche émergentiste à la Bunge se présente comme un programme de recherche de longue haleine.

POSTULAT 3 (LE POSTULAT SUR LES NIVEAUX)
Chaque chose appartient à un niveau ou à un autre. [Bunge, 1977, cf. page 504]

Ce postulat est aussi appelé *Principe de hiérarchie* mais il serait faux de l'interpréter par l'assertion qu'un niveau commande au niveau inférieur, ainsi que nous l'avons déjà vu. La hiérarchie doit être comprise au sens où chaque niveau fournit les composantes du systéme au niveau immédiatement supérieur. Finalement, les niveaux peuvent être composés de sous–niveaux.

Les niveaux et l'évolution chez Bunge: Bunge conçoit le monde comme étant organisé en niveaux. Je vais me référer à l'organisation en niveau telle qu'il l'a représentée dans [Bunge, 1979, cf. page 250](d'où j'ai tiré la figure 8.1) mais il faut savoir qu'il a évolué à cet égard.[47]

Tout d'abord cette organisation est de type pyramidal. Le format pyramidal ne signifie pas la subordination ou la précedence d'un type de niveau supérieur par rapport aux inférieurs mais il représente en revanche, le fait que plus un niveau est élevé, plus il dépend des autres et moins il est peuplé.

A la différence de Morgan pour qui l'apparition d'une nouveauté était imprévisible, Bunge récuse cette dernière caractéristique qu'il qualifie, ainsi que nous l'avons déjà signalé, *d'irrationnelle.* Cela veut dire que les nouveautés seront, du moins en principe prévisibles. Blitz caractérise bien la conception de Bunge :

> The type of emergence which he does accept involves the aspects of lawful occurrence (determinism), explanation (rationalism) and multiple factors (pluralism) [48]. In Bunge's view, change does not involve only a single causative factor, but may involve external causes, self– determination, and change factors, all of these resulting in the emergence of novelty. [Blitz, 1990, page 159]

Quel est le rôle des composantes dans le processus évolutif? Les composantes sont, pour Bunge les précurseurs dans ce processus. Cela veut dire qu'elles ont un double rôle, étant à la foi des composantes et des précurseurs. L'exemple est donné par les acides aminés lesquels sont des précurseurs et des composantes de protéines; les cellules pour leur part sont des composantes des organismes multicellulaires et se situent à l'origine de ces organismes.

Le processus selon lequel les composants donnent naissance aux entités du niveau supérieur est caractérisé d'*auto– assemblage*. Ainsi, l'émergence et les différents niveaux, loin de former un système statique constituent en fait un processus dynamique et évolutif.

Dès lors on peut énoncer le dernier postulat en relation avec l'émergence selon Bunge.

[47] Comparer par exemple avec [Bunge, 1977, page 504].

[48] Le *pluralisme* doit être compris en faisant référence à la variété des choses et des processus, en revanche le *monisme* est en relation à la substance sur laquelle les changements s'opèrent qui possède des propriétés.[Bunge, 1979, cf. page 251]

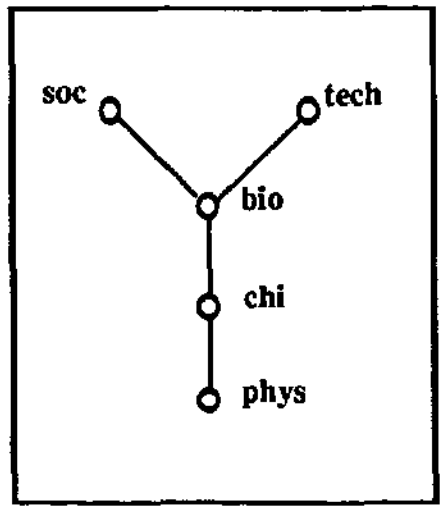

Figure 8.1: Relation de précédence entre niveaux.

POSTULAT *4*
Toute chose complexe appartenant à un niveau donné est l'auto–assemblage des choses appartenant aux niveaux précédents. [Bunge, 1977, cf.page 504]

Ce dernier postulat exprime le caractère à la fois dynamique et naturaliste de son ontologie; or c'est justement ce caractère dynamique qui permet de donner une réponse satisfaisante et scientifique à la question de l'évolution, en lieu et place des autres visions super–naturalistes de l'ontologique.

Ce dernier postulat pour Bunge exprime le caractère à la fois dynamique et naturaliste de son ontologie, c'est justement ce caractére dynamique.

Le but de l'exposition de la théorie de Bunge était de démontrer qu'il existe un concept d'émergence qui non seulement n'est pas une entrave à l'hypothèse de l'unité de la science mais qui de plus permet la caractérisation efficace d'un certain nombre de phénomènes.

Les mérites de cette définition de l'émergence me semblent multiples. Bunge a su récupérer les notions fécondes de la tradition émergentiste comme la notion de différence entre les niveaux qu'il considère comme étant des concepts et non des objets réels; ce faisant il peut récuser l'existence de la causalité descendante dans sa théorie comme non–pertinente. J'ai déjà exprimé les difficultés auxquelles se heurte la causalité descendante du moins dans le cadre des théories évolutionnistes. Les règles qui définissent les différents niveaux sont explicites et leur caractère propre à chaque système montre qu'il faut tenir compte des particularités de ce dernier.

La caractérisation que Bunge fait du concept de système montre le caractère non–réductionniste de son approche tout en étant scientifique; en définitive il montre l'insuffisance des démarches réductionnistes, car un système a non seulement une composition, mais aussi une structure. Ainsi Bunge reprend un des exemples classiques de l'émergentisme britannique pour lui donner une possibilité de description qui s'accorde avec les théories physiques actuelles sans trahir les éléments que les émergentistes du début de siècle voulaient relever. En effet, une étendue d'eau qui est un système a aussi une structure. Ce ne serait pas une description exhaustive que de tenir seulement compte de la composition.

> [...] a body of water is a system, hence something with a structure, not only a composition. And that structure includes the hydrogen bonds among H_2O molecules. The result is a system with emergent properties such as fluidity, viscosity, transparency and others, which its molecular components lack. Surely one can (hope to) understand all of these emergent properties in terms of those of the water molecules and their interactions. That is, one can (hope to) reduce the macroscopic properties of water to the properties of these microproperties. But such explanation –which has yet to be provided– does not accompany an ontological reduction: explained fluidity is still fluidity. [Bunge, 1977, page 506]

Néanmoins, j'aimerais aussi signaler que Mario Bunge soutient une théorie émergentiste de la relation corps–esprit qu'il appelle *matérialisme émergentiste*. Bunge nous met en garde quant au sens qu'il faut donner au mot *théorie* dans ce cadre : il ne s'agit pas d'une théorie dans le sens strict du terme (c'est à dire un système hypothético–déductif) car c'est plutôt une hypothèse programmatique dans la quête des autres théories pour former un corpus.

Bunge exprime le matérialisme émergentiste par le postulat suivant :

POSTULAT 5

(i) *Tous les états mentaux, événements et processus sont des états, des événements et processus dans le système nerveux central (SNC) des vertébrés:*

(ii) *ces états, événements et processus sont émergents lorsque l'on prend comme états relatifs à ces états, événements et processus des composants cellulaires du SNC;*

(iii) *les relations appelées psychophysiques sont des interactions entre les divers sous–systèmes du SNC, ou entre ces derniers et d'autres composants de l'organisme.*
[Bunge, 1977, cf.page 506]

La première thèse affirme le caractère physicaliste de sa position, la deuxième consacre le caractère émergent de sa théorie. La troisième thèse, selon lui doit être comprise comme une version moniste des positions dualistes mystiques comme l'occasionnalisme ou le parallèlisme.[49] Pour nous situer il nous donne l'exemple suivant:

> [...] rather that say that love can color our reasonings, we may say that the right brain hemisphere affects the left one, and that sex hormones can act upon the cell assemblies that do the thinking. In short, ironic as it may sound, the dualistic modes of speech, which encapsulates our undigested experience and which are but metaphorical and vague context of psychoneural dualism, become literal and precise in the context of emergentist materialism. The latter is salvaged from the dualist myth. [Bunge, 1977, page 507]

Je pense que l'on peut considérer la troisième thèse comme celle qui défend le dualisme de propriétés et cette position est cohérente avec un réductionnisme épistémologique et avec la récusation d'un réductionnisme ontologique du mental.

D'autres auteurs comme Roger Sperry défendent une théorie émergentiste de la relation corps–esprit. [50]

Je reviendrai à la théorie de Bunge lors de la conclusion de ma thèse. A mon avis, Bunge et Sperry ont raison de signaler qu'il est nécessaire de tenir compte des conditions d'organisation du cerveau et de soutenir que les propriétés mentales ont un statut ontologique bien qu'elles puissent ou doivent être expliquées.

8.7 Conclusion

Dans ce chapitre j'ai analysé à partir d'une perspective historique le concept d'émergence dans la philosophie des sciences. Tout le long de cet exposé j'ai essayé de mettre en évidence que les deux intuitions fondamentales de l'émergentisme sont, *primo* le caractère imprévisible des entités ou propriétés émergentes conjointement avec le caractère irréductibles des lois, *secondo* l'émergence comme produit d'une complexité spatio–structurelle suffisante des niveaux inférieurs sur lesquels reposent les phénomènes émergents.

J'ai souligné ensuite comment dans certains domaines de la science on a mis l'accent sur l'un ou l'autre de ces traits. J'ai aussi montré que certaines définitions de l'émergence structurées autour de la simple intuition de l'imprévisibilité ne réussissent pas à donner des moyens d'individualisation entre les propriétés émergentes et celles qui ne le sont pas. C'est peut–être par ce fait que

[49] Pour une définition de ces positions dualistes voir chapitre 2.

[50] Pour compte rendu voir [Sperry, 1980] et [Sperry, 1986] aussi [Wimsatt, 1976]. De même voir la controverse entre R.W. Sperry et J. J. C. Smart [Smart, 1981] sur la supériorité d'une position matérialiste émergentiste du mental vis–à –vis de la théorie de l'identité de type.

certains estiment le concept d'émergence d'un intérêt négligeable pour caractériser des phénomènes. Puis j'ai relevé que le caractère d'imprévisibilité fait de l'émergence un concept qui risque soit de devenir une entrave à l'unité de la science, soit de servir simplement de salle d'attente des progrès de la science pour un certain nombre de phénomènes et donc un concept basé sur des faits épistémologiques.

Les définitions d'émergence qui mettent l'accent sur la description des relations totalité–partie et qui réfutent le caractère irréductible des propriétés émergentes semblent plus porteuses d'espoir pour donner un compte rendu des phénomènes dans un cadre donné. L'intuition que je voudrais exploiter dans la conclusion de cette thèse est que lorsque l'on a une description des éléments faite à partir d'une description abstraite de la complexité nécessaire des niveaux de base, on reconnaît par là que cette structure est un exemple d'implantation ou d'une réalisation de la propriété émergente.

Le concept d'irréductibilité dans le cadre des courants émergentistes que j'ai appelés *rationnels* se traduit par le concept d'indépendance ontologique, ce qui n'exclut pas la possibilité d'une explication causale ou d'une réduction épistémologique. Je vais soutenir dans la conclusion de ma thèse que c'est ce concept d'émergence qui s'avère compatible avec une démarche physicaliste non–réductionniste dans l'étude du problème corps–esprit. Dans ce cadre, le concept d'émergence sera convergent à la réalisation et non le contraire comme la position de Beckermann, dont j'ai traité dans ce chapitre, le prétend.

Chapitre 9

L'élan phénoménologique dans les sciences cognitives

Je préfère le domaine des mathématique où on ne sait pas trop bien ce que l'on fait ! C'est la raison pour laquelle je considère aujourd'hui les mathématiques avec un certain détachement et je ne saurais dire s'il existe actuellement un problème strictement mathématique pour lequel j'éprouve un profond intérêt. C'est inévitable, on ne peut pas consacrer toute sa vie aux mathématiques !
René Thom

9.1 Introduction

La plupart des courants physicalistes en sciences cognitives soutiennent que les objets auxquels la cognition est appliquée sont des *morceaux de réalité*. Les versions plus extrêmes du physicalisme considèrent que cette réalité est décrite seulement par des théories de niveau microphysique puisque c'est l'unique réalité qui soit douée d'une structure à part entière. Cependant, nos expériences en général et celles de la perception en particulier dans la vie de tous les jours ne se réfèrent point aux propriétés de la microstructure mais à des macropropriétés et à leurs structures qualitatives. Certains auteurs d'inspiration phénoménologique dont Barry Smith et Jean Petitot ont été ainsi amenés à conclure que les sciences cognitives traditionnelles n'expliquent qu'une partie du problème. Une théorie respectable de l'intentionnalité doit remplir une autre condition que celle d'expliquer le renvoi de l'esprit aux choses du monde extérieur. Elle doit aussi expliquer comment la relation de renvoi peut aller au-delà des propriétés phénoménologiques ou du sens commun pour parvenir à la réalité telle qu'elle est comprise par le réductionnisme physicaliste. Barry Smith expose le programme de ce courant phénoménologique dans le cadre cognitif:

> As philosophers have known at least since Meinong, human cognitive acts are directed towards entities of a wide range of different types. What follows is a new proposal for bringing order into this typological cluster. We shall embrace a broadly naturalistic perspective. [...] We presuppose further that a categorial scheme for the objets of human cognition should be both (1) critical and (2) comprehensive. As to (1), cognitive subjects are liable to error, even to systematic error of the sort that is manifested by the believers in the Pantheon of Olympian gods. Thus not all putative object-directed acts should be recognized as having objects on its own. As to (2), we take it that a categorial scheme should do justice to each sort of object of its own terms, and not attempt to

> eliminate objects of one sort in favour of objects of other, more favoured sorts. Linguistic and other forms of idealism, as well as Meinongian theories, which assign to each and every refering expression or intentional acts an object tailored to fit, yield categorial schemes which fail to satisfy (1). Physicalistic and other forms of reductionism too often yield categorical schemes which fails to satisfy (2). What follows is a categorial scheme that is both critically realistic and comprehensive. Thus it enjoys the benefits of linguistic idealism and physicalism, without [...] the corresponding disavadvanges of each. [Smith, 1994, pages 1–2]

Le programme que Barry Smitb présente n'est pas nouveau hors des sciences cognitives. D'autres auteurs de ce siècle ont mis l'accent sur cette problématique. J'ai déjà exposé dans un chapitre précédent[1] l'ambition de Rudolph Carnap de bâtir un langage susceptible d'expliquer l'accord intersubjectif qui justifie les sciences dans *La Construction logique du monde* et j'ai souligné qu'il prend une position phénoménologique qu'il abandonne très vite par la suite. D'autre part les théoriciens de la *Gestalttheorie* ont eu une grande influence sur les développements postérieurs.

James J. Gibson avait entrepris une position dans ce sens surtout dans la troisième partie de son œuvre où il présente sa conception écologique de la perception. [2].

Pourquoi une théorie qui prend en compte des données phénoménologiques pour caractériser l'objet intentionnel ne va-t-elle pas, du moins à première vue, de pair avec un projet naturaliste *standard* de l'intentionnalité?

La réponse est que les projets naturalistes conduits jusqu'à maintenant dans le cadre des sciences cognitives sont d'inspiration réaliste et physicaliste, en ce qui a trait au problème corps-esprit. Or, ils ne tiennent pas compte des concepts phénoménologiques pour concevoir l'ontologie des choses. Les concepts s'imposent aux sujets qui ont un rôle plutôt passif dans la relation intentionnelle. [3] L'existence des objets n'est pas mise en examen. Le réalisme présuppose leur existence en se posant seulement la question de la détermination de leur référence. Pour répondre à cette question, les projets naturalistes auront recours aux données de la physique scientifique afin de déterminer la référence. Nous avons déjà exposé la position de Putnam à ce sujet dans le chapitre 4 (§4.4) et elle s'avère constituer un bon exposé de cette tendance. On peut qualifier l'attitude traditionnelle, même au risque de la caricaturer un peu, comme étant un curieux mélange de parsimonie ontologique pour ce qui est des propriétés phénoménologiques et de naïveté quant à l'existence des objets.

Des premières et des secondes propriétés: La différence entre premières et secondes propriétés remonte à Locke qui emprunta ces termes à Boyle. Selon Locke, on apprend que la *chaleur* ou le *froid* ne sont pas en l'eau, il s'agit simplement des sensations que l'eau produit dans l'esprit et que ce dernier attribue à l'eau. Ces deux qualités sont foncièrement différentes de celles telles que tailles ou formes qu'il considère comme des qualités premières. Mais qu'est-ce que l'on doit comprendre par propriété seconde? Comme Berkeley l'a déjà signalé, on ne pourrait pas dire que ce sont celles qui relèvent d'un relativisme perceptif parce que cette prémisse impliquerait que toutes les propriétés sensorielles sont des qualités secondes et donc que le monde se verrait démuni de toute propriété sensible.

Alexander [Alexander, 1974] a fourni une autre interprétation de cette distinction qui vise les différents modes d'explication existants pour chacune. L'explication de la différence dans la perception de la taille pour un objet lorsque l'on prend des perspectives différentes est l'angle relatif à l'oeil qui est une fonction non seulement de la taille de l'objet mais aussi de la distance. La démarche est tout-à-fait autre lorsqu'il s'agit de montrer pourquoi un objet semble chaud dans une main et froid dans une autre. L'explication dans ce dernier cas sera que les deux mains ont différents niveaux d'énergie cinétique moléculaire: tandis qu'une des mains prend de l'énergie à l'objet, l'autre lui en donne. On voit bien que la différence entre les deux explications est que, dans le premier cas, elle se rapporte à la taille réelle de l'objet. Dans le second ces en revanche, elle se rapporte à l'énergie moléculaire cinétique de la main du sujet plutôt qu'à la chaleur supposément possédée par l'objet.

[1] chapitre 1 §1.3

[2] Pour un compte rendu de toute l'oeuvre de J. J. Gibson voir [Scaglione, 1991]

[3] Cf. chapitre 4

Selon les explications de la physique scientifique, la suppression des propriétés secondes remonte à Galilée. Dès lors on considère que ces dernières n'ont pas une existence autonome, qu'elles existent seulement dans la mesure où il y a un sujet qui les perçoit. Or, les explications en physique prennent en considération seulement les propriétés quantitatives au détriment des qualitatives. Dans le meilleur des cas, ces dernières sont réduites aux premières.

> Modern physics is, crudely defined, a science of matter. It deals with a rather limited number of ways in which matter manifests itself in phenomenal reality (above all, of course, in controlled contexts of laboratory experiments). Moreover, it deals with these manifestations not as denizens of the phenomenal world but as it were in a purified form, as quantities or magnitudes: qualitative data are treated via mathematical algorithms and concepts. It seeks to use mathematical devices to *explain* the given manifestations by showing how they are consequences of formal laws or principles. Phenomenal reality comes thereby to be filtered entirely through structures of a formal and quantitative sort. The resultant physical models capture only a limited set of the features of phenomenal reality, and many qualitative and morphological structures of phenomenal manifestation are lost to view as such. This is not, as might be supposed, a trivial matter, a consequence of selective attention that is characteristic of all sciences. Rather, [...] the very entities with which physics deals are in certain precise ways shaped and constrained by the filtering structures with which the physicists is compelled to operate. [Petitot and Smith, 1994]

Pour les courants des sciences cognitives standard, j'ai déjà exposé qu'une ontologie respectable sera celle que l'on pourra hâtir en prenant comme base des données de la physique traditionnelle qui font référence aux propriétés appartenant au niveau des microstructures et qui sont caractérisées par des données quantitatives, c'est-à-dire mesurables, voire observables. Toute démarche qui prend comme base de ses représentations l'information projetée au lieu de l'information réelle du monde risque de réduire la théorie cognitive ainsi construite à une forme sophistiquée de solipsisme.[4] Néanmoins, l'ontologie ainsi bâtie n'arrive pas à combler le fossé existant entre le niveau physique et le niveau symbolique. Cette ontologie ne nous donne aucun indice sur la manière de démontrer comment les propriétés macrophysiques qui sont exposées aux sujets percevants dépendent des propriétés microphysiques qui fournissent l'information sur l'environnement par la médiation du système perceptif.[5]

L'unique façon d'éviter l'écueil du solipsisme sophistiqué qui voue à l'échec tout programme visant à hâtir une ontologie fondée sur des données phénoménologiques ou sur le sens commun est de montrer qu'il existe une correspondance entre les données de la physique appartenant au niveau microscopique et celles de niveau macroscopique. Il faudrait prouver que les qualités sensibles des objets peuvent toujours être identifiées avec certaines *variations physiques* appartenant aux concepts scientifiques comme par exemple, dans le cas des couleurs qui peuvent être identifiées au spectre de réflexion de la surface. Il s'agit de développer une physique phénoménale et pour Petitot et Smith ceci est un projet non seulement viable mais nécessaire.

> On the one hand, then, we have objective physical determinations of different modes of manifestation of matter (movement, radiation, etc.), and on the other hand we have phenomenal (qualitative, morphological) manifestations in the sense familiar to us all pre-theoretically. Our thesis here is that phenomenal manifestation is also a mode of manifestation of matter and that there can indeed exist a sort of phenomenal physics. This phenomenal physics is of course different from standard fundamental physics: it is qualitative, macroscopic and emergent. Yet it is, nonetheless, objective. [...] Our task here, therefore, will be that of devising a science of salience in this sense, i.e. a science of the properly qualitative modes of manifestation of matter, with the goal of bridgiag the gap between quantity and quality, or between the physical and the phenomenal modes of manifestation of matter in such a way as to make the latter, too, able to serve as the object of a genuine theory. [Petitot and Smith, 1994]

Bref, la physique phénoménale devra montrer que si l'on utilise des méthodes mathématiques adéquates il est possible d'établir une corrélation entre les qualités sensibles du macroniveau et les variations physiques du microniveau. Seulement ainsi la physique phénoménale sera-t-elle en

[4] Bien que j'aie déjà donné une définition du solipsisme dans le chapitre 7, je me permet de citer la définition de Lalande ici: « Doctrine présentée comme une conséquence logique résultant du caractère idéal (Idéel) de la connaissance, elle consisterait à soutenir que le "moi" individuel dont on a conscience, avec ses modifications subjectives, est toute la réalité, et que les autres "moi" dont on a la représentation n'ont pas plus d'existence indépendante que les personnages des rêves - ou du moins à admettre qu'il est impossible de démontrer le contraires.[Lalande, 1985].

[5] Sauf pour quelques exemples comme la théorie de la vision de David Marr.

mesure de récuser la thèse selon laquelle les données phénoménales ne sont que des apparences subjectives puisqu'elle servira à montrer l'existence d'une dynamique dans le substrat sous-jacent. Cette dynamique est celle qui octroie aux propriétés qualitatives une existence objective.

La démarche phénoménologique telle que Petitot la conçoit sert à faire le pont entre l'environnement et les représentations en montrant de façon explicite les processus intervenants. Il s'agit donc d'un ambitieux programme de recherche qui vise à combler le fossé existant entre la réalité telle qu'elle est décrite par les sciences et telle qu'elle est perçue à partir des macropropriétés.

Pour y parvenir, Jean Petitot propose une outil puissant : les modèles morphodynamiques de René Thom. Cet outil mathématique extraordinaire servira à créer un pont entre l'*être* et l'*apparaître*. Le programme de recherche est long et Petitot croit qu'il exigera les efforts de plusieurs générations.

Dans ce chapitre je vais, tout d'abord présenter les aspects les plus saillants des modèles morphodynamiques. Ensuite, je discuterai la portée philosophique d'un telle modélisation en me concentrant sur deux aspects. Premièrement, je décrirai les différentes interprétations du concept d'*émergence* pour situer la position de Petitot et de Smith à cet égard dans ce contexte. Secondo, j'analyserai les arguments "pour et contre" la possibilité que les modèles morphodynamiques s'avèrent équivalents à la *réduction transcendantale* husserlienne. Je vais soutenir que cette prétention est seulement acceptable au prix d'une lecture ajournée du texte de Husserl *Ideen*.

Finalement et comme conclusion je signalerai deux difficultés que rencontre ce programme. La première difficulté a trait à l'objectivité de la modélisation. En effet, ces explications démontrent l'appropriation de l'information de l'environnement *en fonction* de nos systèmes perceptifs. Je pense que c'est pour cette raison peut-être que leur objectivité est si souvent mise en question.

Deuxièmement, bien que les modèles permettent l'établissement de domaines qualitatifs différents, on sait toujours comment on pourrait définir des ontologies à partir de tels modèles. Dans ce sens, il me semble que les modèles mathématiques, aussi riches soient-ils, restent ontologiquement neutres. D'où la difficulté à combler totalement le fossé entre la réalité et l'environnement.

9.2 Le modèle de la morphodynamique

La *modélisation morphodynamique* a une triple dimension : physique, phénoménologique et structurelle.

Christopher Zeeman et René Thom [6] sont les pères des modèles morphodynamiques. Zeeman a eu l'idée d'exploiter la théorie de Thom dans le cadre de la théorie des systèmes. Ainsi, la théorie des systèmes a connu un tournant nouveau grâce à l'utilisation des résultats mathématiques de la topologie différentielle. Je vais présenter tour à tour les concepts généraux selon les deux points de vue, systémique et mathématique.

9.2.1 Le point de vue systémique

La théorie générale des systèmes standard se consacre essentiellement à définir ce qu'est un système. Selon Thom il faudrait suivre une approche plus naïve que celle de la systémique traditionnelle à la question:

> [...] un système auquel on s'intéresse est nécessairement localisé dans l'espace-temps qu'il occupe. [Thom, 1980, page 81]

Mais qu'est-ce qu'un système dans ce cadre?

[6] Mais ils ont deux approches différentes, le premier utilise ces modèles dans un but pratique tandis que pour Thom il s'agit d'un outil de théorisation. Petitot va encore plus loin et j'entends par la suite exposer ce point de vue selon lequel ces modèles auront un statut philosophique transcendantal au sens husserlien.

DEFINITION 15 (SYSTÈME)
Un système *est le contenu d'un domaine (D) de l'espace–temps où* domaine *de l'espace euclidien est un ouvert*[7] *connexe*[8].

La définition laisse au dehors toute réunion d'objets, comme par exemple un *couvert* qui est l'union d'un fourchette, d'une cuillère et d'un couteau ne sera pas, selon la définition, considéré comme un système. Dans la plupart des cas, le domaine (D ou sa section spatiale à tout instant) est non seulement connexe mais contractile[9], en fait, topologiquement, une boule [10]

Si D est un boîte parallélépipédique, alors D est connexe et contractile. Le bord du domaine (D) constitue la *paroi* du système. La boîte noire idéale sera celle dont la paroi est totalement imperméable à toutes les influences physiques (flux de matière et d'énergie sous toutes ses formes). Alors, on dira qu'un tel système a une évolution rigoureusement autonome du reste de l'univers. Ce système idéal sera aussi inobservable. L'idéal d'isolement complet ne peut pas être atteint puisque par exemple, rien n'est imperméable à la gravitation. L'étude d'un système suppose l'observation de certains événements comme l'entrée et la sortie des flux et c'est en principe à partir de ce bilan que l'on pourra tirer des conclusions sur le système. Or, il n'est pas utile de vouloir séparer totalement le système qui est l'objet d'étude du milieu. En outre, rien n'empêche que la paroi du domaine D soit totalement fictive, il suffit qu'elle serve à contrôler le flux d'échanges. [Thom, 1980, cf. page 82]

Zeeman a proposé une définition de système plus opérationnelle que la précédente [Petitot, 1992, cf. page 1–2]:

DEFINITION 18 (SYSTÈME)
Soit S un système quelconque conçu comme une « boîte noire» et tel que les hypothèses très générales suivantes soient satisfaites:

1. *A l'intérieur de la boîte noire, il existe un processus interne (en général inobservable) X qui définit les* états internes *que le système S est susceptible d'occuper de façon stable. Pour la simplicité, on peut supposer que ceux–ci sont en nombre* fini.
2. *Le processus interne X définit* globalement l'ensemble *des états internes de S. Cette hypothèse est essentielle. Elle signifie que les états internes sont* en compétition *et donc que le choix de l'un d'eux comme état* actuel virtualise *les autres. Autrement dit, ces états n'existent pas en tant qu'entités isolées mais comme composantes d'une structure. Ils s'entredéterminent par* détermination réciproque, *les états virtualisés par le choix de l'état actuel constituant autant de* présupposés *de ce dernier.*

7

LEMME 1 (BOULE OUVERTE)
Dans un espace métrique E de distance d, on appelle boule ouverte *de centre $x_0 \in E$ de rayon $r > 0$, l'ensemble des points de E dont la distance à x_0 est strictement inférieure à r, soit:*

$$B(x_0, r) = \{x \in E \mid d(x_0, x) < r\};$$

de manière analogue, on définit la boule fermée *de centre x_0 de rayon r:*

$$B(x_0, r) = \{x \in E \mid d(x_0, x) \leq r\}.$$

DEFINITION 16 (ENSEMBLE OUVERT)
Soit E un espace métrique de distance d. On dit qu'un sous–ensemble U de E est ouvert *si pour tout $x \in E$ il existe une boule ouverte de centre x contenu dans U.*

8

DEFINITION 17 (ESPACE EUCLIDIEN CONNEXE)
Un ensemble A d'un espace euclidien se dit connexe *si pour toute paire de ses points il existe au moins une ligne polygonale ineluse dans l'ensemble A.*

[9] Soit $U \in \mathbb{R}$ ouvert contractile si $U \to \mathbb{R}^n$ homotope à $u \mapsto a, a \in \mathbb{R}^\kappa$
[10] Dans le cas différentiel un espace ouvert contractile est une boule.

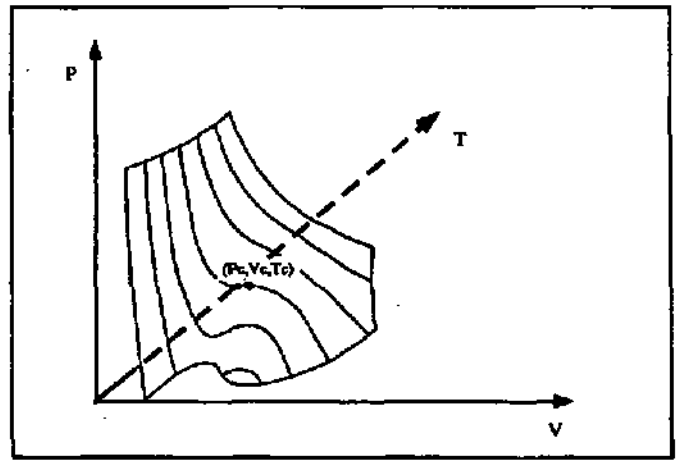

Figure 9.1: Surface de l'équation d'états

3. *Il existe donc* une instance de sélection *I qui, sur la base des critères sélectionne l'état actuel parmi les états internes possibles. Les critères de sélection sont spécifiques au système et peuvent varier considérablement.*

4. *Enfin, autre hypothèse essentielle, le système S est* contrôlé *par un certain nombre de paramètres de contrôle, paramètres variant dans un espace W que, pour l'opposer au processus interne X, on appelle l'*espace externe *(ou espace de contrôle ou encore espace substrat) de S. On suppose de plus que, dans un sens intuitif, non encore spécifié mathématiquement, le contrôle est* continu. *Cela signifie que le processus interne X est un processus interne X_w qui dépend de la valeur w du contrôle, qui varie continûment lorsque w varie continûment dans W et qui, en se déformant, déforme la structure des états internes ainsi que leurs relations de détermination réciproque.*

Mais avant de donner quelque notions mathématiques, prenons un exemple de la thermodynamique: la transition de phases. [11]

EXEMPLE 5
Soit S un système thermodynamique, par exemple de l'eau dans une casserole. Les états internes sont les phases thermodynamiques possibles (solide, liquide, gaz) et X les processus internes qui définissent ces états.

Lorsque l'on observe notre système (la casserole d'eau) on s'aperçoit qu'il existe des changements qualitatifs caractérisés par des valeurs des paramètres de contrôle que l'on appellera des *valeurs critiques*. Par exemple, on observe qu'à pression normale et à température de 100^0 l'eau change de phase. Malgré le fait que les processus X_w soient indescriptibles à cause de leur complexité, ces changements de qualité montrent ou expriment les changements des états internes. Le système externalise sous la forme d'un système de discontinuités, de frontières, la compétence de ses états internes.

Ce problème est traité traditionnellement en utilisant l'équation due au physicien hollandais Van der Waals (1837–1923) qui fournit le graphe de la surface d'états de la figure 9.1. Van der Waals obtient cette équation à partir de celle d'un « gaz idéal» dite de Boyle

$$\left(p + \frac{\alpha}{v^2}\right)(v - \beta) = RT \tag{9.1}$$

[11] [Petitot, 1992, cf. page 2].
Jacques Guélat a fait un intéressant travail de licence de mathématique en 1980 sous la direction du Professeur Sigrist que je remercie de m'avoir fourni une copie de ce manuscrit.

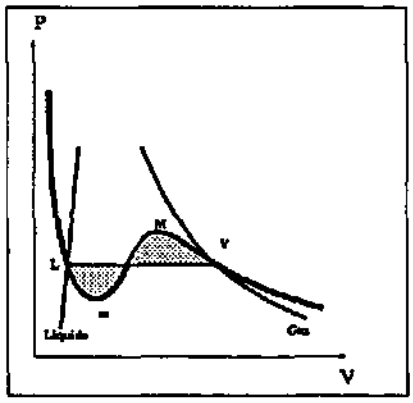

Figure 9.2: Isotherme de Van der Waas ($T < T_c$)

où: p = pression, v = volume, T = température absolue, R qui est la constante universelle des gaz et α, β sont des constantes toutes référentes à l'échantillon de fluide qui est en train d'être considéré. [12]

On peut déterminer à partir de l'équation 9.1 des valeurs de T différentes qui correspondent aux différentes situations:

- Si $T > T_c$ alors il n'y a que l'état gazeux qui est possible, le gaz suit la loi de Boyle–Mariotte (il croît régulièrement) et la fonction $V = V(P)$ ($T = const$) est lisse, c'est-à-dire différentiable.
- Si $T = T_c$ =température critique, V est une fonction continue mais non différentiable de P.
- Si $T < T_c$ la relation $V = V(P)$ n'est pas une fonction. Il y des valeurs de la pression p pour lesquelles *différents* volumes v sont possibles. C'est là que l'on observe le phénomène de transition des phases.

La thermodynamique traditionnelle explique que tout le long de l'arc mM de la figure 9.2 et qui correspond à une suite d'états instables de l'état gazeux est interdite à cause de cette même instabilité. La condensation est le résultat de cette situation. Le système tend alors vers un nouvel état d'équilibre non homogène en se présentant en deux phases. La position de la droite horizontale LV de transition est déterminée à l'aide du second principe de la thermodynamique [13] qui fournit la règle des aires égales due à Maxwel. Cette règle postule l'équivalence entre la chaleur reçue par le système et le travail fourni au milieu extérieur, géométriquement elle consiste à remplacer la

[12] La constante R est souvent notée Nk, où N est le nombre des molécules de l'échantillon et k est la constante dite de Boltzmann 1.380×10^{-6} erg/ deg. En revanche, les constantes α et β sont choisies expérimentalement et elles n'ont pas de signification physique.[Poston and Stewart, 1978, cf. page 328]

[13] Le deuxième principe de la thermodynamique ou Principe de *Carnot-Clausius*. Il est énoncé comme un bilan de la variation d'une fonction du système, appelée *entropie* (S). La variation dS de l'entropie en cours d'une transformation du système peut toujours se décomposer en deux parties: la variation d_eS due à l'échange d'énergie et de matière entre le système et le monde extérieur, et la variation d_iS due à la création ou à la disparition d'entropie au sein du système:

$$dS = d_eS + d_iS$$

Le deuxième principe de la thermodynamique se formule par l'inégalité:

$$dS_i \geq 0$$

Le signe égalité correspond à des transformations réversibles. Dès lors, dans tous les cas, les transformations irréversibles apportent une contribution positive à l'accroissement de l'entropie. [Balescu, 1985, cf. 174]

courbe originale par la ligne droite de façon telle que les deux aires résultantes (qui dans la figure 9.2 apparaissent hachurées) soient égales. Voilà pour l'approche traditionnelle du problème.

Si l'on veut donner à ce même problème un formalisme qui soit en accord avec la théorie des catastrophes, alors on peut procéder de la façon suivante: les variables P (pression) et T (température) sont considérées comme des variables externes et le volume V [14] comme une variable interne. Les états d'équilibre sont donnés selon les équations de l'énergie libre ou enthalpie libre de Gibbs. [15]

Ce principe joue le rôle de critère de sélection dont I sera une instance dans le modèle morphodynamique (cf. plus haut item 3 de la définition 18 de Système).

Dans le but de rendre plus simple l'explication et aussi plus clair le rôle du principe de sélection il me semble utile d'avoir recours à une métaphore que j'emprunte à Thom [Thom, 1983, cf. pages 67]. Donc, pour éclaircir nos idées supposons que la boîte noire qui représente notre système soit habitée par un «démon» dont le but est de maximiser son «gain», représenté par F, et donc de minimiser le potentiel de F.

Plus exactement, les états d'équilibre sont donnés par les minima d'un *potentiel* qui est l'enthalpie.

Alors, le potentiel est représenté par la fonction $F(p, v, T)$ où v est la variable interne et p et T les variables de contrôle. La théorie des catastrophes résoudra le problème en prenant comme représentation l'espace (p, v, T); mais on part aussi de l'équation de Van des Waals.

Ce que nous intéresse ce sont les points singuliers des coordonnées $(p_c, v_c; T)$ de cette fonction.

Pour le moment on va se borner à parler du point d'inflexion sur l'isotherme critique. Notre plan consiste à opérer des transformations locales dans le voisinage du point singulier $(p_c, v_c; T)$. Ces transformations auront les caractéristiques suivantes: *primo* elles devront nous permettre d'appliquer les principes de l'analyse fonctionnelle dans l'espace transformé, *secondo* elles ne changeront pas les propriétés qualitatives tout en ayant la possibilité d'opérer un changement d'échelle et

---

[14] En réalité la densité $\frac{1}{V}$

[15] Le premier principe de la thermodynamique qui reçoit aussi le nom de *principe de la conservation de l'énergie* affirme que pour un système fermé

$$Q = U_1 - U_2 + W.$$

La quantité $U_2 - U_1$ correspond à *l'accroissement de l'énergie* U du système entre l'état initial 1 et l'état final 2. La quantité Q de chaleur reçue par le système, et W est le travail fourni au milieu extérieur. Il y une expression différentielle du premier principe que l'on écrit:

$$dQ = dU + dW.$$

Ainsi l'étude des transformations d'un fluide soumis à une pression uniforme mais dans le cas ou $dp = 0$ c'est-à-dire à pression constante (*isobares*) et l'introduction de la fonction d'état $H = E + pV$, appelée *enthalpie* du système, nous permet d'énoncer le premier principe comme suit:

$$dQ = dH - Vdp$$

A partir du premier et du deuxième principe et en éliminant la différentielle dQ nous pouvons obtenir les relations suivantes:

$$Td_iS = TdS - dE - pdV = TdS - dH + Vdp$$

On en déduit les *critères d'évolution* suivants: pour un système maintenu à V et à S constantes, l'énergie E décroît ($dE \leq 0$) et il en est de même à p et à S constants pour l'enthalpie ($dH \leq 0$). Le signe de l'égalité correspond à une condition d'équilibre. Comme l'emploi des variables (V, T) ou (p, T) est beaucoup plus commode en pratique que celui des variables (V, S) ou (p, S), on opère un changement de variables sur les relations précédentes. En posant:

$$F = E - TS,$$

$$G = H - TS$$

on obtient comme critère d'évolution la décroissance de F, à température T et à volume constants ($dF \leq 0$), ou celle de G, à température et à pression p constantes ($dG \leq 0$). On a donné le potentiel thermodynamique de ces deux nouvelles fonctions d'état F et G. ... Toutefois, les expressions les plus employées actuellement sont celles d'*énergie libre* de Helmholtz pour F et d'*énergie libre* de Gibbs pour G, dite aussi *enthalpie libre*. [Glansdorff and Prigogine, 1985, page 1164]

tertio ce seront des transformations lisses (différentiables) et réversibles [16].

D'abord on doit trouver le point $(p_c, v_c; T)$; pour y parvenir il nous suffit de calculer la dérivée de 9.1 en la considérant comme un $P(V)$ et T fixe. En réalité on travaille sur l'isotherme de la température critique.

$$\frac{\partial p}{\partial v}|_{p_c} = 0$$

et nous savons que

$$p_c = \frac{\alpha v_c - 2\alpha\beta}{{v_c}^3} \tag{9.2}$$

Ensuite, on veut que p_c soit un point d'inflexion. C'est-à-dire;

$$\frac{\partial^2 p}{\partial^2 v}|_{p_c} = 0 \tag{9.3}$$

Alors de 9.2 et 9.3 il résulte que $\alpha({v_c}^3 - 3\beta{v_c}^2) = 0$. D'où

$$\beta = \frac{1}{3}v_c \tag{9.4}$$

$$\alpha = 3p_c{v_c}^2 \tag{9.5}$$

et puis

$$v_c = 3\beta \tag{9.6}$$

$$p_c = \frac{\alpha}{27\beta^3} \tag{9.7}$$

$$T_c = \frac{1}{R}\left[\left(\frac{\alpha}{27\beta^2}\frac{\alpha}{9\beta^2}\right)(3\beta - \beta)\right] = \frac{8\alpha}{27\beta R} \tag{9.8}$$

enfin le point

$$(p_c, V_c; T_c) = (\frac{\alpha}{27\beta^2}, 3\alpha; \frac{8\alpha}{27\beta R}) \tag{9.9}$$

Nous venons d'obtenir les constantes pour la valeur critique. La théorie des catastrophes s'intéresse à l'étude locale autour de ce point critique. Étant donné que l'on s'est permis de changer d'échelle, on normalise de la façon suivante :

$$\bar{P} = \frac{p}{p_c} = \frac{27\beta^2}{\alpha}p \tag{9.10}$$

$$\bar{V} = \frac{v}{v_c} = \frac{1}{3\beta}v \tag{9.11}$$

$$\bar{T} = \frac{T}{T_c} = \frac{27\beta R}{8\alpha}T \tag{9.12}$$

Si on remplace dans 9.1 ces valeurs normalisées nous avons que:

$$RT = \frac{8\alpha}{27\beta}\bar{T}$$

D'où on obtient l'équation réduite de 9.1 comme suit:

[16] Je vais revenir sur ces définitions plus loin dans ce chapitre, il s'agit de difféomorphismes, c'est à dire de transformations différentiables et isomorphes entre les espaces au voisinage du point critique.

$$\left(\bar{P}+\frac{3}{\bar{V}^2}\right)\left(\bar{V}-\frac{1}{3}\right)=\frac{8}{3}\bar{T}$$

Maintenant on remplace le volume $\bar{V}$ par la densité $\frac{1}{\bar{V}}$ qui est une description plus intuitive de la variable d'état, ce qui donne :

$$(\bar{P}+3\bar{X}^2)\left(\frac{1}{\bar{X}}-\frac{1}{3}\right)=\frac{8}{3}\bar{T} \tag{9.13}$$

Lorsqu'on fait la translation à l'origine par :

$$\mathrm{p}=\bar{P}-1 \tag{9.14}$$
$$t=\bar{T}-1 \tag{9.15}$$
$$x=\bar{X}-1 \tag{9.16}$$

L'équation 9.13 devient :

$$x^3+\frac{1}{3}(8t+p)x+\frac{2}{3}(4t-\mathrm{p})=0 \tag{9.17}$$

Si l'on pose :

$$a=\frac{1}{3}(8t+\mathrm{p}) \tag{9.18}$$
$$b=\frac{2}{3}(4t-\mathrm{p}) \tag{9.19}$$

L'équation 9.17 devient :

$$x^3+ax+b=0 \tag{9.20}$$

Mais qu'est–ce que l'équation 9.20 représente au juste? L'équation 9.20 va nous permettre de modéliser la situation étudiée par Van der Waals en la représentant comme une catastrophe que Thom a nommé une *fronce.*[17]

Quelle est la relation entre l'expression 9.20 et la solution traditionnelle des aires égales de Maxwell? La relation n'est pas immédiate. Tout d'abord, les transformations que l'on opére ne garde pas la valeur des aires. Rappelons qu'on s'était permis des transformations qui devaient conserver les propriétés qualitatives tandis que les changements d'écbelle n'étaient guère interdits. L'expression 9.20 ne permet pas l'application de la règle de Maxwell telle qu'elle a été énoncée antérieurement. Ce seront ces variables a, b et x qui seront pertinents à l'analyse des catastrophes.

Le critère de sélection I sera donc plus pertinent pour la détermination de l'ensemble K des points de catastrophe. Thom réénonce le critère de Maxwell sous un autre aspect. Cet énoncé nouveau dira que les états stables correspondent au minimum du potentiel tandis que les états instables reflètent le niveau maximum.

La solution de cette équation cubique dépend de son discriminant $D=4a^3+27b^2$ et comme c'est une équation à coefficients réels, elle a au moins une racine réelle. Les cas possibles sont les suivants :

$D<0$ il y a trois racines réelles distinctes;

$D>0$ il n'y a qu'une racine réelle, les autres deux étant complexes conjuguées

[17] Je vais développer ce concept plus bas; pour le moment l'exemple est cité purement à des fins de motivation.

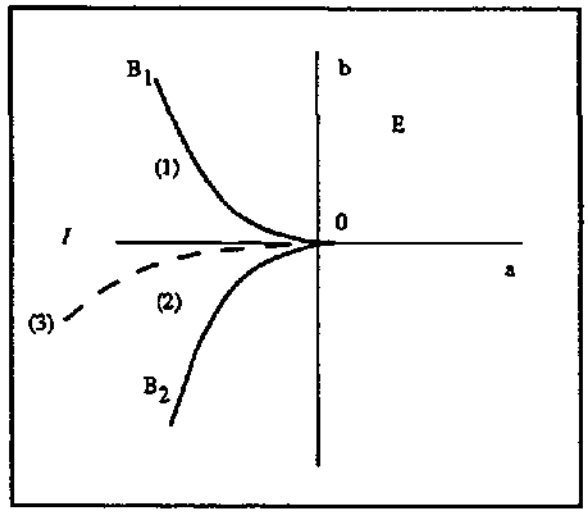

Figure 9.3: Parabole semi-cubique d'équation $4a^3 + 27b^2 = 0$ dans le plan de contrôle (a,b). Dans la région I, la ligne en pointillés (3) sortant de l'origine 0 ($a = b = 0$) indique les points de catastrophes, c'est–à–dire le stade de conflit entre deux régimes (1) et (2).

$D = 0$ il y trois racines réelles mais certaines coïncident; pour $D = 0$ et $a \neq 0$ ou $a \neq 0$ deux racines réelles sont égales; pour $D = 0$ et $a = b = 0$, les racines sont les trois égales.

Géométriquement, cette interprétation de la théorie des catastrophes nous permet de faire le graphique du *plan de contrôle*.

Dans le voisinage de la singularité (p_c, v_c, T_c) nous pouvons voir la situation dans l'espace de dimension 3 comme le montre la figure 9.4. Pour la courbe semi-cubique (voir figure 9.3) de l'équation $4a^3 + 27b^2 = 0$ nous avons les partitions suivantes du plan (a, b): l'origine 0, les deux branches de la courbe B_1 et B_2, la région I *intérieure* à la courbe et la région E *extérieure* à la courbe.

Si le point (a, b) est en E, il n'y a qu'une seule racine réelle C qui corresponde à un minimum d'énergie libre donc un maximum de gain pour le *démon* que nous supposons exister en notre boîte. Dans E il n'y a qu'un seul régime possible.

En I il y a trois racines réelles possibles c_1, c_2, c_3 qui correspondent à deux minima (c_1, c_2) et à un maximum (c_3). Le *démon* a devant lui deux régions qui correspondent aux minima établis pour c_1 et c_2. C'est–à–dire que dans ce cas il y a en I deux valeurs de la variable d'état x, c_1 et c_2 pour le même point de l'espace de contrôle. Notre *démon* choisira en fonction de la *convention de Maxwell*: celle qui correspond au plus petit potentiel. Si F est le potentiel, alors il choisira c_1 si $F(c_1) < F(c_2)$.

Dans notre cas l'espace de contrôle est de dimension 2; néanmoins on peut généraliser à un espace de dimension r. Thom explique les cas possibles de la façon suivante:

> En fait, de la convention de Maxwell il découle qu'un point K de l'espace de contrôle R^r peut être catastrophique uniquement en deux cas: ou bien on atteint le minimum absolu du potentiel $F(x_1 \cdots x_n; u_1 \cdots u_r)$ en deux points distincts $c = (c_1, \cdots, c_n)$ et $\tilde{c} = (\tilde{c}_1, \cdots, \tilde{c}_n)$ avec $F(c_1, \cdots, c_n) = F(\tilde{c}_1, \cdots, \tilde{c}_n)$ (point de *conflit*), ou bien le minimum absolu du potentiel, obtenu en un point unique $c = (c_1, \cdots, c_n)$ cesse d'être stable (*point de bifurcation*).
>
> Appliquons maintenant la théorie de Maxwell à l'intérieur I de la parabole semi–cubique prise en exemple.
>
> Il n'est pas possible de choisir un régime continu à l'intérieur de cette parabole. La strate de conflit, c'est–à–dire l'ensemble de la fonction F pour laquelle $F(c_1) = F(c_2)$, est donnée par un choix de paramètres a, b qui décrivent dans le plan de contrôle (a, b) une courbe (dessinée en pointillés sur la figure [9.3]) sortant de l'origine 0. A l'origine, c'est–à–dire pour $a = b = 0$, correspond un minimum du potentiel non stable: l'origine est donc un point de bifurcation dans le plan de contrôle (a, b). En somme, la *bifurcation engendre la catastrophe!* [Thom, 1983, pages 73–74]

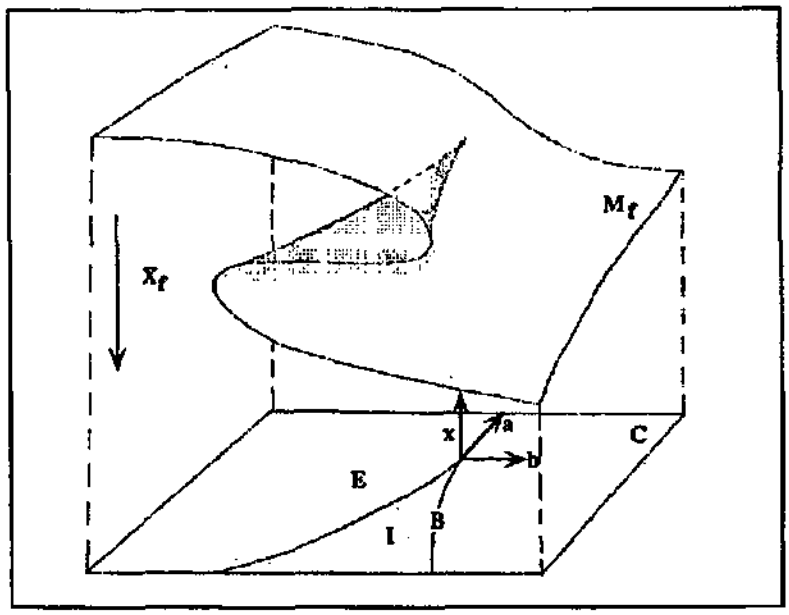

Figure 9.4: Représentation de 3 dimensions

En B_1 et B_2 (sauf à l'origine 0), nous trouvons un minimum et un point d'inflexion; ils seront les points de bifurcation. Enfin, à l'origine 0 comme nous avons vu, $c_1 = c_2 = c_3$

Ces situations sont schématisées dans la figure 9.5

La théorie de la morphodynamique n'apporte rien de nouveau dans ce cas précis, néanmoins elle effectue une reformulation du problème, d'où aussi l'intérêt de l'exemple car il nous permet de faire une évaluation plus claire des différences entre deux points de vue: le point de vue classique et celui des catastrophes.

Thom caractérise même les différences entre la démarche de la physique et la méthode de la théorie des catastrophes comme suit:

> On étudie [dans la théorie de catastrophe] la correspondance entre les entrées et les sorties et, par l'analyse de cette correspondance, on essaye de comprendre les mécanismes en œuvre dans la boîte. Cela indique d'ailleurs clairement que la théorie des catastrophes, sous sa forme la plus pure en quelque sorte, est bien une herméneutique. Elle n'a rien de démiurgique comme la physique. En physique, on dit: il y a des lois, nous allons les découvrir. La théorie des catastrophes dit simplement: il y a continuité, continuité des fonctions, de leurs dérivées. On peut par conséquent traiter l'objet comme un objet analytique et faire des diagrammes, des figures du type des singularités analytiques. C'est la philosophie sous-jacente. [Thom, 1991, page 31]

Je vais présenter maintenant la formalisation mathématique du concept de système.

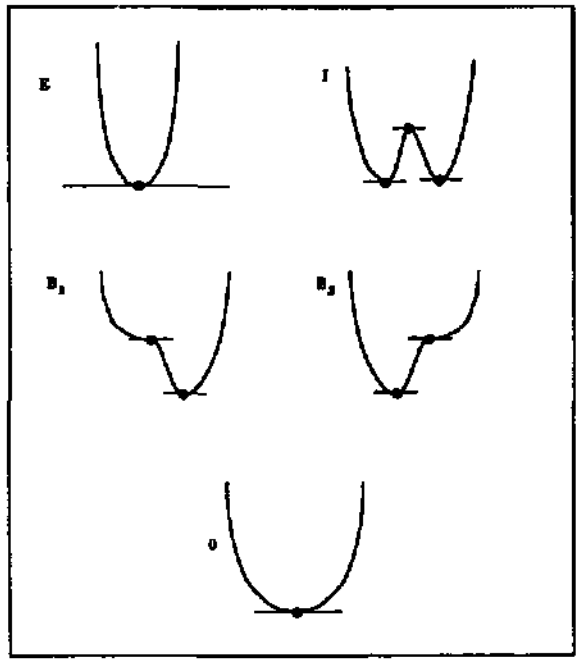

Figure 9.5: Différentes formes de potentiel selon les régions du plan de contrôle (a, b).

9.2.2 Les formalisations mathématiques

L'ensemble des processus internes possibles X sera l'espace χ, qui sera généralement un *espace fonctionnel des champs de vecteurs* [18]sur un espace interne M [19]. Voici un exemple de concept de champs de vecteurs dans une application de la théorie des catastrophes :

EXEMPLE 6

Considérons une enceinte dans laquelle on a mis K substances chimiques $s_1, s_2, \ldots, s_k$, de concentrations respectives $c_1, c_2, \ldots, c_k$. Par suite des réactions chimiques qui ont lieu entre ces substances, les concentrations c_i varient, selon une loi différentielle que nous écrirons:

$$\frac{dc_i}{dt} = X_i(c_1, \ldots, c_k) \, (*)$$

Nous ne chercherons pas à expliquer les seconds membres X_i à l'aide des lois de la cinétique chimique (action de masse, etc.); le seul fait qui nous intéressera est le suivant: les équations $(*)$ définissent dans l'espace euclidien à k dimensions $\mathbb{R}^k$ de coordonnées $(c_1, c_2, \ldots, c_k)$ un champ de vecteurs X de composantes X_i. L'évolution du mélange sera décrite par le déplacement du point représentatif $c_i(t)$ le long d'une trajectoire du système différentiel défini par le système $(*)$.[Thom, 1980, page 13]

L'hypothèse du contrôle continu de l'espace de contrôle W sur les processus internes sera mathématiquement décrite comme une application qui à chaque valeur w assigne un processus de χ (X_w), elle sera donc définie comme un champ continu σ tel quel:

18

DEFINITION 19 (FAMILLE DE CHAMPS DE VECTEURS)
A est une famille de champs de vecteurs si et seulement si

$$A = \{f | f\mathbb{R}^m \to \mathbb{R}^n\}$$

DEFINITION 20 (ESPACE FONCTIONNEL DES CHAMPS DE VECTEURS)
Un espace fonctionnel de champs de vecteurs est une famille de fonctions des champs de vecteurs sur lesquels on a défini une topologie.

[19]La notion d'espace interne M sera expliquée plus tard lors de l'exposition de la théorie généralisée des catastrophes.

$$\sigma : W \to \chi$$

Or, la description du système sera donnée par

$$\mathbb{S} = (\mathbb{W}, \chi, \sigma, \mathbb{I})$$

Manifestation phénoménologique du $S = (W, \chi, \sigma, I)$: Supposons que *l'apparaître* du système S est fournie par les *qualités observables* $q^1 \dots q^n$ qui se manifestent lorsque le système occupe un certain état interne X_w.

Selon l'hypothèse 4 de la définition 18 de système, lorsque le contrôle varie continûment, l'état interne aussi. Les propriétés q^i sont des propriétés dépendantes (voir émergentes) du niveau microscopique, cependant toutes les variations du niveau élémentaire ne vont pas entraîner des variations au niveau macroscopique. Dans le cas de l'exemple thermodynamique du changement de phases, rappelons-nous qu'une diminution de la température à l'intérieur de l'aire E de la figure 9.3 n'entraînera pas un changement de phase; or la caractéristique qualitative liquide ne varie pas. Ces observations motivent les définitions suivantes de point régulier et de point critique ou catastrophique.

DEFINITION 21 (POINTS RÉGULIERS)
Les points w de l'espace de contrôle tels qu'ils varient continûment dans le voisinage de w sans changer les qualités observables où dans lesquelles les propriétés observables restent invariantes sont appelés points réguliers.

DEFINITION 22 (POINTS CRITIQUES OU CATASTROPHIQUES)
Les valeurs w pour lesquelles le système S vérifie un changement d'au moins une des qualités observables s'appellent points critiques. *Lorsque ceci arrive le système S devient le support d'un phénomène critique. On appelle ces valeurs* catastrophiques.

Mais quel est le rôle de I dans la bifurcation ou transition des états internes? Supposons que le contrôle w parcoure un chemin γ dans W et que A_w soit l'état actuel sélectionné par I. Au cours de la déformation de X_w le long de γ et selon l'hypothèse 4 de la définition 18 la structure A_w et les relations de A_w avec les autres états virtuels B_w et C_w aussi se déforment. Il se peut que lors de la traversée d'une valeur critique A_w ne satisfasse plus aux critères de sélection imposés par I. Alors le système bifurque spontanément de A_w vers un autre état actuel et jusqu'alors virtuel B_w.

Dans notre exemple des changements de phases nous avons vu que le point critique est atteint lorsque le système ne vérifie plus le critère de Maxwell et que cela cause un changement d'état. Rappelons que selon ce critère le système tend vers le minimum global. Géométriquement la situation est celle schématisée par la figure 9.6

On peut apprécier le changement discontinu d'état selon les variations continues des variables de contrôle (a et b). Une analyse de l'équilibre du système ne permet pas d'affirmer où de tels sauts se produisent. En principe, ils sont possibles dès qu'il y a plusieurs états d'équilibre pour une même valeur des variables de contrôle (comme c'est le cas dans la région I).

Lorsque le chemin des variables de contrôle est γ on dira que le système obéit à la *règle du retard*: le système reste stable jusqu'au moment où l'équilibre disparaît.

Lorsqu'on analyse l'évolution du potentiel le long du chemin γ, on peut se rendre compte de l'évolution en fonction de la règle de Maxwell: le point de la figure 9.7 représente la variable d'état et on voit qu'il se déplace à chaque instant sur un minimum *global* du potentiel.

Une autre situation intéressante est celle où l'on parcourt le chemin γ mais dans le sens contraire. J'appellerais ce chemin $-\gamma$.

Le saut n'aura pas lieu au même endroit que dans le cas précédent, le système restera le plus longtemps possible dans l'état de la feuille du haut de M_f puis sautera à la feuille du bas quand il ne sera plus possible de faire autrement.

La zone intermédiaire que dans la figure 9.9 on a représentée en sombre est une zone de maxima du potentiel, or elle donne lieu à des équilibres instables. Physiquement de telles situations ne sont pas réalisables, d'où le fait qu'on la considère comme une zone inaccessible.

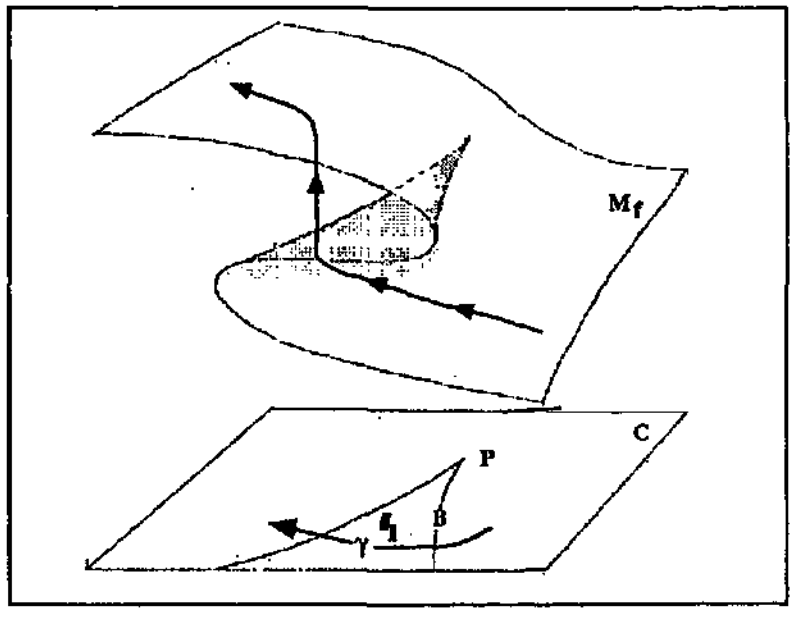

Figure 9.6: Changement discontinu d'état sur un chemin continu dans l'espace de contrôle.

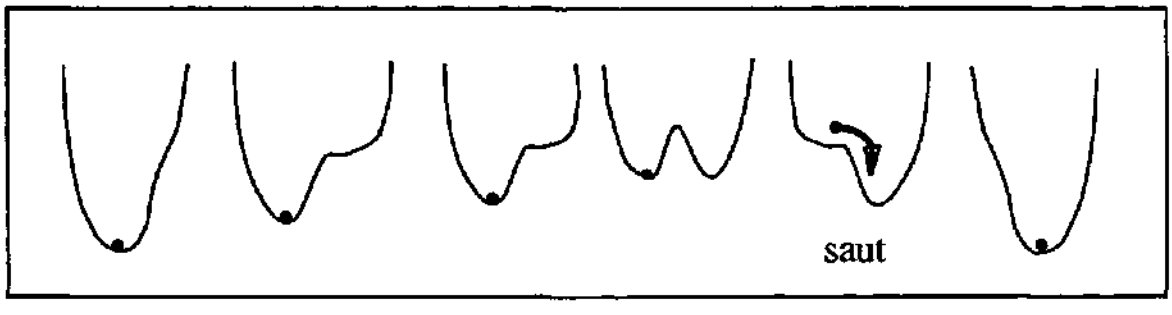

Figure 9.7: Évolution du potentiel le long du chemin γ.

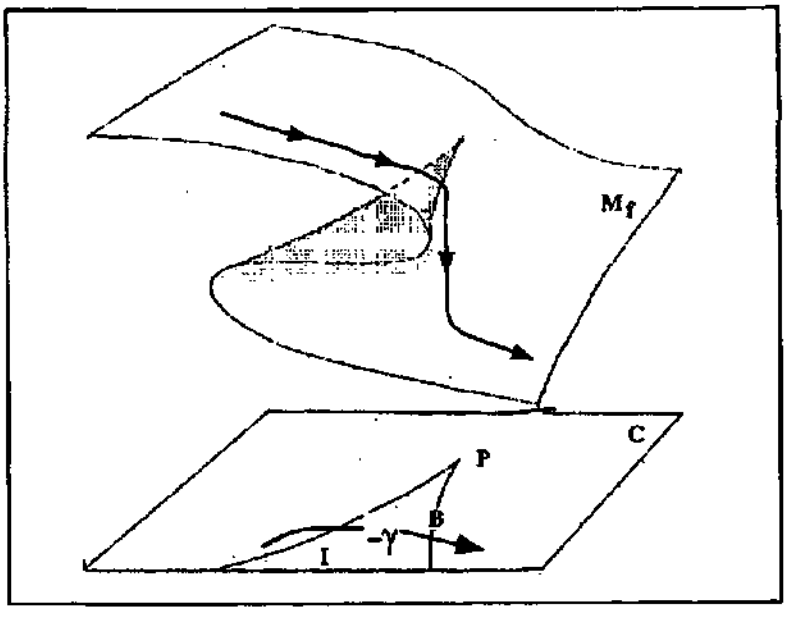

Figure 9.8: Évolution du potentiel le long du chemin $-\gamma$.

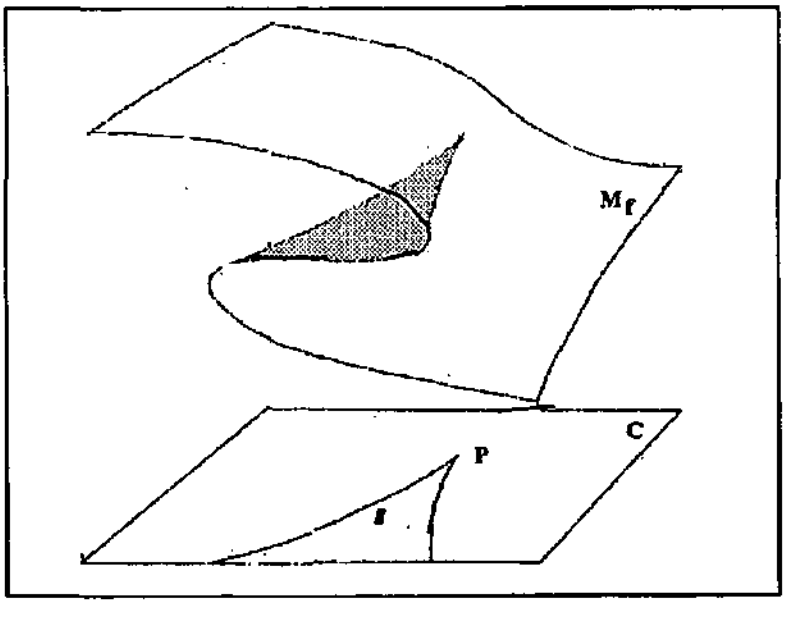

Figure 9.9: Zone inaccessible

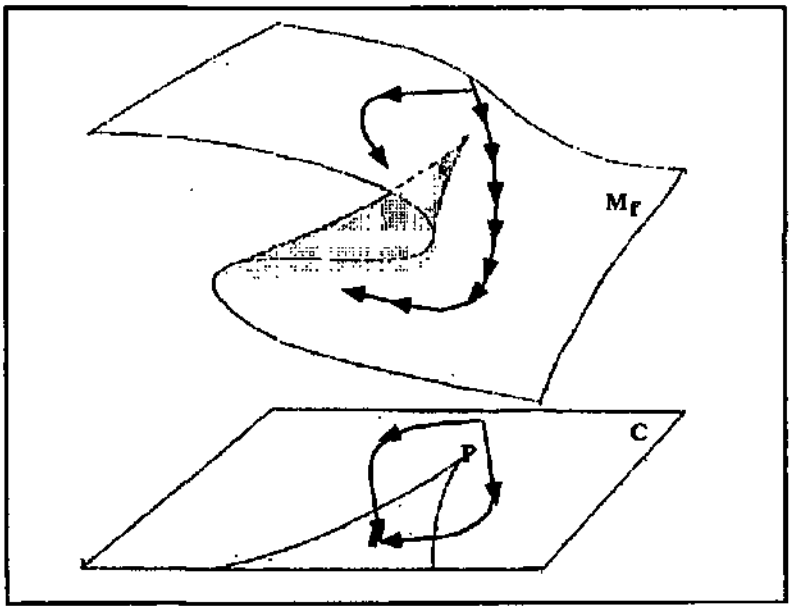

Figure 9.10: Diversion

La figure 9.10 montre qu'il est possible d'avoir des chemins partant du même point et très rapprochés qui produisent néanmoins des différences d'états qualitatifs considérables. En effet on voit bien que l'état dépend du fait que ce chemin passe à gauche ou à droite de P.

Les points critiques qui marquent les changements qualitatifs représentent dans la formalisation mathématique des *singularités*. Ces points présentent des difficultés géométriques et analytiques dans leur voisinage qui ne permettent pas de description facile. En particulier, une simple application du théorème des fonctions implicites à ce point n'est pas possible puisque ce théorème d'interprétation géométrique si riche est seulement applicable aux points réguliers. [20]

L'intérêt d'avoir une géométrie locale de ces points critiques réside dans le fait qu'elle nous permettra une description du système en termes des courbes de niveau des valeur w critiques. Comme le signale Petitot:

> Dans les bons cas (simples), l'ensemble K_w constituera un système d'interface, analogue à un diagramme de phases, partitionnant l'espace externe W en domaines dont chacun correspond à la zone de W où domine l'un des états internes. Les systèmes $S = (W, \chi, \sigma, I)$ *catégorisent* donc naturellement leurs espaces externes. Ce sont des modèles *de l'émergence du discontinu par variation continue*, des modèles de phénomènes critiques. Or, il faut se rappeler que, *bien que phénoménologiquement dominants*, ces derniers ont longtemps été considérés avec suspicion par les scientifiques dans la mesure où ils violent le *principe de continuité* selon lequel une variation infinitésimale des causes ne peut produire qu'une variation infinitésimale des effets. Il y avait là un obstacle épistémologique que vient lever le concept de transition catastrophique. [Petitot, 1992, page 4]

Je vais citer maintenant les définitions et quelques résultats mathématiques nécessaires à la compréhension des concepts des modèles morphodynamiques.

9.2.3 Éléments de la théorie de géométrie différentielle de base

Cette partie du travail n'a en aucun cas, la prétention de constituer un traité formel de géométrie ou de variétés différentiables. Il s'agit seulement de citer quelques résultats qui s'avèrent intéressants pour la compréhension ultérieure des modèles. Un des concepts centraux de la topologie différentielle est celui de *difféomorphisme*. Les transformations difféomorphes nous permettent

[20] Je reviendrai plus tard sur le théorème des fonctions implicites.

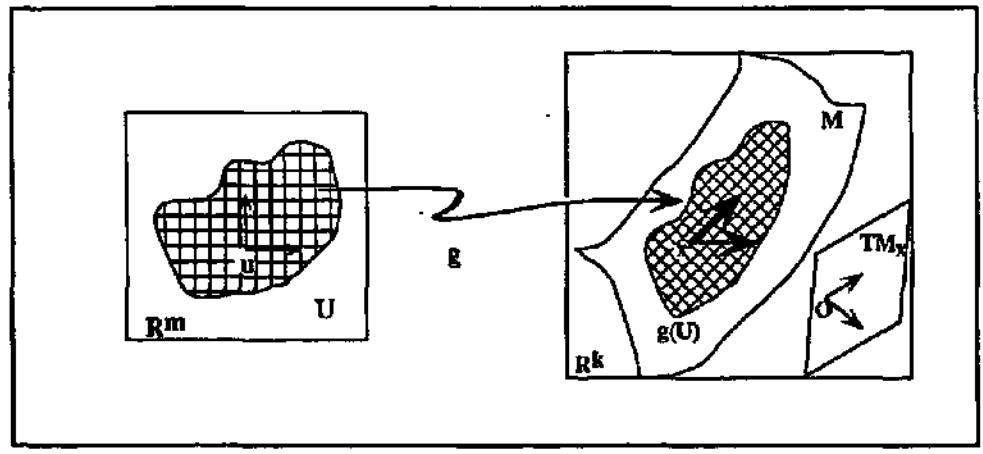

Figure 9.11: Changement de paramètres via difféomorphisme.

d'opérer des changements différentiels (lisses) et réversibles des coordonnées de façon à pouvoir appliquer localement les notions analytiques dans l'espace ainsi transformé.

DEFINITION 23 (DIFFÉOMORPHISME)
Soient X et Y tous deux ouverts en $\mathbb{R}^m$. Une application $f : X \to Y$ est un difféomorphisme si f applique X homéomorphiquement en Y et si f et f^{-1} sont lisses [21]

La topologie différentielle s'intéresse aux propriétés invariantes par difféomorphisme.

DEFINITION 25 (VARIÉTÉ LISSE)
Un sous-ensemble $M \subset R^k$ est une variété lisse de dimension *m si pour chaque $x \in M$ existe un voisinage $W \bigcap M$ qui est difféomorphe à un ouvert U de l'espace euclidien $\mathbb{R}^m$. (cf. figure 9.11)*

Par ailleurs, on appelle *paramétrisation* de la région $W \bigcap M$ à tout difféomorphisme $g : U \to W \bigcap M$. On appelle *système de coordonnées* le difféomorphisme inverse $g^{-1} : W \bigcap M \to U$.

EXEMPLE 7
La sphère unitaire qui est une variété lisse de dimension 2 est formée par tous les $(x, y, z) \in \mathbb{R}^3$ qui vérifient $x^2 + y^2 + z^2 = 1$. Le difféomorphisme

$$(x, y, z) \mapsto (x, y, \sqrt{1 - x^2 - y^2}),$$

pour $x^2 + y^2 < 1$ paramétrise la région $z > 0$ de la sphère. Si l'on change tour à tour les rôles de x et y on obtient des paramètrisations semblables pour toutes les régions qui couvrent la sphère, d'où l'on peut affirmer qu'elle est une variété lisse.

DEFINITION 26 (VARIÉTÉ DIFFÉRENTIABLE DE OIMENSION n)
Une variété différentiable de dimension n est un espace topologique obtenu par recollement des morceaux *d'espace standard $\mathbb{R}^n$. C'est un espace topologique* [22] *qui peut être recouvert par une famille $(U_i)_{i \in I}$ d'ouverts tels que :*

[21]

DEFINITION 24
Soit $U \subset \mathbb{R}^n$ et $V \subset \mathbb{R}^l$ des ensembles ouverts. Une application f telle que $f : U \to V$ se dite lisse *si toutes ses dérivées partielles $\frac{\partial^n f}{\partial x_i \ldots \partial x_n}$ existent et sont continues. (cf. [Milnor, 1965])*

[22] Un espace topologique M est un ensemble où l'on a défini la notion de voisinage entre éléments. Un ouvert U de M est alors un sous-ensemble qui est un voisinage de chacun de ses éléments. Dans l'espace euclidien $\mathbb{R}^n$ les ouverts de base sont les boules ouvertes $B(a, l)$ ensemble des points ce $x \in \mathbb{R}^n$ dont la distance à $a \in \mathbb{R}^n$ est strictement inférieure à $l > 0$ fixé et U est ouvert si et seulement si pour tout $a \in U$ il existe $l > 0$ tel que $B(a, l) \subset U$.

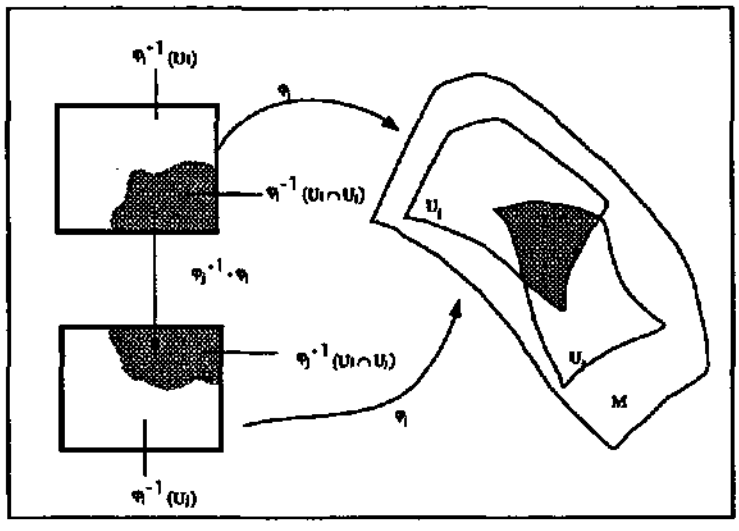

Figure 9.12: Changement de paramètres via diffeomorphisme

1. *chaque U_i soit homéomorphe* [23] *,via un homéomorphisme $\varphi : U_i \to V_i$, à un ouvert V_i de $\mathbb{R}^n$;*

2. *pour tout couple (i,j), les applications induites par $\varphi_j \circ \varphi_i^{-1}$ entre l'image dans V_i de l'intersection $U_i \bigcap U_j$; et son image dans V_j soient différentiables* [24]

Chacune des paires (U_i, φ) ainsi définie est appelée une carte différentiable de la famille de variétés $(U_i)_{i \in N}$

Remarque: Si (U_i, φ_i) et (U_j, φ_j) sont deux cartes différentiables de la variété de classe C^k, alors $\varphi_j^{-1} \circ \varphi_i$ réalise un difféomorphisme de classe C^k de l'ouvert $\varphi_i^{-1}(\varphi_j(U_i))$ inclus dans U_i sur l'ouvert $\varphi_j^{-1}(\varphi_i(U_i))$ inclus dans U_j.

Les U_i et U_j sont des cartes locales et les fonctions $\varphi_j \circ \varphi_i^{-1}$ montrent les changements de cartes locales, voir figure 9.12. A travers les φs, on peut transporter aux U_i systèmes de coordonnées de $\mathbb{R}^n$ et donc raisonner sur U en termes de *coordonnées locales*. Cela permet de prolonger aux variétés différentiables toutes les notions, les entités et les constructions qui relèvent de la structure différentiable des espaces standard $\mathbb{R}^n$ et qui sont de nature locale.

Voilà quelques notions de base nécessaires pour les énoncés qui suivent en référence à l'étude des singularités. Néanmoins, cette étude se révéle loin d'être triviale. Les espaces des fonctions différentiables sont des espaces compliqués. On cherche des descriptions locales qui nous donnent des situations algébriques de dimension finie sur lesquelles il soit possible de faire de la géométrie et donc d'appliquer les résultats de l'analyse des fonctions. Pour y parvenir la notion de variété différentielle est centrale puisqu'elle permet la reconstruction des régions de $\mathbb{R}^n$ par *recollement des morceaux*.

9.2.4 Quelques éléments de la théorie des singularités

Le concept de *singularité* est présenté dans plusieurs domaines de la mathématique (la topologie différentielle, la dynamique qualitative, le géométrie analytique et la topologie algébrique); néanmoins ces concepts ont un dénominateur commun dans tous ces domaines. En effet, il s'agit des

[23] Si M et N sont deux espaces topologiques, les applications $f : M \to N$ qui sont compatibles avec leur structure topologique (i.e. qui préservent la relation de voisinage) sont les applications continues. Une application continue est un *homéomorphisme* si c'est un isomorphisme pour la structure topologique, c'est-à-dire si c'est une bijection continue dont l'inverse est aussi continue.

[24] Une application $f : V \to \mathbb{R}^n$ d'un ouvert V de $\mathbb{R}^n$ dans $\mathbb{R}^n$ est dite différentiable si ses composantes $f_i(x_1, \ldots, x_n) i = 1, \ldots, n$ admettent des dérivés partiels à tous les ordres. Il s'agit d'une notion locale.

points où la dérivé d'une application n'est pas de rang maximal, ou des points où l'espace analytique n'est pas lisse. Souvenons-nous que dans cette situation la simple application du théorème des fonctions implicites ne nous permet pas une description de la *géométrie* du voisinage de la singularité. Ce n'est qu'à partir des travaux de Marston Morse, de Hassler Whitney et de René Thom que la théorie des siogularités des applications différentiables a vu le jour. Petitot décrit les difficultés de l'étude des espace fonctionnels comme suit :

> [Les] difficultés sont nombreuses et en grande partie dues au fait que les espace fonctionnels d'applications différentiables sont des espaces compliqués sur lesquels il est impossible de faire *directement* de la géométrie. La stratégie générale sera donc de réduire, autant que faire se peut, les situations rencontrées *à des situations algébriques de dimension finie* sur lesquelles il devient possible de calculer et de faire de la géométrie. [Petitot, 1992, page 95]

Trivialité du théorème de fonctions implicites: Nous avons vu que dans la théorie des catastrophes nous sommes surtout intéressés par les points singuliers puisque c'est dans ces valeurs que le système vérifie un changement qualitatif. Le théorème des fonctions implicites ne permet pas une géométrie de la situation dans le cas des points singuliers, tout simplement parce que ce théorème est seulement applicable au point où le rang de la matrice Jacobienne est maximal, alors qu'il ne peut être appliqué que pour les point réguliers.

La notion de dérivabilité pour une fonction lisse $f : M \to N$ des variétés lisses consiste aussi à associer à chaque $x \in M \subset \mathbb{R}^k$ un sous-espace linéaire $TM_x \subset \mathbb{R}^k$ de dimension m dit espace tangent de M en x. Alors df_x sera une application linéaire de TM_x en TN_x, où $y = f(x)$. Les éléments de l'espace vectoriel TM_x sont appelés *vecteurs tangents* à M en x. Intuitivement on pense à l'hyperplan m-dimensionnel qui s'approche le mieux de M dans le voisinage de x, alors TM_x est un byperplan à l'origine parallèle à celui-ci.

Petitot parle de la trivialité du théorème des fonctions implicites car celles-ci ne peuvent décrire que des *bonnes situations*. Mais qu'est-ce qu'une *bonne situation* ?

Soient $f : M \to N$ et $a \in M$. Considérons f au voisinage de a et de $f(a)$ c'est-à-dire la situation locale en $(a, f(a))$. Suivant les dimensions respectives m et n de M et de N, les *bonnes situations* sont les suivantes :

1. $m = n$, f est localement l'identité $\mathbb{R}^m \to \mathbb{R}^m$;
2. $m < n$, f est un plongement $\mathbb{R}^m \to \mathbb{R}^n$;
3. $m > n$, f est localement la projection $\mathbb{R}^m = \mathbb{R}^{m-n} \times \mathbb{R}^n \to \mathbb{R}^n$;

Nous dirons que f est localement triviale si elle est, localement en $(a, f(a))$, différentiablement équivalente à la *bonne situation* qu'imposent les dimensions m et n. Le *théorème des fonctions implicites* affirme que la trivialité locale dépend seulement de l'application linéaire tangente df_a de f en a et cette trivialité se vérifie si df_a, qui est une application linéaire de l'espace vectoriel T_aM de dimension m dans le vectoriel $T_{f(a)}N$ de dimension n, est de rang maximal. La valeur du rang maximal dépend des n et m et alors on a les situations suivantes :

1. $m = n$, alors rang maximal $=m=n$
2. $m < n$, alors rang maximal$=m$
3. $m > n$, alors rang maximal$=n$

Nous pouvons maintenant donner une définition des *points réguliers* lorsqu'on parle des éléments de la source (x) et des *valeurs régulières* lorsqu'on parle des éléments du but $(f(x))$.

DEFINITION 27 (POINT RÉGULIER)
Soient $f : M \to N$ et $a \in M$. On dit que a est un point régulier de f si l'opplication linéaire tangente df_a est de rang maximal.

Rappelons que dans des sections précédentes nous avons défini les points réguliers dans le cadre du modèle morphologique comme les valeurs où les propriétés observables restent invariantes. Mathématiquement on voit qu'ils correspondent aux points où l'application linéaire tangente est de rang maximal, c'est-à-dire aux points où le graphe est lisse. Les caractéristiques plus fines du voisinage des points réguliers dépendent des relations entre les dimensions n et m. [25]

Cas $n = m$:

THÉORÈME 3 (THÉORÈME D'INVERSION LOCALE)
Si $m = n$ et si $a \in M$ est un point régulier de f alors f est, localement en a, inversible (i.e. est, localement en a, un difféomorphisme)

Le théorème des fonctions implicites est un corollaire du théorème précédent d'inversion locale.

THÉORÈME 4 (THÉORÈME DES FONCTIONS IMPLICITES)
Soit $f : U_1 \times U_2 \to \mathbb{R}^n$ une application différentiable où U_1 est un ouvert de $\mathbb{R}^k$ et U_2 un ouvert de $\mathbb{R}^n$. Soient $x_0 \in U_1$ et $y_0 \in U_2$ et f_{x_0} l'application $f_{x_0} : U_2 \to \mathbb{R}^n$ donnée par $f_{x_0}(y) = f(x_0, y)$. Si $f(x_0, y_0) = y_0$ et si l'application linéaire tangente $D_{y_0} f_{x_0}$ de f_{x_0} en y_0 est de rang n, alors, quitte à restreindre U_1 et U_2, il existe une application différentiable $g : U_1 \times U_2 \to U_2$ telle que $g(x_0, y_0) = y_0$ et $f(x, g(x, y)) = y$ pour tout $x \in U_1$ et $y \in U_2$ [26].

Cas $m < n$: Soit a un point régulier de f (i.e. $D_a f$ est de rang maximal m). On dit alors que f est une immersion en a. f est une immersion si c'est une immersion en tout point de M. Un résultat important est le suivant :

THÉORÈME 5
Soit $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ une immersion en $a \in U$. Alors par changement de carte dans le but $\mathbb{R}^n$, f peut se ramener à l'injection canonique $\mathbb{R}^m \to \mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^{n-m}$ (restreinte à U). Autrement dit, f est localement triviale et peut être linéarisée.

Cas $m > n$ Soit a un point régulier de f (i.e. $D_a f$ est de rang maximal n). On dit alors que f est une submersion en a. f est une submersion si c'est une submersion en tout point de M.

THÉORÈME 6
Soit $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ une submersion en $a \in U$. Alors par un changement de carte locale en a, f peut se ramener à la projection canonique $\mathbb{R}^{m-n} \times \mathbb{R}^n \to \mathbb{R}^n$ (restreinte à U). Autrement dit, f est localement triviale et peut être linéarisée.

Ces résultats montrent dans le cas des immersions et des submersions qu'au voisinage d'un point régulier, une application est différentiablement équivalente à sa partie linéaire $f(a) + D_a f(x - a)$; par changement de carte locale dans la source ou le but, on peut éliminer les termes du développement de Taylor de f en a de degré plus grand ou égal à 2 dans un voisinage a.[27]

Points et valeurs critiques

Dans le voisinage des points réguliers il n'y a pas de problème pour trouver une application différentiable linéaire qui s'approche de la fonction. Le problème se présente dans le voisinage des points non-réguliers , dits *singuliers ou critiques* et dont les images s'appellent des *valeurs singulières ou des valeurs critiques.*

Selon Chenciner [Chenciner, 1985] le lemme de Sard est l'unique résultat de structure globale applicable à toute fonction C^∞ sur $\mathbb{R}^n$ et il est à la base des théorèmes de transversalité de René Thom.

[25] Les propriétés suivantes sont des résultats présents dans [Petitot, 1992]
[26] Intuitivement, le théorème des fonctions implicites dit que si f est défini sur un produit $U_1 \times U_2$ et sa dérivée à x constant est inversible alors on peut exprimer y en fonction de f.[Petitot, 1992, page 104]
[27] [Petitot, 1992, cf. pages 104-105].

LEMME 2 (LEMME DE SARD)
Soit $f : U \to \mathbb{R}^n$ une application lisse, définie sur l'ouvert $U \subset \mathbb{R}^m$ et soit C l'ensemble des points critiques; c'est-à-dire l'ensemble de tous les $x \in U$ tels que

$$. \; rang(df_x) < p = \min(n, m).$$

Alors $f(C)$ est de mesure nulle. [28]

Si m < n le théorème est trivial; en effet dans $C = U$ car une application différentiable ne peut augmenter la dimension de sa source puisque $f(M)$ (bien qu'éventuellement fort complexe) sera de dimension au plus n dans M et donc de mesure nulle. Le cas intéressant est celui où $m \geq n$. [29]

Une des conséquences importantes du lemme de Sard se réfère à la densité de l'ensemble des points réguliers. Il affirme, en effet que bien que les singularités d'une application puissent être extrêmement complexes, elles sont nécessairement *rares*. Donc ce corollaire important dû à Arthur B. Brown peut être énoncé comme suit:

COROLAIRE 1
L'ensemble des valeurs régulières d'une application lisse $f : M \to N$ est dense partout en N. [Milnor, 1965, ma traduction page 11]

Les résultats que j'ai présentés jusqu'à ici sur les propriétés des points critiques font allusion aux propriétés globales mais on n'a pas encore énoncé un résultat que l'on puisse considérer comme équivalent au théorème des fonctions implicites pour les points ou valeurs critiques. C'est seulement lorsque Morse a énoncé le lemme qui porte son nom que l'on a pu fixer des conditions nécessaires aux changements de coordonnées locales dans le voisinage des points critiques des fonctions f tels quels $f : \mathbb{R}^\kappa \to \mathbb{R}$. Alors on a compris qu'il existe des points critiques *bons* et des *laids*.

Mais avant de continuer sur la théorie des points singuliers, un rappel du concept de forme quadratique s'impose.

DEFINITION 29 (FORME QUADRATIQUE SUR $\mathbb{R}$)
Une forme quadratique en n variables $x_1, \ldots, x_n$ est une expression

$$q(x) = \sum_{ij} \lambda_{ij} x_i x_j, \lambda_{ij} \in \mathbb{R}.$$

Si l'on appelle $\kappa = (\lambda_{ij})$ la matrice de la forme quadratique, alors on obtient l'expression

$$q(x) = x \kappa x^T$$

où

$$x^T = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \tag{9.21}$$

Si l'on remplace κ par $\frac{1}{2}(\kappa + \kappa^T) = M$, la forme quadratique reste inchangée puisque $x_i x_j = x_j x_i$ et M se trouve être *symétrique* ce qui veut dire $M = M^T$. Or toute forme quadratique peut être écrite comme suit:

$$q(x) = x M x^T$$

avec une matrice symétrique M.

[28] La mesure nulle fait référence à la mesure de Lebesgue et je prends pour définition, bien qu'informelle celle de [Milnor, 1965, cf. page 10];

DEFINITION 28
In other words, given any $\varepsilon > 0$, it is possible to cover $f(C)$ by a sequence of cubes in R^n having total n-dimensional volume less than ε

[29] Pour une démonstration voir [Milnor, 1965, chapitre 3]

DEFINITION 30 (LE RANG D'UNE FORME QUADRATIQUE)
Toute forme quadratique en n variables peut être transformée comme suit :

$$z_1^2 + z_2^2 + \ldots + z_r^2 - z_{r+1}^2 - \ldots - z_s^2$$

en appliquant une transformation linéaire non-singulière des variables. [Poston and Stewart, 1978, cf. page 21-23] où $s \leq n$. Le nombre s est appelé le rang de la forme quadratique *et l'on peut démontrer qu'il est le rang de la matrice de cette forme. Or le rang de la forme ne dépend pas de la transformation linéaire choisie.*

Un résultat très intéressant est la loi d'inertie de Sylvester qui affirme que r est aussi indépendant de la transformation linéaire choisie. On appelle le nombre

$$2r - s = r - (s - r) = \sum(coefficients)$$

la *signature* de la forme quadratique.

DEFINITION 31 (FORME QUADRATIQUE DÉGÉNÉRÉE)
Une forme quadratique est dégénérée si son rang r est plus petit que le n nombre de variables.

Par ailleurs, le co-rang, c'est-à-dire la différence $n - r$ indique le nombre de directions indépendantes où la forme est dégénérée. Ces directions ont une expression explicite dans la forme diagonalisée de la forme quadratique.

Nous allons voir que les points critiques non-dégénérés admettent une approximation par une forme quadratique elle aussi non-dégénérée, ce qui peut être considéré comme une généralisation du théorème des fonctions implicites à ces types des points.

DEFINITION 32 (POINTS CRITIQUES NON DÉGÉNÉRÉS)
Soit $f : \mathbb{R}^\kappa \to \mathbb{R}$ on dit que u est un point critique non dégéneré si $Df|_u = 0$ et si $D^2 f|_u$ est une forme quadratique non dégénerée (le rang est égal au nombre de variables n). Autrement dit, ceci équivaut à ce que la matrice Hessien [30] soit non singulière.

Pour les points critiques non dégénérés d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ le lemme de Morse affirme qu'il existe un changement de coordonnées dans le voisinage de ce point dont la fonctioo admet une expression simple.

LEMME 3 (LEMME DE MORSE)
Soit u un point critique non dégénéré d'une fonction lisse $f : \mathbb{R}^n \to \mathbb{R}$. Alors il existe un système local de coordonnées $(y_1, \ldots, y_n)$ dans un voisinage U de u tel quel pour tout i il se vérifie que $y_i(u) = 0$ et

$$f = f(u) - y_1^2 - \ldots - y_l^2 + y_{l+1}^2 + \ldots + y_n^2$$

pour tout $y \in U$. [31]

Comme on l'a déjà anticipé, le lemme de Morse montre que l'on peut transformer par un difféomorphisme tout point critique non dégénéré en une l-selle de Morse. La l-selle de Morse est une forme quadratique qui a l'expression suivante :

$$z_1^2 + \ldots + z_{n-l}^2 - z_{n-l+1}^2 - \ldots - z_n^2.$$

A partir du lemme de Morse on peut tirer quelques conclusions sur des propriétés des points critiques non dégénérés. Comme le signalent Poston et Stewart :

---

[30]

DEFINITION 33
Soit $f : \mathbb{R}^n \to \mathbb{R}$ et soit $u \in \mathbb{R}^n$, la matrice $H(u) \in \mathbb{R}^{n \times n}$ définie comme $H(u)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}|_u$.

[31] Pour un démonstration du théorème voir [Poston and Stewart, 1978, page 54-57].

> Since Morse saddle clearly is an isolated critical point, and since smooth coordinate changes leave isolated critical points isolated, it follows that a nondegenerate critical point is always isolated. The number l is an invariant of the tolegical type of the critical point in the following sense: a smooth reversible coordinate change does not alter l.
>
> At a non-Morse critical point, Hessien is degenerated. We can measure *how* degenerate by computing its corank, the number of 'directions in which it is degenerate'. This number is independent of smooth invertible coordinate changes ... [Postoo and Stewart, 1978, page 56]

Le lemme de Morse montre qu'il existe une forme normale dans le voisinage des points critiques non dégénérés; dans ce sens, on peut le considérer comme une généralisation du théorème des fonctions implicites. Rappelons que le théorème des fonctions implicites montre qu'il existe un changement de coordonnées locales qui permet d'éliminer les termes non-constants du développement de Taylor de f dans le voisinage d'un point u dans le cas où u s'avère régulier. Le lemme de Morse, en revanche applicable à des points singuliers non dégénérés permet de trouver un changement de coordonnées locales en éliminant tous les termes du développement de Taylor de f en u de degré plus grand que 2. Si $(y_1, \ldots, y_n)$ est le système de coordonnées locales alors $f(x) = f(u) + x^t H x$ où H est le hessien de f en u. Il suffit dès lors, par un changement de coordonnées, de ramener la forme quadratique $x^t H x$ à ses axes principaux pour ramener f à sa forme normale de Morse.

Un autre lemme important qui découle du lemme de Morse est le *splitting lemma*. Ce résultat est très puissant puisqu'il nous permet de diviser l'expression d'une fonction autour d'un point critique dégénéré entre une expression de Morse pour l'ensemble des variables et une autre partie qui a une expression sur un ensemble des variables différentes du précédent dont le nombre est égal au co-rang.

LEMME 4 (*Splitting Lema*)
Soit $f : \mathbb{R}^n \to \mathbb{R}$ *une fonction lisse telle que* $Df(0) = 0$ *et dont le Hessien en 0 est de rang* r *(et le co-rang est* $n - r$*). Alors* f *est équivalente autour de 0 à une fonction de la forme*

$$\pm x_1^2 \pm \ldots \pm x_r^2 + \hat{f}(x_{r+1}, \ldots, x_n)$$

où

$$\hat{f} : \mathbb{R}^{n-r} \to \mathbb{R}$$

est lisse.

La puissance de ce lemme est évidente. Il dit que l'étude d'une fonction autour d'un point critique dégénéré peut se faire par l'étude d'une fonction dont le nombre de variables est égal au co-rang du hessien. L'étude d'une fonction de 2001 variables autour d'un point critique dégénéré de co-rang 3 se réduit donc, à l'étude d'une fonction à trois variables [Poston and Stewart, 1978, cf. page 63].

9.2.5 Transversalité et stabilité structurelle

Une notion importante que René Thom a introduite est celle de *transversalité*. Je ne vais pas faire l'exposé de toute la théorie mathématique mais je voudrais seulement signaler l'idée plus ou moins intuitive de cette notion.[32]

Thom proposa la notion de transversalité pour donner forme au concept de stabilité structurale ou topologique. Dans les paragraphes suivants je vais exposer le concept de stabilité structurelle de Morse et le concept de tranversalité pour pouvoir finalement démontrer comment ses deux concepts peuvent être apparentés.

[32] Pour un développement complet voir [Thom, 1980, chapitre II], [Petitot, 1992, chapitre IV] et aussi [Chenciner, 1985]. En outre, je vais suivre dans cette partie les concepts et les motivations tels qu'ils ont été exposés par [Poston and Stewart, 1978, chapitre 6]

La stabilité structurelle

DEFINITION 34 (STABILITÉ STRUCTURELLE)
La définition générale: $f : M \to N$ est structurellement stable si et seulement si pour tout application $g : M \to N$ assez voisine de f il existe $\varphi \in Diff(M)$ et $\Psi \in Diff(N)$ tels que $\Psi \circ f = g \circ \varphi$.[Petitot, 1992, page 114]

Autrement dit, lorsqu'on perturbe f et qu'on utilise à sa place g alors le type topologique est conservé.

Pour fixer les idées, essayons de voir ce qui arrive à une fonction f qui a un point critique en 0 lorsqu'on la perturbe avec une petite fonction p dont les dérivées partielles s'avèrent être faibles dans un voisinage de 0.[33]

Il existe deux cas possibles: soit f est de Morse, c'est à dire une fonction dont tous les points critiques sont non dégénérés, soit elle n'est pas de Morse.

Si f est de Morse, alors le point critique en 0 sera non–dégénéré et donc le déterminant du Hessien sera non–nul. Si la fonction p est suffisamment petite, alors le Hessien de la fonction sommation $(f+p)$ sera non–nul en 0 étant donné que le Hessien varie continûment. En conséquence la fonction $f+p$ vérifie être aussi de Morse, aussi la valeur I de la selle de Morse pour les points critiques en question est la même pour f que pour $f+p$. D'où l'on peut dire que f et $f+p$ sont équivalents après avoir appliqué la translation à l'origine pertinente.

EXEMPLE 8
Soit $f(x) = x^2$ et $p(x) = 2\epsilon x$ où ϵ est une constant petite. La fonction $(f+p)(x) = x^2 + 2\epsilon x = (x+\epsilon)^2 - \epsilon^2$.

Cette dernière fonction a un point critique de Morse en $x = -\epsilon$. Le point critique s'est déplacé de $-\epsilon$ mais ceci sans changer son type.

En revanche la tableau est bien différent lorsque s'agit des points critiques dégénérés.

EXEMPLE 9
Soit $f(x) = x^3$ et $p(x) = \epsilon x$ où ϵ est une constante petite. La fonction $(f+p)(x) = x^3 + \epsilon x$ n'aura aucun point critique si ϵ est positive. Cependant, si ϵ est négative alors on va trouver deux points critiques de Morse: une *maximum* et un *minimun*.[34]

Tout ceci montre que la fonction de $f(x) = x^3$ n'est pas structurellement stable puisqu'en effet une petite variation fait changer non seulement la nature des points critiques mais aussi leur nombre.

Voilà pour la notion de stabilité structurelle, mais ne perdons pas de vue notre objectif qui est de montrer que ces phénomènes de stabilité structurelle seront traduits en termes mathématiques dans des situations de transversalité.

Transversalité: Les propriétés de transversalité sont *génériques* puisque dans ces conditions rien de spécial se passe [Poston and Stewart, 1978, cf. page 67]. Selon Thom

> Toutes ces propriétés étaient bien connues des anciens Géomètres italiens du début du siècle, lesquels employaient souvent la méthode *della piccola variazione* pour simplifier une figure [Thom, 1980, page 45]

Petitot nous invite à comprendre intuitivement la notion de la manière suivante:

> La notion de transversalité est très intuitive et, dans son intuition même, indissociable de celle de la *généricité*. Elle précise la vieille notion géométrique de position générale. Considérons par exemple deux courbes γ_1 et γ_2 dans un plan et soit x un de leurs points d'intersection. γ_1 et γ_2 s'intersectent

[33] J'emprunte cette analyse de [Poston and Stewart, 1978, cf. page 63–64].

[34] En effet, il suffit de calculer la $d(f+p) = 3x^2 + \epsilon$. Alors si $\epsilon > 0$ la fonction n'a point de racine mais si $\epsilon < 0$ alors elle a deux racines dans $x = \pm\sqrt{\frac{\epsilon}{3}}$. Si l'on calcule $d^2(f+p) = 6x$ alors on voit qu'il s'agit d'un maximum et d'un minimum.

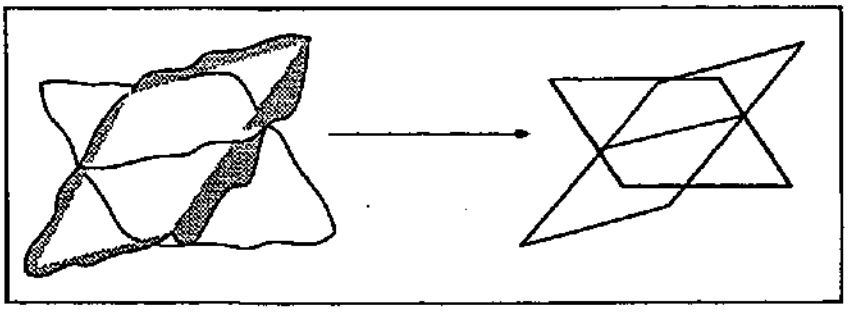

Figure 9.13: Toute intersection transversale entre deux sous-varietés V et N qui verifie $v + n \geq w$ a l'aspect que montre cette figure pour $w = 3$ et l'on peut la transformer sans perdre la propriété de transversalité.

> transversalement en x si elles ne sont pas tangentes. Il est clair que cette propriété est générique dans la mesure où, si γ_1 et γ_2 sont tangentes en x, on peut, par petites déformations, faire « exploser» ce point de tangence en un certain nombre d'intersections transversales. Supposons alors γ_1 et γ_2 plongées dans $\mathbf{R}^3$. Le fait de s'intersecter transversalement n'est plus générique car, disposant d'une dimension supplémentaire, on peut, par petites déformations, disjoindre γ_1 et γ_2. [Petitot, 1992, page 107]

En effet, la position plus générale pour des courbes dans un plan est qu'elles se coupent ou que leur intersection est vide et non qu'elle soient identiques ou qu'elles soient tangentes. Ces deux derniers cas seront des cas *spéciaux* et non *généraux*. En revanche, lorsqu'on considère ces courbes dans l'espace, alors dans le cas le plus général leur intersection sera vide puisqu'il existe une dimension de plus.

La détermination de la transversalité doit prendre en compte non seulement les dimensions des variétés mais aussi la dimension de l'espace où elles se trouvent immergées comme on le verra dans la définition suivante :

DEFINITION 35 (VARIÉTÉS TRANSVERSALES)
Soient V et N deux sous-variétés de W, et l'on pose $dimW = w$, $dimV = v$ et $dimN = n$; on dit qu'elles sont transverses *si, en chaque point x de $V \cap N$, l'espace tangent à W en x est engendré par les vecteurs tangents en x à V et les vecteurs tangents en x à N.*

Si $v + n < w$, il n'est pas possible qu'un espace vectoriel de dimension w soit engendré par la réunion d'un espace de dimension v et d'un espace de dimension n; donc il n'y a pas de point dans $V \cap N$.

Si $v + n \geq w$, le théorème des fonctions implicites entraîne qu'il existe un voisinage Ω de x dans W tel que $(\Omega, \Omega \cap V, \Omega \cap N)$ soit homéomorphe à $(\mathbb{R}^w, \mathbb{X}^v, \mathbb{Y}^n)$, où $\mathbb{X}^v$ (resp.$\mathbb{Y}^n$) est le sous-espace vectoriel de $\mathbb{R}^w$ engendré par les v premiers (resp. les n derniers) vecteurs de la base; il en résulte que $V \cap N$ est une variété de dimension $v + n - w$. [35]

On dit qu'un prolongement différentiable f de V dans W est transversal *à la sous-variété tranverse N de W si $f(V)$ et N sont des sous-variétés transverses de W. On définit aussi des fonctions $f : V \to W$ transverses à N. On peut munir l'ensemble $Hom^\infty(V, W)$ des applications de classe C^∞ de V dans W d'une topologie telle que l'ensemble des fonctions transverses à N soit un ouvert partout dense. [Morlet, 1985, page 178]*

La figure 9.13 illustre la transversalité comme une propriété stable car elle se maintient lorsqu'on lui fait subir de petites transformations.

Quelle est la relation entre stabilité structurale de Morse et transversalité de Thom?

La grande découverte de Thom a été de caractériser les situations de stabilité structurelle de Morse comme des situations de transversalité. Le théorème d'isotopie de Thom démontre que l'on peut déduire la stabilité structurelle de Morse à partir des entrecroisements transversaux avec la

[35] Voir figure 8.1

fonction nulle. La démonstration du théorême fait appel aux *espaces de jets* et elle est loio d'être triviale. Je me contenterai ici de montrer intuitivement la relation entre les deux notions.

La condition nécessaire pour le lemme de Morse est que le Hessien soit non singulier. Cette dernière condition équivaut à la condition que la matrice du Jacobien soit non-singulière, c'est-à-dire que

$$Df = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right) : \mathbb{R}^n \to \mathbb{R}^n$$

s'avère non-singulière. Ceci assure que le graphe de la fonction Df rencontre la fonction nulle transversalement. Ainsi, on peut déduire non seulemeot la stabilité de Morse pour une fonction à partir du fait que sa fonction dérivée est transversale à la fonction nulle mais aussi de sa *généricité*.

Pour préciser reprenons les deux fonctions précédentes doot une se vérifie être de Morse

$$f(x) = x^2$$

et une autre qui ne l'est pas

$$f(x) = x^3$$

Pour $n = 1$ les conditions pour que Df soit transversale à la fonction nulle signifient que $\frac{d^2f}{dx^2} \neq 0$. Ces conditions qui caractérisent un point de Morse avec $n = 1$ sont les mêmes. Dans la figure 9.14 on voit que Df est transversale à la fonction nulle et que sa $\frac{d^2f}{dx^2} = 2 \neq 0$.

Dans la figure 9.15 représente $f(x) = x^3$ qui, ainsi que nous l'avons vu dans les sections précedentes, n'est pas structurellement stable. Dans ce cas, nous constatons aussi que Df n'est pss transversale à la fonction nulle puisque les deux fonctions (Df et la fonction nulle) sont tangentes; le croisement est donc non-générique. Par ailleurs, la $\frac{d^2f}{dx^2}$ est singulière ou nulle (dans le cas de $n = 1$) dans ce point.

> From the full statement of the Thom transversality theorem, which says that typically a function f has the graphs of Df and of its higher derivatives meeting a particular mainfold transversality, we may deduce the typicality of Morse functions. [Poston and Stewart, 1978, page 71]

J'énoncerai maintenant le théorème central de la stabilité de Thom-Mather qui montre que les différents types de stabilité se trouvent être équivalents entre eux.

THÉORÈME 7 (THÉORÈME DE STABILITÉ DE THOM-MATHER)
Soient M compact et $f : M \to N$. Alors les conditions suivantes sont équivalentes.

1. *f est structurellemeot stable;*
2. *f est infinitésimalement stable;*
3. *f est stable par déformation;*
4. *f est homotopiquement stable;*
5. *f est transversalement stable;*

[Petitot, 1992, page 121]

9.2.6 Théorème de la classification de Thom

Dans cette sectioo je vais donner uoe idée intuitive de la manière doot la classification des catastrophes élémeotaires de Thom est apparentée aux conditions de transversalité.

Comme [Poston and Stewart, 1978, cf. page 99] l'indique il s'agit de répondre aux deux questions converses suivantes :

1. Etant donné une famille de fonctions r-paramétrées, quelle seront les types locaux que l'oo va trouver?

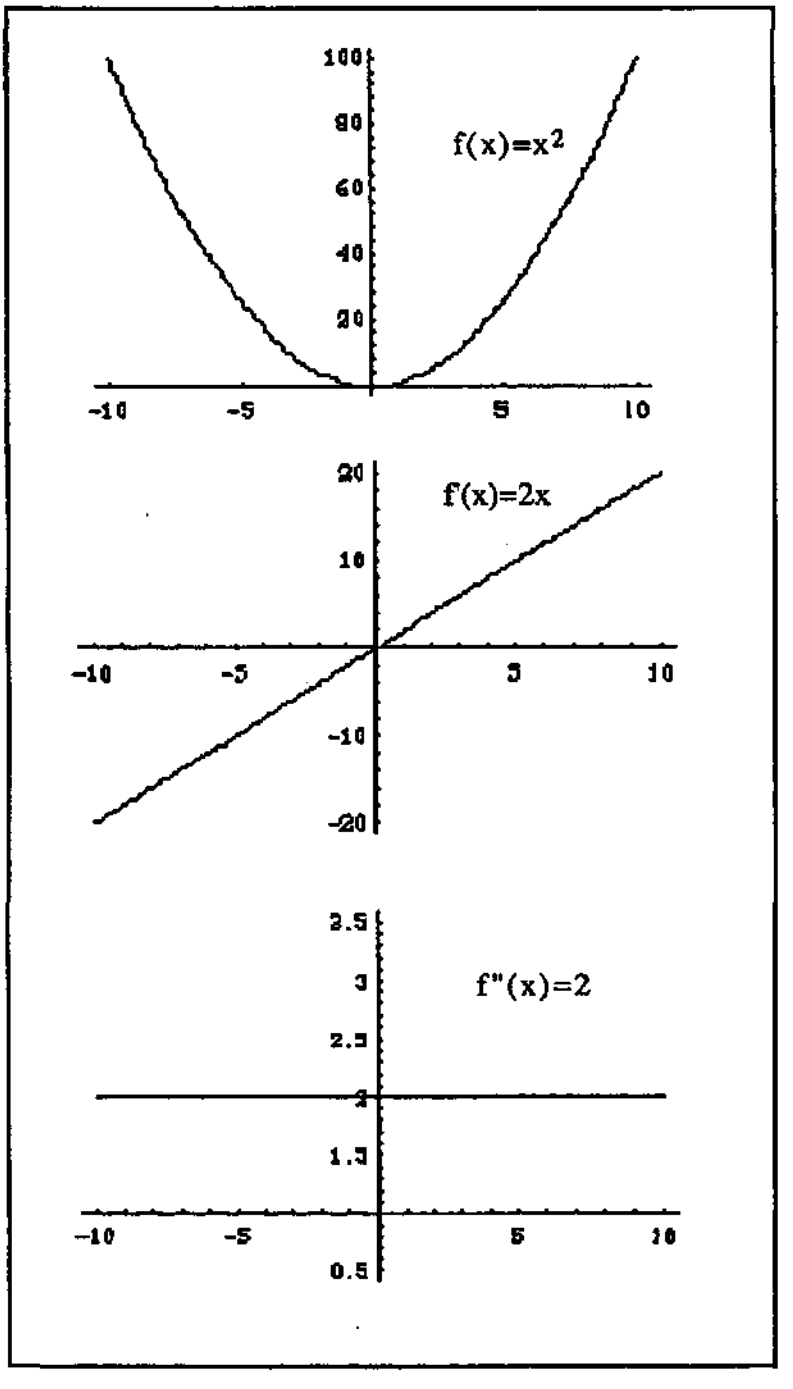
100
80
60
40
20
f(x)=x2
-10
-5
5
10
20
10
f'(x)=2x
-10
-5
5
10
-10
-20
2.5
3
f''(x)=2
2.5
2
1.5
-10
-5
5
10
0.5

Figure 9.14:

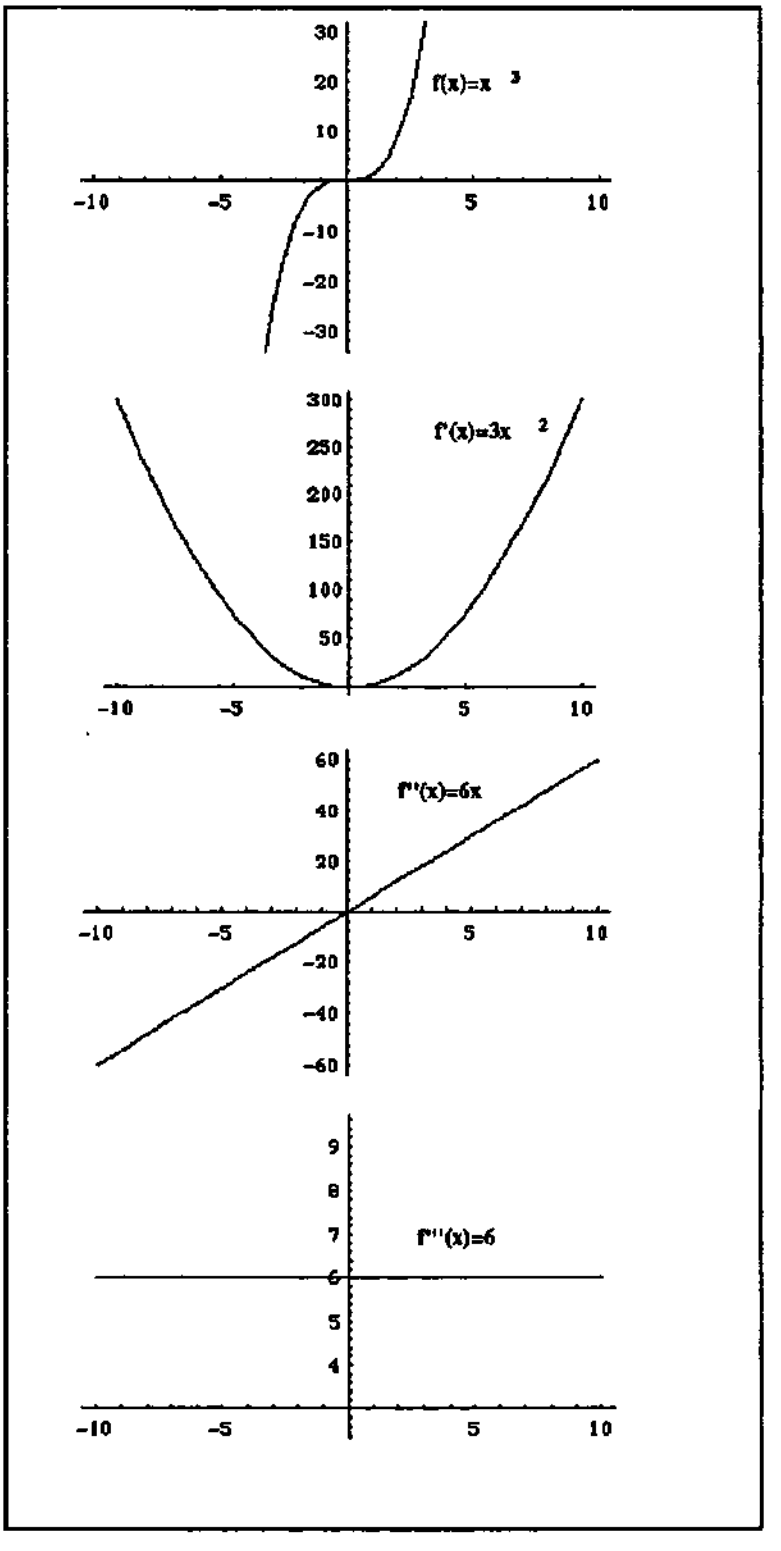

Figure 9.15:

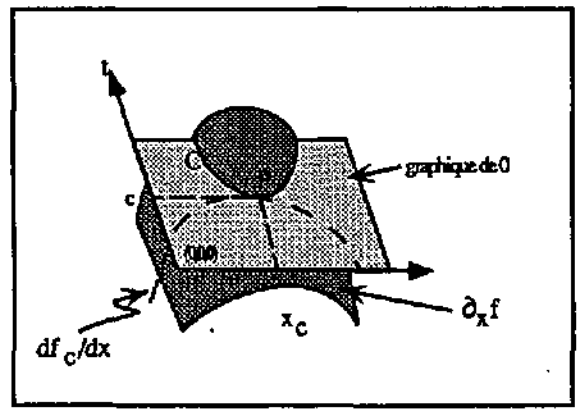

Figure 9.16:

2. Etant donné une fonction, quel sera l'aspect de la famille la plus proche à laquelle elle appartient?

Prenons comme exemple le cas d'un famille de fonctions uniparamétrée. Souvenons-nous qu'on peut déduire la stabilité structurelle de Morse d'une fonction à partir de la transversalité. Lorsque l'on perturbe une fonction très peu, sans qu'elle soit paramétrée, les points singuliers restent du même type.

Ce qui nous intéresse maintenant est de caractériser une séquence complète de perturbation. Ceci veut dire étudier les perturbations d'une fonction en relation avec un paramètre : une fonction $f : X \to \mathbb{R}$ qui change selon un paramètre équivaut à considérer que cette fonction change en fonction d'une variable additionnelle que j'appellerai t.

Une famille de fonctions de ce type $f_t : X \to \mathbb{R}$ avec $t \in \mathbb{R}$ peut être notée comme un seule fonction lisse comme suit :

$$f : X \times \mathbb{R} \to \mathbb{R}$$

$$(x,t) \to f(x,t) = f_t(x).$$

Une observation importante à faire est qu'une famille de fonctions peut contenir des fonctions que s'avérent ne pas être de Morse; néanmoins ceci n'empêche pas la famille d'être une *famille* générique.

On dèsire voir si dans cette famille il y a une fonction f_t dont la dérivée ne soit pas transversale à la fonction nulle. On peut caractériser la famille des fonctions dérivées $Df_t : X \to \mathbb{R}$ comme suit :

$$\partial_x f : X \times \mathbb{R} \to \mathbb{R}$$

$$(x,t) \mapsto \partial_x f(x,t) = Df_t(x).$$

On calcule seulement la dérivée partielle dans la direction X car, pour le moment, nous sommes intéressés par le point critique pour une valeur fixe mais arbitraire de t.

Il faut noter qu'il se peut que pour une valeur donnée de $t = c$, la fonction df_c s'avère non transversale à zéro en x_c mais néanmoins $\partial_x f$ est transversale au plan nul en (x_c, c) de même que partout. C'est justement la généricité de cette transversalité que l'on va exploiter.

Il est intéressant de regarder de plus près la relation entre les dimensions des fonctions transversales.

Supposons que la fonction nulle et $\partial_x f$ soient deux surfaces bi-dimensionnelles en $\mathbb{R}^3$; elles ne vont pas se rencontrer génériquement dans des points isolés pour des raisons de dimension. Elle vont se rencontrer, en revanche dans une courbe C (voir figure 9.16)

Prenons maintenant le cas où $X = \mathbb{R}^n$; alors la fonction nulle et $\partial_x f$ seront des applications

$$\mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1} \to \mathbb{R}^n$$

alors les graphes de ces fonctions vont appartenir à $\mathbb{R}^{n+1} \times \mathbb{R}^n = \mathbb{R}^{2n+1}$. Elles vont se rencontrer donc de façon générique ou transversalement en un ensemble de dimension $(n+1)+(n+1)-(2n+1)=1$, qui sera une courbe que nous appellerons désormais C.

Lorsque l'on étudie la fonction dérivée il y aura trois cas possibles :

1. des endroits où $\partial_x f \neq 0$ c'est-à-dire où f_t est régulier.
2. des endroit où df_x rencontre la fonction nulle transversalement, c'est-à-dire comme une singularité de Morse.
3. des point où la singularité n'est pas de Morse.

Dans les deux premiers cas, une petite variation du paramètre ne va pas changer fondamentalement les fonctions de la famille.

Lorsque l'on voyage sur la courbe C on trouve aussi des points critiques et dans eux le développement de Taylor n'a pas de terme linéaire. Notre étude concerne les termes quadratiques. On va négliger la valeur constante de Taylor puisque ce qui nous intéresse est la forme des fonctions et non les valeurs.

L'étude de la partie quadratique nous montre qu'elle ne peut pas avoir une seule direction dégénérée dans le cas où elle est dégénérée, ce qui signifie que le co-rang du Hessien dans cette singularité non-Morse est maximum 1. [36]

Or, si l'on applique le *spliting lema* dans ce point la fonction f à la forme suivante :

$$f(u_1, \ldots, u_n; t) = \tilde{f}(u_1, t) \pm u_2^2 \pm \ldots \pm u_n^2$$

Sans perdre le caractère de généralisation, on peut se borner à l'étude des fonctions $\tilde{f}_t$ d'une seule variable pour caractériser la famille des fonctions monoparamétrées.

On va changer $\tilde{f}_t$ par f et u_1 par $\tilde{x} = x - x_s$. Lorsque l'on voyage sur la courbe C nous pouvons calculer le développement de Taylor de la nouvelle fonction unidimensionnelle f jusqu'à l'ordre 4 comme suit :

$$f_t(\tilde{x}) = k_s + p_s\tilde{x}^2 + q_s\tilde{x}^3 + r_s\tilde{x}^4 + Taylor$$

Les coefficients p_s, q_s, r_s sont trouvés par l'évaluation des dérivées sur les valeurs de C. L'ensemble de ces valeurs possibles donne une courbe $\tilde{C}$ dans l'espace de coordonnées tridimensionnelles (p,q,r). [37] On va appeler $\tilde{P}$ le point où la courbe $\tilde{C}$ coupe le plan (q,r). Dans ce point le terme quadratique p s'annule. La question est de savoir s'il sera possible d'avoir p et q simultanément nuls. La réponse sera négative et la raison repose sur un argument de généricité. Si l'on suppose que p et q peuvent être simultanément nuls, alors la courbe $\tilde{C}$ rencontrera l'axe r, mais ceci n'est pas possible parce que génériquement deux courbes (dimension 1) ne se retrouvent pas dans un espace de dimension 3 (elles ne sont pas transversales). Alors, lorsque f_t n'aura pas la forme $p\tilde{x}^2 + Taylor$, elle aura la forme $q\tilde{x}^3 + Taylor$. Dans le premier cas, la fonction aura la forme $\pm u^2$ dans une paramétrisation de Morse, dans le deuxième nous allons voir que sa forme sera u^3.

En vertu d'une des conséquences du *Splitting Lema* on peut la réduire exactement à la forme u^3. [38]

[36] Voir [Poston and Stewart, 1978, page 102-104].

[37] Il s'agit d'étudier l'espace des coefficients du polynôme c'est-à-dire l'espace dual.

[38] Le résultat auquel je me réfère est le suivant :

THÉORÈME 8
Soit $f : \mathbf{R} \to \mathbf{R}$ une fonction lisse, telle quelle

$$f(0) = Df|_0 = \ldots = D^{k-1}f|_0,$$

mais

$$D^k f|_0 \neq 0$$

Alors il existe un changement lisse des coordonnées sur lesquelles f prend la forme

$$x^k$$

Résumons : A partir de la notion de transversalité et de co–dimension dans notre analyse nous avons trouvé que la fonction $f : \mathbb{R}^n \to \mathbb{R}$ a:

- Dans le voisinage des points non–singuliers, la forme

$$(u_1, \ldots, u_n; t) \mapsto u_1$$

- Dans le voisinage des points de Morse, la forme

$$(u_1, \ldots, u_n; t) \mapsto \pm u_1^2 \pm u_2^2 \ldots \pm u_n^2$$

- Dans des points isolés, f_t a la forme

$$(u_1, \ldots, u_n; t) \mapsto u_1^3 \pm u_2^2 \pm \ldots \pm u_n^2$$

Les développements tels qu'ils ont été présentés jusqu'ici servent de motivation pour la notion de *jet*, concept cardinal par exemple, dans la démonstration du théorème d'isotopie de Thom.

DEFINITION 36 (LA NOTION DE JET)
Soit $f : \mathbb{R} \to \mathbb{R}$ de classe C^∞. La meilleure approximation locale de f en a par les polynômes de degré n est donnée par:

$$j^n f(a) = f(a+h) = f(a) + hf'(a) + \ldots + h^n \frac{f^{(n)}(a)}{n!} + \ldots$$

Bien que les dérivés des fonctions ne soient pas des entités intrinsèques, la propriété de deux fonctions d'avoir le même développement de Taylor en a jusqu'à un certain ordre k est en revanche une propriété invariante par changement de variables locales. *L'on peut définir l'équivalence de f et g en a à l'ordre k; la classe d'équivalence de f s'appelant jet d'ordre k de f en a et se notant $j^k f(a)$ [Petitot, 1992, cf. page 99]*

DEFINITION 37 (ÉQUIVALENCE DE FONCTIONS ET GERME)
Sur un ensemble de fonctions de $\mathbb{R}^q$ en $\mathbb{R}^m$, définies au voisinage d'un point x (par exemple $x = 0$); on peut définir une relation d'équivalence de manière suivante: $f \sim g$ si, et seulement si, il existe un voisinage de x où f et g coïncident. La classe d'équivalence de f pour la relation $\sim$ (c'est–à–dire l'ensemble des fonctions g telles que $g \sim f$) est dite germe *de la fonction f. Pour $m = q$, on $E(q,q)$ on indiquera avec $B(q)$ l'ensemble des germes réversibles $\mathbb{R}^q$ en $\mathbb{R}^q$ qui appliquent 0 sur 0. [Thom, 1989, Cf. page 170]*

Deux applications $X : M \to N$ et $X' : M' \to N'$ seront dites *équivalentes* s'il existe des difféomorphismes ϕ et ψ tels que le diagramme suivant soit commutatif:

$$\begin{array}{ccc} M & \xrightarrow{\phi} & M' \\ \downarrow{\scriptstyle X} & & \downarrow{\scriptstyle X'} \\ N & \xrightarrow{\psi} & N' \end{array} \tag{9.22}$$

Le concept de germe analytique qui permet d'exprimer une équivalence entre fonctions facilite aussi la mise en évidence de la structure dans le voisinage d'une singularité en utilisant le concept de déploiement universel.

Comme le dit Thom lui–même:

(k impair);

$$\pm x^k$$

(k pair);
et dans ce dernier cas le signe est déterminé par celui de $Df|_0$. [Poston and Stewart, 1978, page 58]

Il est difficile d'expliquer en deux mots de quoi il s'agit, même si, en un certain sens, ce n'est pas trop compliqué à comprendre. L'idée est celle–ci: quand on a un germe d'une fonction on peut toujours l'immerger dans une famille maximale. Ce germe analytique engendre une famille qui est la famille de toutes ses déformations. Donc, de par sa propre structure, il engendre quelque chose *qualitativement*. Le déploiement universel est tout simplement une manière de « déployer» toute l'information intrinsèque renfermée en une singularité. Selon moi, la singularité d'une application est toujours une chose qui concentre toute une structure globale en une structure locale. ... L'idée de déploiement universel contient en un certain sens la partie qualitative de la formule de Taylor: quand on a un germe de fonction différentiable, il y a localement un développement de Taylor et l'on peut légitimement se demander si en tronquant le développement de Taylor à un certain point, ce développement tronqué continue d'avoir la même allure, le même type topologique que le germe de fonction différentiable initial. Cette théorie des singularités permet de résoudre le problème posé par la nécessité de définir des conditions suffisantes pour que le germe tronqué jusqu'à l'ordre k soit équivalent, par un changement de variable, au germe de fonction initial. A l'évidence, il s'agit d'une manière intellectuellement très satisfaisante, dans la mesure où elle permet de réduire localement un germe, décrit par une série infinie, à une série que se compose d'un nombre fini de termes, ce qui représente une grande économie de pensée: un moyen unique pour décrire toutes les déformations possibles de ce germe.[Thom, 1983, page 28–29]

Il nous reste à étudier la dépendence de la paramétrisation en relation à t. La reparamétrisation de $(x_1, x_2, \ldots, x_n)$ à $(u_1, \ldots, u_n)$ qui dépend de t dans le voisinage d'une singularité non–Morse prend l'expression suivante:

$$f(u_1, \ldots, u_n; t) = u_1^3 + tu_1 \pm u_2^2 \pm \ldots \pm u_n^2$$

On peut voir que cette famille est structurellement stable puisque les entrecroisements avec la fonction nulle restent transversaux. Cela veut dire, aussi qu'une petite perturbation de ces fonctions ne change pas la typologie des points critiques.

Les points critiques de f_t sont donnés par la solution $m_+(t)$, $m_-(t)$ de $\frac{d}{du_1}(u_1^3 + tu_1) = 0$ Alors ces valeurs critiques pour f_t appartiennent à une des paraboles semicubiques

$$f(m_\pm(t)) = \pm\frac{4}{3\sqrt{3}}(-t)^{\frac{3}{2}}$$

Ces situations sont universelles et montrent toutes les façons dont une famille avec un seul paramètre passe à travers une singularité non–Morse; le cas qu'on a traité comme exemple ici est la catastrophe dite *fronce*.

En faisant une analyse dans l'esprit de celle–ci, Thom est arrivé à dresser une classification des catastrophes.

Nous pouvons maintenant énoncer ce théorème de Thom.

THÉORÈME 9
Soit $f : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ *une application différentiable* (C^∞). *Soit* $M_f \subset \mathbb{R}^{n+r}$ *l'ensemble des points* $(x_1, \ldots, x_n, u_1, \ldots, u_r)$ *où* $\frac{\partial f}{\partial x_1} = \ldots = \frac{\partial f}{\partial x_n} = 0$. *Soit* $X_f : M_f \to \mathbb{R}^{n+r} \to \mathbb{R}^r$ *l'application induite par la projection* $\Pi_r : \mathbb{R}^{n+r} \to \mathbb{R}^r$, $(x_1, \ldots, x_n, u_1, \ldots, u_r) \longmapsto (u_1, \ldots, u_r)$. *On appelle* X_f *l'application catastrophe. Soit d'autre part* $\mathfrak{F}$ *l'espace de fonctions lisses de* $\mathbb{R}^{n+r}$ *dans* $\mathbb{R}$, *muni de la topologie* C^∞ *de Whitney.* [39] *Si* $r \leq 5$, *il existe un sous–ensemble ouvert* $\mathfrak{F}_*$ *de* $\mathfrak{F}$ *tell que l'on appelle ses éléments* fonctions génériques. *Si* f *est générique, alors*

1. M_f *est une variété de dimension* r;
2. *toute singularité de* X_f *est équivalente à une* catastrophe élémentaire; *il n'y a qu'un nombre fini de catastrophes élémentaires;*
3. X_f *est localement stable en tout point de* M_f. *Le nombre de catastrophes élémentaires ne dépend que de* r:

[39] Dans l'espace de germes $E(q)$ des fonctions $\mathbb{R}^n \to \mathbb{R}$ indéfiniment différentiables, c'est–à–dire C^∞, en 0. Cet espace peut être doté d'une topologie, dite topologie de Whitney adaptée au niveau différentiable, à savoir la topologie de la convergence uniforme d'un élément F de $E(q)$ et de toutes ses dérivées sur les compacts de R^q.

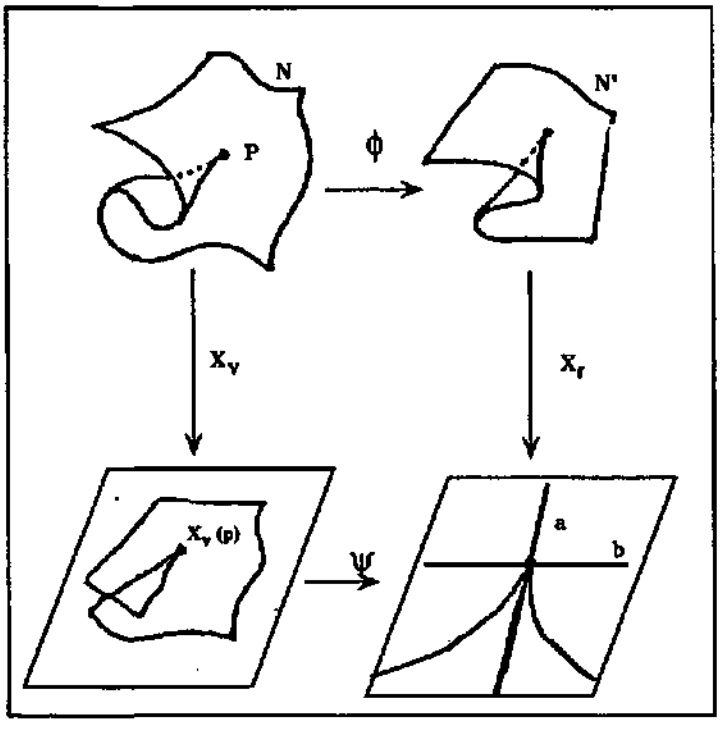

Figure 9.17: Représentation du voisinage de la singularité P dans l'exemple de la thermodynamique

r	1	2	3	4	5	6
cat. élément	1	2	5	7	11	∞

Thom explique admirablement l'idée de ce théorème de la façon suivante :

> Les espaces que l'on considère généralement sont des espaces homogènes, localement homogènes. Ces espaces sont ce que nous appelons des variétés. L'espace euclidien est une variété. Mais les singularités apparaissent lorsque l'on soumet en quelque sorte l'espace à une contrainte. La manche de ma veste, si je la comprime, fait apparaître des plis. C'est une situation générale. Cela ne relève pas de la mécanique des matériaux. J'énonce en réalité un théorème abstrait : lorsqu'on le projette sur quelque chose de plus petit que sa propre dimension, il accepte la contrainte, sauf en un certain nombre de points où il concentre, si l'on peut dire, toute son individualité première. Et c'est dans la présence de ces singularités que se fait la résistance. Le concept de singularité, c'est le moyen de subsumer en un point toute une structure globale. [Thom, 1991, page 23]

Lorsqu'on regarde de plus près la démarche que l'on a suivie dans l'analyse de notre exemple de la thermodynamique, nous observons que c'est justement une démarche qui nous amène à trouver les diagrammes commutatifs de la figure 9.22. Il est maintenant plus clair de comprendre les bases mathématiques de ce développement que j'ai voulu introduire de façon heuristique pour motiver l'importance du modèle de catastrophes. Nous avons décrit $F : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}$, C^∞ à l'origine O et qui constitue le déploiement d'un germe $\eta(x) = x^3$ de fonction $\mathbb{R}$ en $\mathbb{R}$ et dont le déploiement est l'équation 9.17 $F(v, p; T)$.

Le concept de germe sert à Thom pour étudier jusqu'à quel point des petites perturbations causées par les variables de contrôle génèrent des grandes perturbations dans la fonction F.

La figure 9.17 montre bien l'idée maîtresse de la théorie de catastrophes :

> Un modèle catastrophique n'est pas fondé sur des équations; ce sont des équations sur lesquelles on se permet un certain nombre de déformations : des changements de variables, en particulier, des perturbations, des déformations. C'est ce qui reste invariant, en somme, lorsque l'on fait des perturbations, qui est contenu solide de la théorie des catastrophes. Or, ce contenu solide est qualitatif et pas quantitatif. [Thom, 1991, page 42]

Une fois les outils mathématiques présentés, nous pouvons revenir aux modèles de la morphodynamique.

9.2.7 Les spécifications du modèle morphodynamique

Jusqu'à maintenant j'ai présenté le modèle morphodynamique général. Néanmoins ce modèle admet trois spécifications. Premièrement le concept de stabilité structurelle, deuxièmement la théorie des catastrophes généralisée et finalement la théorie des catastrophes élémentaires.

Le Concept de stabilité structurelle

La supposition de base pour ce premier modèle est que les processus internes X_w ont une *topologie* T naturelle. Les types qualitatifs que l'on observe dans le système S déterminent en χ une relation d'équivalence. Soit $\tilde{X}$ la classe d'équivalence de X pour un type qualitatif. La variation de X appartenant à la même classe n'entraîne pas de changement qualitatif, voilà pourquoi Petitot dit que:

> [...] la variation dans une classe d'équivalence $\tilde{X}$ *se réduit à l'identité.* Il n'y a donc de variation qualitative que lorsqu'une déformation dans $\tilde{\chi}$ *fait changer de classe d'équivalence.* La variation se manifeste, alors souvent par une *discontinuité* de la valeur de certains invariants ... [Petitot, 1992, page 5]

Petitot affirme que cette démarche permet une catégorisation de l'espace généralisé χ en assimilant la paire (classe d'équivalence $\tilde{X}$, élément de χ) à la paire scolastique de («genre», «espèce»).

La richesse du modèle réside en ce que pour chaque système $S = (W, \chi, \sigma, I)$ les processus internes X_w ne sont pas considérés comme des entités isolées. La morphodynamique introduit un double déplacement de point de vue.

> D'abord, elle considère comme objet d'étude non seulement les processus X_w mais également les *familles paramètres* $(X_w)_{w \in W}$m et cela en se focalisant sur la géométrie des ensembles catastrophiques K_w induits dans les espaces externes W par la déstabilisation des états internes définis par X_w. Ensuite, et surtout, elle considère ces familles comme l'image de champs $\sigma : W \to \chi$ « plongeant» l'espace externe W qui, en général, sera un morceau d'espace standard $\mathbf{R}^n$ dans l'espace généralisé χ. [Petitot, 1992, page 5]

Cette dernière caractéristique est centrale pour Petitot car elle permet l'expression des conflits internes par les *morphologies externes* en utilisant une dialectique du type *variation / invariant.*

DEFINITION 38 (PROCESSUS INTERNE STRUCTURELLEMENT STABLE)
Soit $X \in \chi$. On dit que X est structurellement stable si tout Y assez voisin de X (dans la topologie qui caractérise S) est équivalent à X.

Soit K_χ le sous-ensemble fermé de χ constitué des $X \in \chi$ *structurellement instables.* K_χ est une sorte d'ensemble catastrophique *intrinsèque.* Alors K_χ permet d'établir une morphologie discriminante qui caractérise et classifie les types qualitatifs de ses éléments structurellement stables.

> Autrement dit, K_χ *géométrise* la classification interne à χ. [Petitot, 1992, page 6]

En effet, un des résultats de cette géometrisation est la détermination des traces de K_χ sur W.

DEFINITION 39 (TRACE)
Soit $\sigma : W \to \chi$ le champ caractéristique d'un système $S = (W, \chi, \sigma, I)$. Soit $K'_W = \sigma^{-1}(K_\chi \cap \sigma(W))$ la trace *de K_χ sur W par l'intermédiaire de σ.*

> L'hypothèse de la modélisation est que l'ensemble catastrophique K_W de S est *déductible* de K'_W à partir de l'instance de sélection I. Elle signifie qu'une valeur w du contrôle appartient à K_W (*i.e.* est une valeur critique) si et seulement si la situation en w est en corrélation de la façon réglée par I à une situation appartenant à K'_W.[Petitot, 1992, page 7]

Les centre de l'analyse sont les ensembles catastrophiques intrinsèques. Si le champ σ s'avère structurellement stable, alors il est possible d'accéder à une classification des structures locales des K'_W et de cette façon aux morphologies externes.

Selon Petitot toute cette démarche a des racines foncièrement philosophiques :

> [...] l'*apparaître* d'un phénomène (*i.e.* la morphologie externe) est essentiellement le *contour apparent* sur un substrat de la forme décrivant son *être* (*i.e.* du processus interne). [Petitot, 1992, page 7]

Cette dernière considération est très importante. Je vais démontrer par la suite qu'elle est à la base de l'amalgame que Petitot fait entre les modèles morphodynamiques et la *géométrie du vécu* husserlien.

La théorie de système dynamique et la théorie de catastrophes généralisées

La théorie dynamique de système considère que le processus interne X est un système dynamique différentiable sur l'espace M de paramètres internes et qui caractérisent le système S considéré. Rappelons nous que W, à la différence de M est l'espace externe.

Chaque état instantané de S est décrit par un nombre fini de paramètres internes $x_1, \ldots, x_n$ qui parcourent un variété différentiable.

Prenons un exemple de la mécanique classique emprunté à [Petitot, 1992, cf. page 7–8]. Un état instantané d'un système de N points matériels est décrit par $6N$ coordonnées : les trois premières étant les coordonnées spatiales de chaque point et les trois suivantes, les composantes de la vitesse de chaque point. Toutes les coordonnées ne sont pas forcément cartésiennes, certaines peuvent être angulaires et décrire des cercles .

Un système dynamique X sur M consiste à assigner à chaque point x de M un vecteur de vitesse $X(x)$ qui sera tangent à M en x; ce vecteur varie différentiablement avec x. Or X est un champ de vecteurs différentiables sur M. En termes de coordonnées locales, il en résulte un système d'équations différentielles ordinaires:

$$dx_i/dt = f_i(x_1, \ldots, x_n)$$

où les composantes du champ f_i sont des fonctions différentiables des x_j et t est le paramètre temporel.

L'intégration d'un tel champ donne comme résultat les *courbes différentiables* qui sont les trajectoires paramétrées par le temps t dans M

$$\gamma : \mathbb{R} \to M$$

$$t \longmapsto \gamma(t) = (x_1(t), \ldots, x_n(t))$$

qui admettent en chaque point pour vecteur de vitesse $dx/dt = d\gamma(t)/dt$ le vecteur du champ $X(x) = X([\gamma(t)]$

DEFINITION 40 (SYSTÈME DYNAMIQUE)
Dans ce cas, on dira que le système est dynamique si:

1. *si les trajectoires sont intégrables sur un temps infini.* [40]
2. *si par tout point il ne passe qu'une et une seule trajectoire.* [41]
3. *Les trajectoires varient différentiablement en fonction des conditions initiales.[Petitot, 1992, page 8]*

Comme conséquence de la première condition de la définition 40 on peut définir l'application φ_t telle que: $\varphi_t : M \to M$ qui associe à tout point x de M le point au temps t de la trajectoire issue de x au temps $t = 0$. Alors, φ_t s'avère être un difféomorphisme de M.

[40] Cette condition élimine, par exemple la possibilité pour les trajectoires de sortir de M.

[41] Cette condition est en relation avec le principe de déterminisme selon lequel la condition initiale $x(0)$ au temps $t = 0$ détermine univoquement l'évolution future $x(t)$ pout $t > 0$ et l'évolution passée $x(t)$ pour $t < 0$.

DEFINITION 41 (FLOT DU SYSTÈME)
Soit l'application $\varphi : \mathbb{R} \to DiffM$, c'est-à-dire, $t \longmapsto \varphi_t$ qui va du groupe de réel additif au groupe de difféomorphismes de M et qui est un morphisme de groupe[42] *φ s'appelle le* flot *du sytème et il est la version intégrale du champ de vecteurs X.*

Malgré le fait que du point de vue mathématique la théorie ait atteint une certaine maturité, les modèles de la théorie de catastrophes présentent des difficultés lors de leurs applications. Deux caractéristiques de ces modèles sont à l'origine des ces difficultés:

- **Dynamique lente et dynamique rapide:** René Thom a observé que les modèles dynamiques sont soumis à deux types de dynamiques selon deux échelles de temps différentes: *la dynamique lente et la dynamique rapide.* Il a observé que:

  > L'idée philosophique essentielle sous-jacente à la théorie des catastrophes est que tout phénomène, toute forme spatio-temporelle doit son origine à une distinction qualitative des modes d'action du *temps* dans les choses. Toute distinction d'apparence qualitative dans un espace W (le substrat), peut être attribuée à deux modes d'action du temps: un mode « rapide » qui crée dans un espace interne des « attracteurs » qui spécifient la *qualité* phénoménologique locale du substrat; et un mode « lent » agissant dans l'espace substrat W lui-même. [Petitot, 1992, page 8-9]

  Ainsi la dynamique interne des états instantanés est *infiniment* rapide par rapport aux dynamiques externes d'évolution dans l'espace externe W. Alors ce qui compte pour l'analyse ce sont les états asymptotiques ($t \to +\infty$) définis par X_w. L'analyse de ces états présente une difficulté inattendue et redoutable.

- **Déterminisme mathématique et prédictibilité:** Le déterminisme mathématique n'implique pas la prédictibilité. Le problème dans les systèmes dynamiques réside en leur formidable complexité. Notamment, la définition de la condition initiale ne peut être faite qu'approximativement. En effet, le point x_0 qui représente la condition initiale n'est pas un point sinon qu'il appartient à un domaine *épaississant* que l'on appelle U. Pour qu'un système dynamique soit déterministe il faut que les trajectoires γ issues des x_0 soient stables. Il ne faut pas, néanmoins confondre la stabilité des trajectoires avec la stabilité structurelle du système. Les trajectoires sont stables si le domaine U où évoluent ces trajectoires est un petit tube *épaississant.* Techniquement on dira que la trajectoire de γ est relativement stable relativement à des petites perturbations de sa condition intiale.[43]

La notion d'attracteur et de bassin sont à la base de la description du système. Cependant il s'agit de concepts pour ainsi dire délicats puisqu'ils généralisent la notion d'équilibre stable. Les attracteurs de X_w sont des états internes de S et il est lui même un régime asymptotique stable.

DEFINITION 42 (ATTRACTEUR ET BASSIN)
*Une attracteur A de X est un ensemble fermé, X-invariant et indécomposable pour ces deux propriétés (*i.e. *minimal) qui attire voire qui capture toutes les trajectoires issues des points d'un de ses voisinages. Le plus grond voisinage de A qui a cette propriété s'appelle* bassin *de A et on le note $B(A)$*

Dans les cas les plus simples les attracteurs auront une structure topologique aussi simple; le nombre des attracteurs sera fini et leurs bassins seront des formes simples séparées par des frontières de délimitation mais ceci est rarement le cas. En général le nombre des attracteurs est infini, les bassins sont imbriqués les uns dans les autres et les attracteurs ont des topologies très compliquées.

Malgré toutes ces difficultés on suppose que l'hypothèse d'équilibre local est valable: la dynamique interne *rapide* entraîne le système vers un régime esymptotique stable correspondant à un état interne.

[42] En effet, il suffit de voir que φ_0 est l'identité de M, que $\varphi_{t+t'} = \varphi_t \circ \varphi'_t$ et que $\varphi_{-t} = (\varphi)^{-1}$.

[43] Je laisse de côté les considérations de Thom sur l'indéterminisme concret et sur le déterminisme mathématique. Pour une intéressante discussion du problème voir [Pacherie, 1992].

Je vais à présent présenter la troisième spécification des modèles morphodynamiques. Face à la complexité du modèle général, Thom a proposé une simplification : la théorie des catastrophes simplifiées.

La théorie des catastrophes élémentaires

La simplification porte sur le niveau d'analyse de la topologie des attracteurs, étant donné que la structure que l'on cherche le modèle général est très fine. Il s'agit de se restreindre aux ensembles catastrophiques K_w qui sont, en général plus simples que ceux induits par les bifurcations de système dynamiques généraux.

L'étude se fonde sur l'analyse des gradients. L'idée intuitive est d'étudier non la topologie mais les variations des champs sur leurs variétés. Cette simplification est la cible des critiques. Thom lui-même les reconnaît :

> Une dynamique de gradients est une dynamique très spéciale. Une dynamique de points pesants jetés dans l'espace, ce n'est pas une dynamique de gradients. Lorsque l'on jette un corps solide, pesant, dans l'espace, son énergie est composée de deux parties : l'énergie potentielle et l'énergie cinétique. La première donne effectivement lieu à une dynamique de gradients, par dissipativité. L'énergie cinétique non. Elle donne plutôt lieu à ce que l'on appelle une perte d'énergie par dissipation. L'aspect des trajectoires en est notamment modifié. Selon que l'on adopte une hypothèse où il y a une sorte de frottement infini (c'est au fond l'hypothèse de la physique aristotélicienne) ou une hypothèse sans frottement, on bien les choses se passent absolument librement, sans déperdition d'énergie ou bien il peut y avoir une certaine dégradation. Cette remarque peut être faite pour le frottement: il y a aussi conservation de l'énergie. Mais comme on ne s'intéresse pas à l'énergie thermique, la chaleur qui sort de ce frottement est négligée. L'énergie que l'on appelle libre, donc diminue.
>
> Il y donc une objection. [Thom, 1991, page 41]

Petitot résume la démarche comme suit :

> L'idée est de tenter de généraliser aux systèmes généraux ce qui se passe dans le cas des *systèmes de gradient*, à savoir l'existence de lignes de pente et de variétés de niveau. Pour cela, on utilise le fait que, si A est un attracteur d'un système dynamique X sur une variété M, on peut construire sur le bassin $B(A)$ de A une fonction positive f –dite *fonction de Liapounov* – qui décroît strictement sur les trajectoires dans $B(A) - A$ et qui s'annule sur A. Cette fonction est une sorte d'entropie locale exprimant que, au cours du temps, $B(A)$ se « contracte » sur A de façon analogue à un système de gradient. Mais elle ne permet de rien dire sur la structure interne de l'attracteur. [Petitot, 1992, page 11]

En définitive, il s'agit de faire la supposition que la dynamique interne X_w est en fait la dynamique de gradient associée à une fonction potentielle différentiable $f_w : M \to \mathbb{R}$. Les états internes déterminés par f_w sont alors ses *minima*; cela veut dire que si f est assimilée à une *énergie* alors la supposition que l'on a fait correspond au principe de minimisation de l'énergie. Une des causes majeures d'instabilité structurelle est l'existence de *singularités*, d'où les mathématiques utilisées qui visent à caractériser l'entourage de ce type de points.

Je voudrais avancer que l'utilisation des gradients pour l'étude des propriétés phénoménologiques perceptives a été originairement proposé par James Jerome Gibson en 1952.

Dans [Gibson and Dibble, 1952] il arrive à postuler que les impressions élémentaires comme les surfaces et les contours, au contraire de ce qu'a postulé la *Gestalttheorie* ne suffisent pas pour expliquer les cas où la réalité est très complexe, c'est-à-dire quand les surfaces ont différents caractères de dureté, de distance et de couleur. Il faut changer la façon d'analyser les choses et Gibson a commencé à mettre l'accent sur le rôle de la lumière dans l'image rétinienne plutôt que sur l'objet-lui même.[44]

[44] Ainsi, Gibson arrive ainsi à formuler l'hypothèse de la texture pour la perception d'une surface. Il postule qu'il y a perception d'une surface quand les gradients d'intensité lumineuse dans l'image, entre des petites régions d'intensités différentes, ont une pente maximale. En clair, pour la détermination d'une surface, le concept de contour est remplacé par l'hypothèse de la texture, qui ne prend pas en compte l'objet mais l'intensité de gradient de lumière.
Quelle est la définition de gradient?
Gibson définit le concept de gradient de la façon suivante :

> The word gradient means nothing more complex than an increase or decrease of something along a given axis or dimension ... [Gibson, 1950, page 73].

9.3 La porté philosophique du modèle

L'emergence de propriétes du macroniveau

De tout ce que l'on vient d'exposer, il découle que les modèles morpbodynamiques sont des candidats de choix pour combler le fossé existant entre la réalité et les représentations dont on a parlé au début de ce chapitre.

Premièrement, les changements qualitatifs nous permettent la détermination ou la délimitation des zones de stabilité qui sont constituées par la permanence d'une qualité donnée et dont les limites extériorisent des changements brusques au oiveau de la microstructure.

Secondo, les bypothèses émises daos la définition 18 de système nous permettent d'affirmer que les aspects qualitatifs du macroniveau sont des propriétés *émergentes* du niveau de microscopique. [45]

L'idée générale est bien résumée par Smitb et Petitot comme suit

> Intuitively speaking we say that 'qualitative' structures exist where certain fine-grained microstructures are just sufficiently smooth to admit a coarse-grained morphological organization via discontinuities (boudaries) on the macroscopic level.[Petitot and Smith, 1994]

Dans le chapitre précédent j'ai exposé les différentes notions du concept d'émergence, mais quelle est la conception de Petitot à cette égard?

Une propriété qualitative est émergente *en relation* avec un comportement physique du substrat matériel sur une structure si elle vérifie les trois propriétés suivantes:

1. There must be two levels of reality, a microlevel ond a macrolevel, and the emergent property needs to be a property of objects on the macrolevel.
2. Objects ou the macrolevel must be made up of objects on the microlevel and their parts, so that we must be able to explain causally the emergent structure exclusively by appealing to phenomena on the microlevel (causal reductionism).
3. But on the other hand we must be able to show that there are holistic and structural or organizational features (morphological properties, properties of self-maintenance, etc.) which are distinct from those structures or organiestional features whith are proper to the microlevel sciences. [Petitot and Smith, 1994]

Tout d'abord, le concept de propriété émergente pour ces auteurs est relatif aux comportements du niveau inférieur et il doit vérifier trois caractéristiques que l'on pourrait résumer comme : dualité des niveaux, réductionisme causal du macroniveau au microniveau, existence d'une organisation ou des structures propres au niveau macroscopique.

Cette définition de l'émergence est bien plus claire que les réflexions de Petitot concernant la relation entre les differents oiveaux dans son livre *La physique du sens*. Dans ce texte [Petitot, 1992] il semble récuser le réductionisme causal qu'il professe pourtant dans un texte postérieur cosigné avec Barry Smith.

> ... on peut admettre que chaque niveau d'organisation possède *une autonomie et une syntaxe propre* : le rapport entre deux niveaux n'est pas , comme le point de vue réductioniste l'affirme, un rapport *causal de dépendance unilatérale* allant du niveau physique de base à celui de la manisfestation morphologique, mais bien un rapport *expressif de dépendance bilatérale*. René Thom a toujours beaucoup insisté sur ce point dans sa dénonciation de la « prétention» réductioniste d'expliquer par le niveau le plus fin «les formations apparues aux niveaux plus grossiers». [Petitot, 1992, page 16]

Puis il dit que le gradient peut changer abruptement et former des marches (*steps*), et ajoute :

> These concepts appear to be admirably adapted for describing the retinal Image, since both gradients and steps of stimulation can be found within it.[Gibson, 1950, page 73]

Il affirme qu'il existe deux types de gradients, suivant la surface qu'ils occupent sur la rétine :

micrograclients : se référant à des contours et à des textures visuelles;

macrogradients : pistes pour la profondeur et l'inclinaison de la surface.

Pour l'étude de l'évolution du concept des gradients dans les sciences cognitives voir [Scaglione, 1991].

[45] Cette conception des qualités comme des qualités émergentes reprend de façon scientifique l'idée de Samuel Alexander selon laquelle l'apparition des nouvelles propriétés correspond à l'existence des certaines constellations ou colocations du niveau inférieur. [Alexander, 1920]

Le Petitot (en communication personelle) affirme que les deux définitions sont équivalentes.

Néanmoins, je dois reconnaître que dans le texte qu'il a rédigé en collaboration avec Barry Smith son opposition au réductionisme causal semble plus nuancée. Il semble accepter le réductionnisme causal de macroniveau au microniveau sans réserve.

C'est seulement dans la troisième condition qu'il fait état de l'appartenance de chacun de ces niveaux à des domaines ontologiques différents et de l'impossibilité de conclure par une réduction ontologique à partir d'une réduction causale ou épistémologique.

Maintenant, j'aimerais expliquer les caractéristiques des modèles morphodynamiques qui les font apparaître si aptes pour l'explication de l'émergence des macropropriétés, mis à part celles que j'ai déjà exposées. La richesse réside dans la double voie méthodologique que ces modèles nous octroient.

Comme le Petitot l'affirme, ces deux voies méthodologiques sont, en réalité des stratégies inverses.

Première voie: l'*être* physique détermine causalement l'*apparaître morphologique* Cette voie, qui est plus en relation avec la tradition positiviste, c'est la voie *physicienne*. Toutes les applications que l'on a faites du modèle morphodynamique à des applications physiques rigoureuses et exactes en sont partie. Quelques exemples: catastrophes de diffraction et dislocations de fronts d'ondes en optique ondulatoire, théorie des transitions de phases et phénomènes de rupture spontanée de symétrie dans les milieux ordonnés; stabilité de défauts dans les phases mésomorphes.

La dynamique de cette voie va de l'intérieur vers l'extérieur. Ce sont les lois de la physique qui gouvernent la dynamique interne du système; en appliquant des méthodes mathématiques on cherche à dériver mathématiquement les ensembles catastrophiques K_w de la connaissance explicite des champs $\sigma : W \to \chi$. En ce faisant nous pouvons reconnaître ded zones de stabilité dont l'ensemble met en évidence l'*apparaître morphologique*.

La seconde voie: l'*apparaître* comme déterminant pour l'*être* A mon avis, la première voie légalise la seconde. Si l'on peut dériver causalement la morphologie externe de la dynamique interne, rien ne nous empêche de prendre la stratégie inverse: à partir des morphologies externes faire des déductions sur la dynamique interne, même si ces dynamiques que nous proposons ne sont pas que de complexité minimale.

Petitot appelle cette seconde stratégie la voie *morphologique–structurelle*. Elle est vraiment utile dans le cas des boîtes noires « vraiment noir » selon la qualification de Petitot.

> En particulier, elle conduit à cherche un dynamique interne Y_w pour la morphologie observée K_w qui soit de *complexité minimale*. On pourra alors dire que la dynamique interne « réelle » (inobservable) X_w est une complexification de Y_w qui est phénoménologiquement non pertinente. Cette *réduction* (en général drastique) de la complexité de la dynamique interne correspond, ... à un changement du niveau d'organisation et d'observation, Elle signifie que les dynamiques internes sont en général hautement *surdéterminées* relativement aux morphologies externes. [Petitot, 1992, page 16]

C'est très important de remarquer que cette seconde voie est possible non seulement en vertu de la modélisation mathématique mais aussi par le fait que le niveau morphologique a un *ordre de légalité propre et autonome*.[46]

Lorsque Petitot parle de l'existence d'un ontologie propre il veut dire l'existence de structures spécifiques à chaque niveau. Selon le cadre théorique ou explicatif où on se place on peut choisir le niveau d'explication avec la granularité pertinente.

La seconde voie dite *morphologique–structurelle* permet de faire des inférences sur les phénomènes du microniveau, mais ces caractérisations ont une valeur descriptive plus qu'explicative ou causale.

En observant les propriétés émergentes, on peut postuler des hypothèses sur le niveau des microstructures et à partir de ces hypothèse préfigurer une explication causale (du microniveau

[46] Pour des exemples de cette voie voir dans [Petitot and Smith, 1994] un compte rendu étendu par les sujets qu'il évoque, entre autres, l'application des modèles morphodynamiques à la géometrie de la perception.

au macroniveau) *a posteriori*. La validation de cette démarche sera faite par la cobérence entre les phénomènes observés au macroniveau et les lois de la physique du microniveau que l'on considère comme intervenantes. Je pense que c'est dans ce cadre que l'on doit comprendre le rapport entre les deux niveaux que Petitot caractérise d'*expressive de dépendance unilatérale*.

La première voie, dite *physicienne* comporte des explications par les *causes* des phénomènes du macroniveau à partir de ceux du microniveau; or elle est bel et bien une réduction causale ou épistémologique. Cependant, on n'est nullement obligé d'accepter la réduction ontologique; la preuve est que l'on accepte la seconde voie comme une méthode valable de recherche, dans le pire des cas heuristique. Petitot semble être d'accord avec la position de Bunge. Oui à la réduction épistémique, non à la réduction ontologique.

Peut-on comptabiliser le concept de propriété émergente de Petitot et Smith [47] avec la théorie de Bunge que j'ai exposée dans le chapitre précédent.

La réponse à mon avis est positive; non seulement elle est possible sinon qu'elle est soubaitable.

Tout d'abord je montrerai pourquoi il n'y pas point de contradictioo entre les deux positions, en suite je signalerai les bénéfices que l'on peut tirer de cette compatibilisation.

L'émergence telle que Petitot et Smith la comprennent est, en effet, compatible avec le caractère nomologique de ces occurreoces, l'existeoce d'une explicatioo ou réduction épistémologique, sans que cela entraîne necessairement une réduction ontologique et l'existence des variété ou pluralisme des facteurs causaux. Ces trois dernières caractéristiques étant centrales dans l'approche bungèenne.

De l'autre côté, il y a l'existence des niveaux indépendants. Dans le cas de Bunge ils sont plusieurs, dans le cas de Petitot et Smith ils sont seulement deux. Néanmoins on peut dire que ce ne sont pas des vues incompatibles; je pense que l'on peut attribuer au fait que les phénomènes qui intéressent Petitot et Smith appartiennent à une zone plus restreinte mais comprise daos la région considérée chez Bunge.

Les bénéfices que l'on peut retirer d'une compatibilisation des deux positions consiste en la précisioo dans la détermination de la relation totalité-partie que les modèles morphodynamiques peuvent apporter à la théorie bungénne.

En effet, la catégorisation des zones de stabilité autour des attracteurs et la dynamique dont ces mêmes attracteurs sont issues donnent non seulement une catégorisation des parties de stabilité sinon une structuration entre elles. Dans la mesure où cette structuratioo se base sur une type de dynamique que l'on décrit à l'intérieur d'une théorie, il découle qu'il ne s'agit pas d'une simple description d'inclusion, sioon qu'elle s'avère un outil pour trouver cette relation *suis generis* à laquelle Bunge fait référence. En outre, l'attachement à une théorie donnée est cohérent avec le fait que les propriétés émergeotes sont des propriétés relatives à cette théorie comme Bunge le signale.

Finalement, la comptabilisatioo des deux approches sert à mettre au clair le rôle intuitif que Francisco Varela entre autres donne aux attracteurs dans le concept d'émergence et que j'ai cité au début du chapitre précédent. Probablement une compatibilisation de deux théories nous amènera à une concept d'émergence unifié avec la possibilité d'applications plus au moins immédiates et transversales à différents domaines.

9.3.1 Les modèles morphodynamiques et la réduction eidetique

Le terme *eidétique* fait référence aux essence pures. Dans *Ideen III* §3 il dit

> *D'abord* le mot « essence» a désigné ce que dans l'être le plus intime d'un individu se présente comme son « Quid» (sein Was). Or ce Quid peut toujours être « *posé en idée*». *L'intuition empirique (erfahrende)* ou *intuition de l'individu* peut être convertie en *vision de l'essence* (Wesens-Schauung) en *idéation* -cette possibilité devant elle-même être entendue non comme possibilité empirique mais comme possibilité sur le plan des essences. Le terme et la vision est alors l'essence *pure* correspondante ou Eidos, que ce soit la catégorie de degré supérieur ou une forme plus particulière, en descendant jusqu'à l'ultime concret. [Husserl, 1950, page 19-21]

[47] Bien que le concept de pmpriété émergente dans le cas de ces auteurs n'a pas donné lieu, pour le moment, à une théorie complète ou exhaustive de l'émergence je vais me référer aux caractéristiques déjà énoncées.

C'est une chose que d'affirmer que les systèmes morphodynamiques sont des outils excellents pour expliquer l'émergence des propriétés du macroniveau à partir de celles du microniveau et c'est une autre que de dire qu'ils sont capables de servir de fondements à l'*être ultime des choses*. Petitot défend aussi ce second fait. Il met sur un pied d'égalité ces modèles et la réduction eidétique husserlienne.

Je vais en ce qui suit, d'abord, développer ce dernier point pour ensuite exprimer mes réserves à cet égard.

La géométrie du vécu : La phénoménologie comme toute autre science eidétique a un caractère descriptif. La phénoménologie qui traite de vécus ayant subi la réduction phénoménologique et des corrélations qui appartiennent à leurs essences opère sur des *simples présentifications*[48] portant sur des exemples individuels.

Mais la question que l'on se pose est la suivante : peut-on proposer un eidétique descriptive, autrement dit, une *géométrie du vécu*?

Proposer une eidétique descriptive revient au même que proposer une mathématique de ces phénomènes. Or, Husserl récuse cette possibilité pour des raisons différentes.

Une des raisons est la différence fondamentale entre les sciences eidétiques *concrètes* et *abstraites* dans la démarche husserlienne. Les science eidétiques, en effet comportent ces deux types différents ce qui correspond à la fois à la division entre *genres abstraits* et *genres concrets*. Toute science doit remonter au genre suprême pour trouver l'unité absolue. Pour y arriver elle doit, à partir de la région considérée comme objet d'étude trouver l'unité des régions composantes et éventuellement montrer comment ces régions se fondent les unes dans les autres. Il s'agit de trouver la structure du genre suprême. [cf. *Ideen* §72.134]

Les composantes du genre suprême peuvent avoir un caractère régional (concret) ou être des composantes du genre suprême : dans le premier cas il s'agira des sciences *concrètes*, dans le second de sciences *abstraites*. La phénoménologie sera une science eidétique concrète. [cf. *Ideen* §73]

La géométrie des vécus (science concrète) est différente de la géométrie euclidienne (science abstraite) qui forme partie des mathématiques. Cette dernière a pu surmonter le niveau *symbolique* des représentations grâce à son axiomatisation mais il n'en est pas de même pour la géométrie du vécu. La géométrie du vécu est faite de morphologiques aexactes, concrètes et singulières. Husserl soutient qu'il n'est pas possible de géométriser, ni de décrire conceptuellement les essences aexactes, on ne peut pas passer d'elles à des essences génériques, aux concepts univoques.

En suite, la différence que Husserl fait entre *abstraction* et *idéalisation*. Husserl oppose ces deux notions :

> Les concepts géométriques sont des concepts «idéaux»; il expriment quelque chose qu'on ne peut pas «voir»; leur « origine» [49], et donc aussi leur contenu diffèrent essentiellement de ceux des *concepts descriptifs* en tant que concepts exprimant des essences issues sans intermédiaire de la simple intuition, et nullement des essence «idéales». Les concepts exacts ont pour corrélat des essences qui ont le caractère *«d'idées» au sens kantien* du mot. A l'opposé de ces idées ou essences idéales nous trouvons les *essences morphologiques* qui sont les corrélats des concepts descriptifs.
>
> Cette idéation (*Idéation*) érige les essence idéales en *«limites» idéales* que l'on ne saurait par principe découvrir dans aucune intuition sensible et dont se «rapprochent» plus au moins, sans jamais les atteindre, les essences morphologiques considérées; cette idéation diffère fondamentalement de la saisie des essences a la simple «abstraction» [50], par laquelle un «moment» est détaché et élevé dans la région des essences comme quelque chose de vague par principe, de typique. Les *concepts génétiques*, ou les essences génériques, qui ont leur champ d'extension dans le fluent ont une *consistance* (*Festigkeit*) et une *aptitude aux distinctions pures que ne doivent par être confondues avec l'exactitude des concepts idéaux*, et des genres qui ont exclusivement des objets idéaux dans leur extension. Il faut bien voir en outre que les *sciences exactes* et les *sciences purement descriptives* ont bien entre elles une liaison, mais qu'elles ne peuvent jamais être prises l'une pour l'autre et quel que soit le développement d'une

[48] *Ideen* §70

[49] Le mot «origine» – comme plus loin le mot « abstraction» – est pris au sens de la psychologie génétique: extraction de l'expérience. (Note du traducteur)

[50] [...] Le §23 précise que «l'abstraction» ne produit pas l'essence mais la conscience de l'essence. C'est de cette « abstraction» psychologique, de ce passage à l'essence qu'il s'agit ici; elle porte donc sur toutes les essences inexactes, qu'elles soient concrètes ou abstraites. ...

science exacte, c'est-à-dire opérant avec des infrastructures idéales, elle ne peut résoudre le tâches originelles et autorisés d'une description pure. [Husserl, 1950, pages 236-237]

On voit bien qu'il semble très difficile pour Petitot d'affirmer que la modélisation mathématique morphodynamique *est* la réduction transcendantale et sauver toute la théorie de Husserl. La division que Husserl fait entre les sciences descriptives et sciences exactes ne se base pas seulement dans un aspect métbologique. Husserl a des raisons ontologiques que ne lui permettent pas d'accepter cet amalgame.

Cependant, Jean Petitot relativise ces arguments en signalant que la conception busserlienne du *formel* reste très restreinte à la lumière de la science actuelle. On pourrait caractériser cette conception comme « hilbertienne» : opérationnelle, structurale et axiomatique. Ensuite il soutient que Husserl serait tombé dans un confusion *majeure* et qu'il aurait négligé le fait que la géométrie avant d'être formalisée a dû repousser les essences morpbologiques.

[...] en définitive, l'idée d'une géométrie morphologique représente une *béance*, un vide centrale, « un manque» dans la conception husserlienne. C'est ce manque absolu qui constitue le ressort et le moteur implicite du parcours phénoménologique devient *constitutif*. Cette discussion nous montre qu'en cette affaire si décisive pour son projet, Husserl commet une *confusion majeure*. Il confond l'exactitude des idéalités mathématiques, qui est un caractère d'essence, avec le fait que, *pour se constituer*, la géométrie a dû originairement refouler les essences morphologiques vagues. *Or rien ne permet ici de conclure de l'origine à l'essence.* Rien n'interdit *a priori* à la géométrie de devenir un jour à même de se « se retourner» sur son « refoulé originaire» et *de se réapproprier l'univers des essences morphologiques singulières et fluentes.* [Petitot, 1992, page 90, les parties en italique appartiennent au texte original]

Petitot récuse la différence entre abstraction et idéalisation; il dit qu'elle n'est plus acceptable.

A mon avis Petitot a deux chemins possibles; soit il accepte que sa position est très proche de celle de Husserl, sans pour autant la respecter mot par mot, soit il accepte que la modélisation matbématique est *en un certain sens* une réduction eidétique ceci par exemple en acceptant que ces modélisations ne donnent pas la description ultime des choses sinon qu'elles peuvent servir à bâtir une ontologie des objets dans le cadre d'une conception naïve du monde extérieur.

Barry Smitb est partisan de cette position; il prône l'existence d'une ontologie des objets du sens commun, dont un des moyens pour la démontrer est la modélisation mathématique. (voir [Smitb, 1994]) Ce que je viens d'exposer marque le clivage entre Barry Smith et Jean Petitot. Selon le premier, il existera une ontologie des objets de vie de tous les jours, ce qui veut dire qu'il existe des structures invariantes qui servent à décrire le monde à partir de leurs formes et de leurs caractéristiques qualitatives. Selon Petitot, bien que ces structures invariantes existent bel et bien, elles ne font que confirmer et servir de moyen pour une description phénomènologique *eidétique* des formes.

9.4 Conclusion

J'ai commencé le chapitre en montrant le besoin légitime exprimé par certains chercheurs d'avoir un compte rendu de l'intentionnalité ou du renvoi intentionnel qui prend en compte les données phénoménales.

En suite j'ai décrit les modèles morphodynamiques comme étant les ponts entre les données qualitatives et les propriétés de microniveau telles qu'elles sont décrites par les physique scientifique.

Plus tard, j'a essayé d'exposer une description de la théorie des catastrophes pour mettre en évidence les deux caractéristiques suivantes: *primo* le but d'une telle théorie est de caractériser le singulier et l'universel à la fois et *secondo* les descriptions issues sont indépendantes du substrat.

Finalement, j'ai montré les deux voies de la morphodynamique : physicienne et morphologique–structurelle. La première qui va de l'*être* à l'*apparaître*, la seconde, plus audacieuse, qui soutient que l'*apparaître* est déterminant pour l'*être*. Cette seconde voie serait selon Petitot l'équivalent de la réduction eidétique de Husserl.

Dans l'évaluation que je me propose de faire tout à l'heure, je ne me prononcerai pas sur la valeur des modèles de la morphogènese et en particulier je ne vais pas traiter les limites de leur capacité prédictive.

Par contre, je vais formuler mes réserves en référence à l'utilisation de méthodes matbématiques comme outils aptes à la détermination de l'être ultime des choses. Aussi, je vais soutenir que bien que les méthodes morphodynamiques jouent un rôle important comme lien entre les macropropriétés et les micropropriétés, elles ne seront qu'un maillon dans l'explication de la chaîne de perception. Dans ce sens je vais signaler que la stratégie telle que Petitot la veut, est menacée par le prohlème humonculaire. Je vais soutenir deux argumentations pour donner un fondement critique à l'utilisation des méthodes mathématiques pour la détermination de l'être ultime des choses. La première argumentation fait référence à la possibilité d'utiliser des modèles mathématiques pour la description ontologique et je vais la discuter plus bas. La seconde vise le fait que la détermination des singularités autour desquelles les déploiements universels sont bâtis dans le but de décrire *la structure ultime* des choses sont hautement dépendants des systèmes perceptifs qui les perçoivent. Une chauve-souris, par exemple aurait une ontologie de choses différents à la nôtre (si toutefois vous me permettez de prendre la licence d'octroyer des capacités intentionnelles supérieurs à un animal). On voit mal comment elles serviraient pour décrire leur être ultime.

En relation à la première critique, je vais soutenir avec Mario Bunge les propositions suivantes : premièrement les mathématiques sont ontologiquement neutres, deuxièmenent les matbématiques comportent des modèles qui sont à la fois trop riches et aussi trop étroits pour pouvoir donner une description de l'ontologie des choses [Bunge, 1994, cf. page 169ss]. Bunge affirme, avec justesse, que les théories mathématiques ne rencontrent pas par elles-même le monde réel, d'où la neutralité de son ontologie.

Elles peuvent rencontrer les faits réels s'il s'agit des théories mathématiques interprétées. Une théorie mathématique M peut être dite *complètement interprétée en termes factuels* si à chaque concept primitif (non défini) de M on assigne un élément de l'ensemble F des réalités factuelles[51].

Néanmoins, les théories *complètement* interprétées sont très rares; dans la pluspart des cas, une foule de concepts mathématiques d'une théorie interprétée n'ont pas de correspondance avec des objets fartuels.[52]

Bunge synthétise bien cette idée.

> En somme, les théories, en mathématiques pures, « ne disent rien » de la réalité. C'est seulement en enrichissant un formalisme mathématique avec une fonction d'interprétation qu'elles peuvent devenir des théories factuelles. [Bunge, 1994, page 171]
>
> ...
>
> Les mathématiques sont trop pauvres en un sens, trop riches en un autre. Il manque quelques doigts à un gant trop grand. [Bunge, 1994, page 172]

Réciproquement, le fait que tous les concepts ou toutes les idées ayant un minimum de clarté puissent être exprimés en termes des formalisations matbématiques ne peut pas être considéré comme certaine. Nous avons que pour Husserl dans *Ideen* ce n'était pas le cas. Il n'est pas exclu qu'il existent des concepts importants qui soient réfractaires à ce type d'expression.

Derrière cette considération sur le rôle de la matbématique dans la quête philosophique se trouve un postulat meta-théorique importante : les constructions mathématiques sont considérées comme ayant une réalité indépendante du monde, elles ont une existence autonome. Ce postulat semble être accepté par Petitot et récusé par Bunge. Nous sommes de l'avis de ce dernier : les concepts comme attracteur, déploiement universel entre autres sont des concepts qui n'ont pas une existence autonome; il jouent seulement un certain rôle dans le cadre d'une théorie. Voilà pourquoi ils ne me semblent pas les plus aptes pour fonder l'ontologie des choses.

Cependant, même si l'on accorde à Petitot que les descriptions morphodynamiques révèlent les structures ultimes des choses, sa position est encore réfutable. L'interprétation qu'il fait de

[51] Un item factuel peut être une chose, une propriété, un état, un événement ou un processus.

[52] L'exemple de Bunge se réfère à la mécanique classique. En effet, la dérivée quatrième des coordonnées de position respectivement au temps peut apparaître dans un calcul sans que pour autant elle représente une propriété physique connue.

Husserl est une interprétation fregéenne; or la relation intentionnelle est conçue comme une simple juxtaposition, une relation dans le sens mathématique entre le sujet et l'objet intentionnel. Malgré le fait que les descriptions que les modèles font des propriétés qualitatives puissent être considèrées comme reflétant les propriétés ultimes des choses, il manque encore quelqu'un pour les interpréter et donner le pas du renvoi intentionnelle. Ceci est aussi une conséquence du caractère ontologiquement neutre de ces modèles.

Cette difficulté a été recontrée chez David Marr dans sa théorie de la vision; on a beau montrer les propriétés qualitatives en utilisant un décomposition mathématique, encore faut-il montrer comment le sujet arrive à la référence en partant des premières. [53] Cette dernière observation souligne le problème humonculaire dont j'ai fait référence au début de la conclusion.

À l'appui de ce type d'approche l'on peut dire qu'il s'agit d'une nouvelle démarche, que le programme est long et pas totalement défini à préseot. Néanmoins, si l'on se horne à l'application des modèles morphodynamiques comme outil pour faire le lien entre les propriétés phénoménales et les propriétes de microniveau, il s'avère être un programme porteur. C'est une voie qui vaut la peine d'être explorée tout en gardant un optimisme mesuré.

[53] J'ai déjà traité ce problème dans [Scaglione, 1991].

Partie IV

Conclusion

Conclusion 10

Emergence et réalisation : Deux côtés de la même pièce

Saint Augustin pensait à la manière de comprendre le myetère de la Saint Trinité au bord de la mer. Tout près de lui, il y avait un enfant qui amenait infatigablement de l'eau de mer dans le creux d'un tout petit coquillage dans un tout petit trou qu'il avait creusé dans le sable. Après quelque temps, Saint Augustin interpella l'enfant de la façon suivante.

– Que fais–tu, mon fils ?

– Je vais mettre toute l'eau de mer dans mon trou ? –répond l'enfant.

–Mais, mon enfant –répond Saint Augustin– ceci est impossible; regarde l'immensité de la mer et la petitesse de ton coquillage, tu n'arriveras jamais, même si tu avais une éternité pour accomplir la tâche que tu t'es proposé.

Alors l'enfant regarda Saint Augustin dans les yeux et lui dit

– Tu vois Augustin, de la même façon que je ne vais jamais arriver à mettre toute l'eau de la mer dans ce petit trou, tu ne vas jamais arriver à comprendre le mystère de la Sainte Trinité.

10.1 Introduction

Tout le long de ce travail j'ai essayé de retracer les efforts tendant à rendre une théorie de l'intentionnalité compatible avec le présupposé physicaliste non–réductionniste.

J'ai choisi de poser ce problème de deux manières différentes. La première était en termes du trilemme classique (chapitre 2 §2.2). Selon ce trilemme il s'agit de rendre compatibles le monisme de substance et le dualisme de propriétés (physiques et mentales) tout en octroyant aux propriétés mentales des pouvoirs causaux sans trahir le principe physique d'interaction causale. Je me suis mise dans les rangs de ceux qui tentent de rendre compatibles entre eux les trois termes du trilemme.

L'autre manière était de placer le problème dans le cadre de la théorie de Franz Brentano étant donné que c'est lui qui a lancé le défi connu comme la thèse qui porte son nom sur l'irréductibilité du psychique. L'étude de la conception des phénomènes psychiques et physiques de Brentano m'a donné l'occasion de remonter l'histoire des termes clés que l'on utilise dans la philosophie de l'esprit contemporaine comme ceux de "contenu" et de "représentation"(chapitres 3 et chapitre 4).

Ensuite j'ai présenté les stratégies visant à naturaliser l'intentionnalité, c'est à dire à donner une explication du renvoi intentionnel dans des termes autres que mentaux.

J'ai présenté les stratégies éliminativistes[1] (chapitre 2, §2.4.1.2 et §2.6.3) que j'ai ensuite récusées parce qu'elles vident les entités mentales de toute pertinence causale ou explicative et ne

[1]Le but des positions éliminativistes qui considèrent la psychologie ordinaire erronée est de rejeter toute plausibilité ou explication aux concepts qui en découlent. Elles arguent que la conception des états ou processus auxquels ces énoncés font référence n'est qu'une erreur ontologique, d'où le besoin de remplacer cette conception erronée par une autre qui soit correcte.

cherchent donc pas à rendre les termes du trilemme compatibles mais plutôt à réfuter la pertinence causale des états mentaux.

Subséquemment, j'ai présenté des alternatives visant à établir une corrélation forte entre les états mentaux et les états physiques.

Ces alternatives reposent sur les stratégies suivantes: l'identification entre types mentaux et types physiques (chapitre 2 §2.4.1.1), la survenance entre des états/propriétés physiques (chapitre 2 §2.4.1.2) sur le mental et la réalisation des propriétés mentales par le niveau physique (chapitre 2 §2.4.1.3).

Ces stratégies sont en principe valables, mais on ne peut se prononcer sur leur réussite avant d'avoir éclairci les critères d'individualisation des états ou propriétés sur lesquelles on entend fonder ladite corrélation. Après, j'ai montré l'usage que l'on en a fait dans les théories fonctionnalistes et j'ai aussi signalé leurs limites. Ces théories sont des stratégies du type *top-down* car elles partent des entités intentionnelles et elles essaient d'établir la corrélation des propriétés ou états mentaux avec les états physiques.

Le fonctionnalisme de *types* plaide pour une identification *a posteriori* (chapitre 7) des types mentaux avec des types physiques, les premiers étant individualisés en termes de leur rôles fonctionnels. On fait à l'égard de cette théorie deux critiques différentes. La première qui met en évidence le caractère réductionniste de l'approche est réfutable si l'on est d'accord pour établir une différence entre les concepts de réduction épistémologique et de réduction ontologique (chapitre 2 §2.4.1). En effet, j'ai soutenu à plusieurs reprises au cours de ce travail que la réduction épistémologique n'amène pas forcément à une réduction ontologique. La seconde critique repose sur le fait que le physicalisme de type exclut la possibilité de multiréalisation de la cognition. Le monisme anomal de Donald Davidson (chapitre 6) qui utilise la stratégie de la survenance constitue une autre tentative dont j'ai fait la description. Il est non-réductionniste d'emblée parce qu'il nie l'existence des lois psychophysiques. J'ai montré que malgré son élégance cette méthode a deux défauts importants. D'abord elle ne réussit pas à construire une corrélation forte entre les états physiques et les états mentaux parce qu'elle n'est susceptible d'assurer qu'une survenance faible[2]. Le second défaut est que dans cette théorie les propriétés mentales sont menacées d'épiphénoménalisme.

La troisième des tentatives que j'ai présentées et illustréees dans le cadre de deux théories différentes vise à éviter le réductionnisme du physicalisme en prônant la multiréalisation de la cognition. Il s'agit d'une solution basée sur une survenance en deux pas. D'abord les états/propriétés mentaux individualisés en fonction de leur rôle fonctionnel surviennent d'un niveau inférieur dont la description est indépendante du substrat. Pour la première de ces tentatives, on a choisi la machine de Turing et pour la seconde les propriétés computationnelles. Ensuite, ce sont les propriétés de ce deuxième niveau qui surviennent du niveaux neuronal.

J'ai montré les limites du fonctionnalisme turingien (chapitre 7 §7.1 –§7.2) et j'ai aussi discuté de l'autre alternative, celle du fonctionnalisme représentationnel de Jerry Fodor pour conclure qu'elle n'est pas satisfaisante. En effet, j'ai démontré que même si l'on accepte l'hypothèse du langage de la pensée qui est à la base du parallélisme syntaxique-causal, Fodor n'arrive pas à prouver la nécessité métaphysique d'une corrélation forte entre les propriétés intentionnelles de contenu (compris selon une perspective informationnelle) et les propriétés computationnelles indispensables à la justification de la multiréalisation des secondes sur les premières. Ainsi, il n'est pas en mesure de justifier une corrélation entre les états intentionnels et les états physiques et de ce fait, il ne peut pas bâtir un cadre justificatif pour la multiréalisation.

Dans la troisième partie de mon travail j'ai présenté des stratégies du type *bottom-up*. En particulier j'ai fait état du nouvel élan phénoménologique des sciences cognitives (chapitre 9). J'ai montré comment cette école entend fonder la relation intentionnelle sur les données phénoménologiques ou de sens commun. La vertu de cette tendance est que, au lieu d'assimiler les fonctions cognitives à des fonctions computationnelles, elle préconise une conception systématique du système cognitif. J'ai expliqué comment les modèles morphogénétiques de Thom permettent de faire une description abstraite du substrat tout en partant d'une description de la dynamique spécifique

[2] Définition de survenance faible chapitre 2, §2.4.1.3.

de celui–ci.

Mon objectif dans la conclusion de cette thèse est de prendre de ces deux approches, celle qui est *top–down* et celle qui procède *bottom–up* ce qu'elles contiennent d'éléments positifs.

Les fonctionnalistes de types ont raison d'affirmer l'importance du substrat physique sur lequel sont basés les états intentionnels, mais ils ne peuvent pas dans ce cadre trouver place pour la multiréalisation de la cognition.

Le monisme anomal et le fonctionnalisme computationnel ont raison de proposer une relation de survenance à la place d'une identification mais ils ne sont pas à même d'attribuer une individualisation précise aux états physiques. Le fonctionnalisme computationnel en particulier n'a peut–être pas tort de formuler le concept de réalisation des propriétés mentales en des propriétés physiques par l'intermédiaire d'un niveau computationnel au lieu de parler d'une identification, mais il a tort de concevoir ce concept de réalisation comme totalement indépendant du substrat physique. Cette conception de la survenance du physique sur le mental implique que la définition de multiréalisation prend comme antécédentes et conséquentes les propositions disjonctives.

L'élan phénoménologique qui conduit à la notion de renvoi intentionnel à partir des données du sens commun a raison de procéder de cette manière mais il ne me semble pas suffisant pour tisser les liens intentionnels. J'ai déjà formulé mes critiques dans le chapitre 9. Néanmoins, l'outil qu'ils utilisent, les modèles morphodynamiques, se révélent pertinents à la description des états neuronaux. Mon objectif se précise donc. J'entends montrer que le concept d'émergence est compatible avec celui de réalisation. Ensuite je montrerai qu'à l'aide de ces deux concepts il sera possible de produire un concept de multiréalisation libéré des contraintes des antécédentes disjonctives[3].

Il s'agit d'illustrer le fait que lorsqu'on prend en compte la dynamique du substrat et qu'on applique les modèles de morphodynamique on peut obtenir une description abstraite des états physiques qui est issue de ces derniers. Une fois cette description obtenue, elle nous servira comme outil d'individualisation pour d'autres substrats qui présentent la même dynamique.

10.2 Réalisation et émergence: sont–ce des soeurs ennemies?

Dans le chapitre 8 j'ai exposé certaines propositions qui consistent à opposer les concepts de *réduction* et/ou de *réalisation* d'un côté et d'*émergence* de l'autre.

J'ai aussi remarqué que le caractère d'imprévisibilité ou de non–réductibilité ou d'inexplicabilité que l'on a prêté aux phénomènes dits émergents doit être rejeté parce qu'il les prive de toute pertinence scientifique et que ces caractéristiques ont leur origine historique dans l'impossibilité pour l'émergentisme britannique de s'adapter aux nouvelles données émanant des découvertes scientifiques du début du siècle.

Mon objectif est d'établir un parallèle entre les concepts de réalisation et d'émergence permettant à ce dernier de mériter la caractérisation de rationnelle. Ces caractéristiques je vais les

[3]La définition de multiréalisation d'une propriété P_L du niveau L au niveau $L-1$ est obtenue lorsqu'il existe une disjonction des propriétés du niveau $L-1$ dont l'instantiation d'un des termes est suffisante à l'instantiation de P_L et lorsque l'instantiation de P_L est suffisante à l'instantiation de la disjonction mais non pas à l'instantiation d'un de ces termes. Le schema 10.1 suivant montre la multiréalisation d'un loi d'une science spéciale, p. ex. la psychologie

$$\begin{array}{ccc} M_F & \longrightarrow & M_G \\ \uparrow & & \uparrow \\ NF_1 \vee NF_2 & \longrightarrow & NG_1 \vee NG_2 \end{array} \qquad (10.1)$$

Néanmoins, pour que réalisation il y ait, il est nécessaire que la relation du niveau réalisateur $NF_1 \vee NF_2 \longrightarrow NG_1 \vee NG_2$ puisse faire office de loi. C'est justement cela qui pose des problèmes.

Traditionnellement, les prédicats qui sont des antécédents des lois doivent se vérifier comme étant des espèces scientifiques moniques (*kinds*); autrement dit, les antécédents doivent représenter ou être caractérisés par une seule propriété. Dans la multiréalisation le problème se présente lorsque l'on s'interroge sur le caractère nomologique de l'énoncé réalisateur.

En effet, il s'agit de voir si la disjonction des espèces scientifiques différentes ou hétérogènes est acceptable comme antécédente d'une loi stricte ou non. (Voir Chapitre 7 §7.6.1.)

signaler par la suite. Le concept de propriété émergente doit être défini en relation avec un système et une théorie de ce même système. Les conditions dont il faut tenir compte dans ce cadre sont :

- Une définition claire et explicite des niveaux que comporte le système. Les critères pour la définition des niveaux devront non seulement prendre en compte les relations du type fonctionnel ou les compositions méréologiques mais encore et c'est fondamental ils devront faire appel à des caractéristiques structurelles dynamiques. Un système n'a pas qu'une composition; il a aussi une structure et une dynamique.

- Une propriété émergente est une propriété globale du système ou sous–système que l'on prend en considération et qui n'appartient à aucune des composantes qui lui servent de base. Cette condition assure, par ailleurs, le caractère non réductionniste des explications basées sur l'émergence.

- Toute propriété émergente est explicable à partir des propriétés des composantes du système ou de la dynamique des couplages entre ces derniers.

- Les niveaux ne sont pas des objets, mais des concepts. Ainsi la causalité descendante doit être récusée.

Quelles sont les différences et les coïncidences entre les deux concepts? [4]

L'émergence aussi bien que la réalisation s'adresse à des objets que l'on peut considérer comme constitués de différents niveaux. Les explications données par l'une et par l'autre partent du principe que les propriétés des niveaux supérieurs dépendent des propriétés du niveau immédiatement inférieur. Pour la réalisation il s'agit de révéler les mécanismes qui assurent la relation nomologique selon laquelle l'existence de propriétés du niveau inférieur se trouve suffir à l'existence d'une ou plusieurs propriétés du niveau supérieur. Dans le cas de l'émergence, c'est la complexité et l'organisation du niveau inférieur qui permettra l'existence des propriétés du niveau supérieur.

L'émergence que je défends est à la fois explicative et non–réductionniste. Cette première caractéristique la rapproche du concept de réalisation contrairement aux auteurs qui essaient de les opposer. En effet, j'ai signalé plus haut qu'une conception rationnelle de l'émergence permet d'expliquer les propriétés émergentes en relation à une théorie sans pour autant donner lieu à une réduction ontologique. Dans le chapitre 2 (§2.4.1) j'ai expliqué la différence entre la réduction ontologique et la réduction épistémologique et j'ai fait la même observation maintes fois au cours de ce travail. La réalisation a aussi ces deux caractéristiques, c'est–à–dire qu'elle est explicative et non réductioniste.[5] Finalement on voit que les deux théories sont réalistes vis–à–vis des propriétés d'ordre supérieur auxquelles ces concepts s'appliquent.

Pour ce qui est de la compatibilité avec la survenance des propriétés du niveau inférieur sur le niveau supérieur, je l'ai déjà démontrée dans le chapitre 2 (§2.1.4.4) concernant la réalisation. L'émergence est aussi compatible aver la survenance, du moins lorsqu'on définit cette dernière en termes généraux. Rappelons ici que la survenance exige, qu'une fois établies les propriétés du niveau inférieur, les propriétés du niveau supérieur covarient avec elles; une autre condition est la dépendance des propriétés émergentes par rapport à celles de la base et la troisième exigence est que les propriétés émergentes ne soient pas réductibles aux propriétés de base. Il est évident que l'émergence vérifie ces conditions. Le type de survenance, soit–elle forte, faible ou globale dépendra du système et de la théorie qu'on applique à l'émergence.

[4] Dans [Kim, 1992a] cet auteur propose aussi le conciliation de ces deux concepts. Néanmoins ma démarche est foncièrement différente de la sienne. Il y a deux aspects qui les distinguent. Le premier est que Kim se réfère à une conception ancienne de l'émergence parce que cet auteur se fonde sur des définitions de l'émergentisme britannique du début du siècle. La seconde différence tient au but de sa démarche. Kim donne des arguments en faveur de la compatibilité des deux concepts (émergence–réalisation) dans le sens qu'ils sont l'inverse l'un de l'autre pour illustrer le fait que les stratégies du physicalisme computationnel rencontrent les même difficultés que l'émergentisme britannique. Selon Kim, ni l'un ni l'autre ne peuvent octroyer des pouvoirs causaux aux propriétés du niveau supérieur. Voilà pourquoi tous deux sont obligés d'admettre le principe de causalité descendante. J'ai déjà critiqué cette position de Kim dans le chapitre 7. En outre, je récuse la pertinence de la causalité descendante pour l'émergence.

[5] Malgré le fait que la dernière n'est pas explicitement non–réductionniste elle est compatible avec une telle position (cf. chapitre 2 §2.4.4).

Dans le tableau suivant on voit un schéma du parallèle que je me suis proposé d'établir.

Les différences entre une conception rationnelle de l'émergence et la réalisation tiennent aux diverses manières d'expliquer les propriétés et aussi à la méthodologie appliquée par les théories qui les utilisent.

Je vais discuter ces deux différences par la suite, mais auparavant je voudrais faire remarquer que bien que je n'aie pas encore totalement justifié que la réalisation puisse être considérée comme l'inverse de l'émergence, je crois que j'ai déjà au moins réfuté la thèse qui veut voir dans l'émergence l'opposé de la réalisation.

Il me semble que l'argument qui a le plus de poids parmi ceux qui sont cités pour établir l'opposition entre émergence et réalisation est l'affirmation que la première fait appel aux propriétés physiques du niveau inférieur tandis que la réalisation permet d'établir des corrélations entre les propriétés de niveau supérieur et celles du niveau inférieur selon les mécanismes d'implantation. Ces mécanismes peuvent être *en principe détachés*, libres de toute référence aux propriétés physiques de premier ordre. Un exemple de cette démarche est la multiréalisation des propriétés mentales ainsi que j'en ai déjà fait mention dans le chapitre 7.

Émergence	Réalisation
non réductionniste	compatible avec le non réductionnisme
explicatif	explicatif
différence des niveaux qui justifie l'existence des sciences spéciales	différence des niveaux qui justifie l'existence des sciences spéciales
les propriétés du niveau supérieur dépendent de la complexité du niveau inférieur	les propriétés réalisées du niveau supérieur répondent aux liaisons nomologiques entre ces propriétés et des propriétés du niveau inférieur réalisateur
le propriétés physiques du niveau inférieur, dont une est le niveau de complexité, sont suffisantes pour l'existence des propriétés émergentes dans le niveau supérieur	l'instantiation des propriétés pertinentes de base est suffisante pour la réalisation de la propriété
réaliste vis-à-vis des propriétés émergentes	réaliste vis-à -vis des propriété réalisées
Compatible avec le concept de survenance	Compatible avec le concept de survenance
Méthode d'analyse: systémique et *bottom-up*. *Le tout est plus la somme des parties*	Méthode d'analyse: analytique et *top-down*. *Le fonctionnement de la totalité est explicable par la description des parties.*
les propriétés d'un même substrat sont définies comme émergentes en relation à une théorie	les propriétés réalisées sont définies par rapport à un mécanisme de réalisation

Je vais maintenant traiter des différences pour montrer que ce sont justement ces différences que me permettent d'affirmer que les deux notions sont inverses et non opposées. Je pourrai par la suite ainsi établir que les démarches fondées respectivement sur l'une ou l'autre de ces notions ne sont point antagonistes mais peuvent se révéler complémentaires.

Ensuite je présenterai une possibilité théorique de réfuter l'argument invoqué pour opposer l'émergence à la réalisation. En plus, si cette possibilité se vérifie elle permettra aussi de résoudre le problème des disjonctions bétérogènes suscité par les lois de la multiréalisation.

10.2.1 La méthode d'analyse

Les modèles qui utilisent la stratégie de la réalisation doivent en principe invoquer et décrire un mécanisme qui explique l'instantiation d'un état mental. Dans le cadre théorique j'ai donné l'exemple d'une telle démarche lorsque j'ai présenté le fonctionnalisme computationnel, aussi bien turingien que représentationnel. Nous avons vu que dans ce cas la caractérisation du problème est de type *top-down* parce qu'on part d'une interprétation intentionnelle du problème, c'est-à-dire d'une caractérisation en termes de contenus. Ainsi, en prenant à l'origine des propriétés, des états ou des lois de niveau supérieur on tente de décomposer le problème en se basant sur deux critères : l'économie des moyens et la décomposition fonctionnelle. On construit des explications en termes de sous-états ou de sous-modules plus simples qui rendent compte d'un aspect du problème jusqu'à arriver à des modules dont les fonctions sont si simples qu'elles ne sont plus décomposables et peuvent être accomplies par des unités ou modules très élémentaires. La description fonctionnelle ou modulaire interprète très souvent les modules comme étant encapsulés dans le sens que la tâche qu'ils accomplissent peut se réaliser sans l'intervention des autres modules.[6] Bref, il existe un principe de collaboration minimum entre ces modules d'un même niveau.

Les modèles qui ont pour but d'expliquer l'émergence ont une stratégie *bottom-up*. Ils mettent l'accent sur les modes d'organisation et une décomposition souvent autre que fonctionnelle ou méréologique des niveaux. Au contraire des théories qui se basent sur la notion de réalisation, les modèles émergentistes partent soit de la dynamique du substrat, soit des descriptions et du fonctionnement d'unités élémentaires très simples. La tâche consiste à montrer comment, à partir de cette description très fine et de bas niveau, les propriétés de haut niveau que l'on vise à expliquer se manifestent. Pour les modèles qui tiennent compte de la dynamique du substrat, l'application des modèles de la morphodynamique permet d'obtenir une caractérisation de la topologie des invariants morphodynamiques. En ce qui concerne les modèles des connectionnistes, le système montre les régularités que l'on espère expliquer par l'architecture que l'on a créée.

C'est cette différence de méthodologie entre les deux modèles (*bottom-up* et *top-down*) qui me permet de penser qu'il est possible de considérer l'émergence et la réalisation comme étant inverses. Les disputes et les controverses qui ont eu dans le passé sur l'incompatibilité des deux approches sont dépassées de longue date. [7] Les recherches en psychologie montrent que parfois les méthodes se combinent et se complètent pour permettre le retour à des données expérimentales.

La méthode d'explication des propriétés ou des états appartenant au niveau supérieur
Les explications données selon les deux méthodes semblent incompatibles pour deux raisons. La première tient à la présomption que les cognitivistes représentationnels entretiennent à tort quant à l'impossibilité pour les méthodes connectionnistes d'expliquer les fonctions cognitives supérieures. L'autre est que les modèles fondés sur la description partant du niveau inférieur s'avèrent difficilement compatibles avec la multiréalisation.

La présomption d'impossibilité est mise à mal par les récents travaux de Paul Smolensky notamment. Ce chercheur a développé les principes de ce qu'il appelle une architecture cognitive pour effectuer l'intégration connectionniste symbolique (*Integrated Connectionnist / Symbolic (ICS)*). Ce modèle qui vise à modéliser d'une part une structure abstraite sur laquelle les algorithmes tournent et de l'autre un type différent de structure à la base de l'explication du comportement du système se distingue des architectures traditionnelles : la symbolique, la connectionniste et l'hybride. Smolensky décrit ainsi son approche :

> Strictly speaking, this approach seeks to literally *implement* symbolic computation in networks of connectionnist units; such a network can then be described at two levels. At the lower levels there are algorithms for passing activity between units. At the higher level, there are symbolic representation and algorithms. Formal, precisely correct, and complete description are available at both levels. Such an inter-level reduction is what defines the implementation relation between the levels of description (or of virtual machines) in conventional computers. ...

[6] Voir par exemple la troisième partie de [Fodor, 1986] *La modularité de l'esprit*
[7] cf. [Fodor and Pylyshyn, 1988] et tout le volume N^0 28 de *Cognition*(1988) est dédié à cette controverse.

Rather than a literal implementation of symbolic computation, *ICS* relies on a *partial embedding* of symbolic computation in PDP^8 networks.[Smolensky, 1994]

Ce modèle permettrait l'implantation de la productivité d'uo langage. Ainsi, on pourra réfuter la critique cognitiviste représentationnelle la plus répandue selon laquelle aucune modèle comportant une structure inspirée des principes connectionnistes ne saurait donner un compte–rendu de cette qualité cardinale pour la représentatioo symbolique. En utilisant de manière astucieuse des outils mathématiques comme la tbéorie des tenseurs, Paul Smolensky semble avoir franchi avec son modèle la césure qui sépare le niveau inférieur ou algoritbmique du niveau supérieur symbolique. L'autre critique ressort de l'affirmation que les modèles qui présentent une explication émergente sont incompatibles avec la multiréalisation. Je l'aborderai dans le paragrapbe suivant pour montrer qu'il y a une possibilité de la contourner.

Sur l'incompatibilité entre les modèles émergentistes et la multiréalisation Comment peut–on affirmer qu'une théorie émergentiste qui part de la description du niveau inférieur, surtout lorsqu'il s'agit d'une théorie prenant comme niveau le plus bas l'organisation du système nerveux pourra–t–elle se détacher de cette caractéristique de manière à donner une description indépendante du substrat? Voici la question.

En définitive, dans cette question l'idée est implicite que toute explication donnée compte tenu des caractéristiques d'une microstructure est foncièrement dépendente de cette dernière.

Mon objectif dans cette section est double car j'entends montrer que la notion de microstructure–standard peu être élargie et enrichie. Je vais défendre l'idée selon laquelle les relations entretenues par les composantes de la microstructure sont conçues dans la plupart des cas comme des liens entre totalité et partie, ou dans le meilleur des cas comme des composantes des structures caractérisées seulement en fonction des dispositions spatiales. J'entends aussi démontrer que les descriptions qui se basent sur une conception restreinte des types physiques ne suffisent pas à établir une corrélation avec des états mentaux. Je vais donner un exemple où cette description statique ne s'avère pas assez fine pour donner un compte rendu des activités perceptives visuelles.

J'espère ainsi poser des bases pour pouvoir émettre la conjecture qu'une analyse de la dynamique neuronale peut servir de fondement à une description de types physiques indépendante du substrat. Il y aura donc une chance de multiréalisation parce que l'on aurait trouvé un front commun à tous les membres de la disjonction de manière à justifier l'existence d'une espèce scientifique nomique. Si une telle conjecture se vérifie on aura mis au point une description de types physiques qui tout en partant du substrat s'avèrera indépendante de lui. Cette description a, selon moi, plus de chances de réussir que l'approche *décharnée* du physicalisme computationnel en termes de rôles fonctionnels. En définitive il me semble possible avec les outils mathématiques que l'on a maintenant de parvenir à une caractérisation qui pourrait prendre la forme suivante: tous les substrats ayant cette dynamique sont des réalisations de la propriété X.[9]. En conséquence, à la différence de Joëlle Proust [10] qui semble attribuer une place très restreinte à la multiréalisation, je laisse la porte ouverte à une possible vérification de la multiréalisation entre les différentes espèces d'êtres vivants. Cependant je me rallie à son plaidoyer pour une étude des neurosciences tendant à une investigation approfondie des types physiques et conduisant à l'affaiblissement des arguments basés sur l'existence de la multiréalisation qui consistent à récuser d'emblée toute hypothèse jugée en désaccord avec cette dernière.

J'aimerais maintenant revenir sur le concept de microstructure. Dans le chapitre 8 (§8.4.5.) j'ai remarqué que la démarche de Beckermann qui visait à caractériser les microstructures même restreintes à un seul substrat donnait lieu à une caractérisation extensionnelle. Cette caractérisation d'extension qui semble *a priori* la seule possible découle des énoncés des lois qui comportent des antécédents disjonctifs. J'ai soutenu que cette proposition ne répond pas aux besoins justificatifs

[8] *PDP*: *Parallel distribute processing* qui fait référence au processus que l'on construit à aide de réseaux neuronaux.

[9] J'ai déjà montré dans le chapitre 8 que les modèles morphodynamiques sont capables de dégager des invariantes qui permettent de caractériser une topologie morphodynamique.

[10] cf. [Proust, 1993] et chapitre 7 de ce travail

de la notion de réalisation pour deux raisons: la première étant l'impossibilité de formuler des énoncés nomologiques à partir de cette situation et la seconde tenant à la valeur explicative douteuse de cette proposition. Rappelons que la définition de réalisation requiert aussi une relation nomologique entre éléments de niveau inférieur et supérieur qui soit explicative.

Comment caractériser en général la notion d'état physique en tant qu'état du cerveau?

A mon avis les sciences cognitives représentationnelles ont une forte tendance à concevoir l'individualisation des états physiques en fonction des caractéristiques structurelles et de localisation spatiale, elles sont ancrées dans le clivage structure/fonction. J'ai aussi montré que si la seconde position se révèle trop libérale, la première, en revanche, est insuffisante. Je vais essayer d'illustrer cette insuffisance par l'exemple suivant.

Petitot cite l'hypothèse du *labeling hypothesis* dans le *binding problem*. Le *binding problem* se réfère au passage de la cohérence locale à la cohérence globale dans la perception d'un objet, c'est-à-dire la manière dont à partir de la perception des différentes caractéristiques d'un objet qui a lieu aux niveaux inférieurs de la vision on passe à la perception de l'ensemble. L'hypothèse dite *labeling hypothesis* soutient que ces différentes caractéristiques de l'objet (comme par exemple la segmentation des images) dont la perception est distribuée au niveau neuronal sont réalisées en fonction d'un codage temporel. La *cohérence* des caractéristiques et de composants sera encodée par la *synchronisation* des réponses neuronales aux stimuli oscillatoires. En revanche, les phases différentes ou asynchroniques serviront à encoder les composantes différentes. L'évidence expérimentale montre que les oscillations appartenant au niveau des hypercolonnes corticales sont sensibles à la cohérence du stimulus. [Engel et al., 1992]

Petitot explique très bien la valeur empirique mais aussi théorique de l'hypothèse.

> The binding problem is evident. At the early stages of perception the features of the objets are extrated in a local, distributed and parallel manner. How these localized features and modularited constituents (parts) can be re-integrated in spite of their distributed encoding? One must avoid the "superposition catastrophe", that is their mere linear superposition.
>
> For this one uses a new parameter: temporal coincidence, synchronization. The binding resulting from functional coupling becomes *dynamic*, purely transient. It is more the consequence of a fixed anatomical wiring. [Petitot, 1993]

L'hypothèse que je viens de citer se vérifie et prouve qu'il y a certains états physiques intégrant ou constituant la réalisation d'un état mental perceptif qui ne sont exhaustivement explicables ni de manière structurelle, ni de manière fonctionnelle.[11]

Le passage de la cohérence locale à la cohérence globale du stimulus est le résultat d'une dynamique. Je crois que cette dernière caractéristique est centrale à la description des *modes d'organisation* que j'ai citée ainsi que Proust comme alternative à l'individualisation des états physiques.

La nécessité de prendre en compte la dynamique et aussi de travailler à une description qui fasse état des niveaux d'organisation nous engage à tirer parti des résultats des neurosciences du point de vue systémique. Cette approche systémique ne consistera pas seulement en une description en termes de systèmes et de sous-systèmes mais devra tenter de décrire mathématiquement les phénomènes observés dans le but de trouver des invariants issus de la théorie morphodynamique.

D'ailleurs, dans certains domaines des neurosciences comme par exemple la vision on a déjà commencé des recherches dans ce sens. En effet, des cellules appartenant au Corps Genouillé Latéral qui ont une forme de champ concentrique dont le centre est ON et le pourtour OFF, constituent un type de filtre servant à déterminer les bords dans le champ perceptif. David Marr a montré que ces filtres appliqués aux données d'entrée décrites par la fonction d'intensité bi-dimensionnelle $I(x, y)$ effectuent une analyse multi-scalaire et de localisation spatiale. Il s'agit de trouver les taux maximaux de changement; donc la fonction à analyser n'est pas celle d'intensité, mais celle qui produit les changements, c'est-à-dire la fonction dérivée. Néanmoins, l'input dont nous partons est une matrice discrète d'intensités $I(x, y)$ et par conséquent le transformer en l'algorithme continu de Marr-Hildreth requiert l'utilisation de la formule de Gauss en deux dimensions.[12]

[11] Alain Connes fait la même observation dans [Changeux and Connes, 1989, cf. page 194]

[12] Le choix de la fonction Gaussienne est dû, entre autres raisons, à sa symétrie qui permet d' exprimer sa dérivée

Donc les dérivées secondaires seront prises en terme Laplacien et ses racines seront les points de passage par zéro. Il faut préciser qu'on parle du Laplacien de G*I. C'est le critère connu comme *zero-crossing critérium.*

Les *zero-crossing* sont en fait des invariants dans la géométrie différentielle (2-jet)[13]. Jan Koenderink a avancé l'hypothèse que les colonnes du cortex visuel (air 17 (V1)) implantent une analyse de jets multi-scalaires.[14]

Cet exemple que je viens de citer illustre une autre façon de caractériser les états physiques qui est différente de la structurelle et de la fonctionnelle.

En appliquant la géométrie différentielle à partir des données des neurosciences dans ce cadre, on arrive à une description qui est indépendante du substrat tout en ressortant de la dynamique de ce dernier. L'idée que j'ai essayé d'accréditer est qu'il serait raisonnable de favoriser la quête de propriétés physiques des états neuronaux en les considérant comme des systèmes dynamiques et d'appliquer à cet effet les modèles morphodynamiques comme on fait normalement dans l'étude des phénomènes de la physique.

Il me faut toutefois signaler que ce que je préconise ici comme étant une méthode raisonnable de recherche doit être distinguée des positions que j'ai moi-même critiquées dans le chapitre 9 lorsque j'ai parlé de fallacie *chaussure-mouche.* Dans cette partie du travail j'ai critiqué certaines approches de la vie artificielle qui, en érigeant des systèmes pour simuler certains phénomènes instantient une loi mathématique qui caractérise ce phéoomène. Cette seule caractéristique suffit pour que les chercheurs dans le domaine de la vie artificielle parviennent à l'audacieuse conclusion que le système artificiel est une réalisation et non une simulation du phénomène en question.

La proposition que je défends ici est bien plus large, il ne s'agit pas de décrire un phénomène isolé par une caractérisation mathématique mais d'en faire une description globale qui prenne en compte la dynamique du système et qui fournisse comme élément de caractérisation non point une formule mathématique mais une description topologique eo termes d'attracteurs et de bassins.

Évidemment la partie est loin d'être gagnée d'avance. Très probablement la dynamique cérébrale est extrêmement compliquée et possède une structure topologique constituée d'attracteurs très complexes mais la piste que je viens d'iodiquer me semble pour toutes les raisons citées pleine d'espoir.

Finalement, je suis d'accord avec Petitot quand il dit qu'une telle démarche mettra des générations pour aboutir. Il n'est pas impossible que l'étude des sous-systèmes spécifiques du point de vue fonctionnel tels que la vision puissent donner quelques résultats dans un avenir plus proche.

Néanmoins, si une telle description des états physiques en termes d' attracteurs et de bassins est possible, alors la présomption selon laquelle toute description des modèles *bottom-up* partant du substrat biologique est dépendante de ce dernier a une chance d'être réfutée. Pour la même raison, l'incompatibilité entre l'émergence et la multiréalisation pourra aussi être démentie.

Ainsi, il sera possible d'affirmer que deux substrats différents du point de vue de leur composition spacio – structurelle, présentant néanmoins des dynamiques reposant sur une même structure topologique en termes des invariants sont des réalisations des mêmes types de macroproprietés, macropropriétés qui sont des propriétés émergentes du substrat. En outre, cette description sera commune à tous les états physiques réalisateurs, ce qui démontrera que la disjonction de ces états est une espèce scientifique nomique.

---

en fonction du rayon au centre à la place d'abscisses et d'ordonnées -(x, y)-. De cette façon on peut considérer l'opérateur Laplacien de la fonction de Gauss comme un filtre.

[13] Pour une définition de la notion de jet est de ses conséquence cf. Définition 29 du chapitre 8.

[14] Pour plus de détails et d'exemples cf. [Petitot, 1994b]

10.3 La topologie des invariants et la multiréalisation

La topologie qui est l'étude des invariants, soit des espaces topologiques[15], soit à homotopie[16], soit à homéomorphie[17] se présente comme un outil de choix pour s'appliquer à l'étude de la multiréalisation.

Comme Alain Connes le signale en s'adressant à Jean–Pierre Changeux :

> ... je ne prétends avoir aucune sorte de compréhension nouvelle du fonctionnement cérébral. Je pense qu'il serait bon que certaines notions élémentaires de topologie ... soient mieux connues des scientifiques neurobiologistes comme toi. Mais pourquoi la topologie? Comme tu l'as expliqué, la conformation du cerveau n'est pas identique d'un cerveau à l'autre, pas plus que le perception d'un objet extérieur. Mais les propriétés sur lesquelles on s'accorde ont un caractère d'invariance, de «stabilité structurelle» pour citer Thom, dont le cadre théorique comme la topologie rend asses bien compte. [Changeux and Connes, 1989, page 171]

et puis il ajoute,

> Nous parlons de l'existence d'une grande diversité, mais aussi d'une certaine invariance, dans la manière dont le cerveau est construit d'un individu à un autre. La topologie est précisément le cadre idéal pour comprendre ce genre de phénomène, puisque le même objet topologique peut avoir de nombreuses réalisations différentes. ... [Changeux and Connes, 1989, page 180]

Alain Connes parle de la topologie comme d'un outil que le cerveau utilise pour le codage des formes stables. Ici mon but est différent car il s'agit d'appliquer la topologie à l'ensemble des

---

15

DEFINITION 43 (ESPACE TOPOLOGIQUE)
We say a tological structure *or, simply, a* topology, *is defined on a set X if there is given a class of subsets of X which contains the union of any collection in the class and the intersection of any finite collection in the class. A set endowed with a topological structure is called a* topological space, *its elements–* points, *and the sets of the given class –* open sets. *[Fuks and Rokhlin, 1984, page 5]*

16

DEFINITION 44 (HOMOTOPIE. DÉFINITIONS GÉNÉRALES)
A continuous map $f' : X \to Y$ is homotopic to the continuous map $f : X \to Y$ if there is a continuous map $F : X \times I \to Y$ such that $F(x,0) = f(x)$ and $F(x,1) = f'$, for all $x \in X$. Every such map F is called a homotopy form f to f' *(or* connecting f to f'*. One says also that F is a* homotopy of f. *A map homotopic to a constant map is also said to be* null bomotopic. *Often a homotopy $F : X \times I \to Y$ is interpreted as a family of continous maps $f_t : X \to Y$, related to F via $f_t(x) = F(x,t)$ $(0 \leq t \leq 1)$.*

...

Obviously, the constant *homotopy F of a continuous map $f : X \to Y$, given by $F(x,t) = f(x)$, connects f to f; if the homotopy F connects f to f', then the* inverse *homotopy, F', defined by $F'(x,t) = F(x,1-t)$, conects f' to f.*

... *Thus* homotopy is an equivalence relation, *which yields a partition of $C(X,Y)$ into equivalence classes, called* homotopy classes. *We denote the set of these classes by $\pi(X,Y)$. [Fuks and Rokhlin, 1984, page 57]*

EXEMPLE 10
An example is the rectilinear homotopy. *Namely, let f and f' be continuous maps of a space X into a subspace Y of $\mathbf{R}^n$. If for each $x \in X$ the segment joining $f(x)$ to f' is entirely* contained *in Y, then $F(x,t) = (1-t)f(x) + tf'$ defines a homotopy from f to f', referred to as rectilinear. [Fuks and Rokhlin, 1984, page 57]*

DEFINITION 45 ((*))
Let the maps $f, f' : X \to Y$ be homotopic. Then given any contiuous maps $g : Y \to Y'$ *and* $h : x' \to X$, the maps $g \circ f \circ h$ and $g \circ f' \circ h$ *are homotopic. In fact , let $F : X \times I \to Y$ be a homotopy from f to f'. Then $g \circ f \circ h$ and $g \circ f' \circ h$. [This shows the] the mapping $C(h,g) : C(X,Y) \to C(X',Y')$ induced by two continuous maps $h : X' \to X$ and $g : Y \to Y'$ transforms homotopy classes into homotopy classes. The* resulting *mapping fact $C(h,g) : \pi(X,Y) \to \pi(X',Y')$ is denoted by $\pi(h,g)$, and [*] implies that it depends only on the homotopy classes of h and g. [Fuks and Rokhlin, 1984, page 57]*

17

DEFINITION 46 (HOMEOMORPHISME)
Soit $f : X \to Y$ continue est un homeomorphisme *si f^{-1} est une fonction continue.*

attracteurs et des bassins obtenus par une description dynamique d'un état physique de manière à utiliser les invariants qui en découlent pour caractériser cet état physique. Ainsi on pourra arriver à individualiser les états physiques en termes des invariants topologiques appliqués aux invariants dynamiques des attracteurs et des bassins. Il me semble possible d'arriver à une formulation de la multiréalisation qui ait l'allure suivante : deux états physiques qui sont en corrélation avec le même état fonctionnel sont des réalisations de cet état s'ils possèdent les mêmes invariants topologiques.

L'état physique qui sert de témoin sera celui qui appartient à l'espèce pour laquelle nous avons la certitude que l'état fonctionnel en question se vérifie. Reprenons le cas de la douleur que nous avons discuté dans le chapitre 7 (§7.6.2) lorsque j'ai critiqué le concept de multiréalisation. Dans ce cas nous nous sommes accordés à accepter qu'il est pertinent de parler de la douleur pour des espèces proches des primates et nous avons mis en doute que l'état fonctionnel de la douleur puisse être attribué aux individus appartenant à des espèces qui en sont phylogénétiquement très éloignées. L'individualisation en termes de la topologie des invariants que je défends pourra nous servir de moyen de comparaison dans ce cas.

Par ailleurs, si cette caractérisation en termes de topologie est possible, elle servira de dénominateur commun à tous les états physiques qui réalisent le même état fonctionnel en libérant la théorie du problème des antécédents disjonctifs et en montrant par le même fait qu'il s'agit d'un même type nomique.

10.3.1 Une stratégie raisonnable de recherche

Cependant, comme le signale Jean-Pierre Changeux, il ne suffit pas de disposer d'un modèle formel théorique très général. "Ce qu'il faut, pour lors, c'est faire en sorte que ces propositions aboutissent à des expériences réalisables en laboratoire."[Changeux and Connes, 1989, page 188]

Voilà pourquoi un programme de recherche raisonnable à besoin des deux stratégies : *top-down* et *bottom-up*. La redéfinition du concept rationnel d'émergence montre qu'il est l'inverse de la réalisation, ce dernier fait a une autre conséquence d'ordre méthodologique; les méthodes d'analyse qui en découlent ne seront pas concurrentes dans le sens que la réussite de l'une entraînera l'échec de l'autre. Elles seront complémentaires. L'observation de Changeux va dans le sens de la complémentarité des méthodes. Le neurobiologiste cherche les bases neurales des fonctions supérieures du cerveau humain. Dans ce sens, il part des (macro) propriétés psychologiques (par exemple la perception de visages, la discrimination auditive des syllabes en écoute dichotomique, certaines tâches de reconnaissance visuelle) qu'il essaye d'expliquer en termes d'organisation, de règles d'interaction et des propriétés des éléments qui composent le niveau neuronal. Dans ce sens il s'agit d'une approche *top-down*. Une fois les données obtenues une approche *bottom-up* peut commencer. On pourra fournir des interprétations formulées mathématiquement sur les caractéristiques des niveaux inférieurs. La complémentarité des deux approches me semble évidente.

La finesse descriptive et le progrès des deux méthodes devront favoriser la convergence des deux explications. Si la méthode *top-down* vise à expliquer la réalisation des propriétés de niveau supérieur en fonction du niveau inférieur car elle a comme but la découverte des mécanismes d'implantation, la seconde *bottom-up* montre la complexité des organisations neuronales et rend possible de postuler des hypothèses sur la dynamique de manière à obtenir une explication de l'émergence des propriétés supérieures à partir du substrat étudié.

10.4 Conclusion

Dans ce travail j'essaye de poser le problème de la naturalisation de l'intentionnalité et de faire le tour, selon une perspective historique, de quelques unes des solutions les plus importantes de ces dernières années. J'ai signalé les limites de ces solutions et ce faisant j'ai montré les difficultés que l'on éprouve en voulant dépasser la thèse brentanienne de l'irréductibilité du mental. On a vu que les objections suscitées par la tentative de rendre compatible les trois propositions du trilemme classique sont énormes.

Le positivisme a eu parfois une influence heureuse dans la quête constante de rigueur des propositions et des théories, mais cette même rigueur a causé la confusion de la réduction ontologique avec la réduction épistémologique ou explicative parce que l'on s'attachait fondamentalement au modèle naguélien de réduction. Je pense que j'ai décrit d'autres possibilités qui illustrent l'énoncé suivant: *explain does not mean explain out.*

J'ai aussi souligné que la menace d'être considéré comme réductionniste a donné lieu non seulement à une récusation du physicalisme de types, mais aussi au postulat de multiréalisation de la cognition.

A cette dernière hypothèse on a attribué un poids excessifs. On a souvent traité cette possibilité logique comme si elle constituait un fait empirique irréfutable. Cette attitude a engendré une conséquence méthodologique dramatique: on a relégué les neurosciences à l'arrière plan pour leur substituer des modèles artificiels. En ce faisant, on n'a pas su profiter de résultats des neurosciences et on a cru que la valeur des modèles artificiels était bien plus qu'heuristique.

Je pense avoir rendu manifeste que les théories à la recherche d'une description neutre de la corrélation entre les états mentaux et les états physiques en termes de leur rôle fonctionnel sans passer par le substrat nerveux ne parviennent pas à donner une solution au trilemme. Cette insuffisance m'a amenée à défendre la pertinence du niveau biologique pour la caractérisation des états physiques.

Je crois que la conjecture que j'ai défendue sur la caractérisation des états physiques en termes de topologie des invariants dynamiques (des attracteurs et des bassins) est non seulement plausible mais qu'elle peut permettre d'établir l'isomorphisme tant recherché entre états mentaux et états physiques sans s'avérer pour autant réductionniste.

Finalement, je ne pense pas que l'on soit trop gourmand lorsqu'on tente de résoudre le trilemme. Néanmoins, la tâche que l'on se fixe est immense, mais je crois qu'un jour nous arriverons à expliquer en vertu de quelles propriétés physiques nous possédons les propriétés mentales qui nous permettent de croire, de désirer, de percevoir, d'aimer. Peut-être qu'un jour nous pourrons mettre toute l'eau de la mer dans un petit trou de sable; les quelques gouttes que l'on perd lorsqu'on se met à l'ouvrage ne doivent donc point nous décourager dans notre tâche.

Le Locle, le 27 juin 1996

Index des noms

Index des notions

Bibliographie

[Achim, 1992] Achim, S. (1992). *The historical facets of Emergence*, pages 25–48. In [Beckermann et al., 1992].

[Alexander, 1974] Alexander, P. (1974). Boyle and Locke on primary and secondary qualities. *Ratio*, 16: 51–87.

[Alexander, 1920] Alexander, S. (1920). *Space, Time, Deity*, volume I–II. Macmillan, London.

[Andler, 1985] Andler, D. (1985). Logique mathématique. In *Encyclopædia Universalis*, volume Corpus 11, pages 185–196. Encyclopædia Universalis, Paris.

[Andler, 1992] Andler, D., editor (1992). *Introduction aux sciences cognitives*. Ed. Gallimard, Paris.

[Aquila, 1995] Aquila, R. E. (1995). Intentionality. In [Kim and Sosa, 1995], pages 224–145.

[Armstrong, 1970] Armstrong, D. M. (1970). The nature of Mind. In [Block, 1980a], pages 190–199. Aussi dans C. V. Borst (Ed.) (1970) *The Mind \ Brain Identity Theory*, London, Macmillan.

[Assad and Packard, 1991] Assad, A. and Packard, N. (1991). *Emergent colonization in an artificial ecology*, pages 143–152. In [Varela and Bourgine, 1991].

[Austin, 1962] Austin, J. L. (1962). *How to do things with words*. Oxford Universtiy Press, New York.

[Ayer, 1982] Ayer, A. J. (1982). *Language, Truth, and Logic*. Victor Gollancz, London.

[Balescu, 1985] Balescu, C. (1985). Irréversibilité. In *Encyclopædia Universalis*, volume Corpus 10, pages 174–175. Encyclopædia Universalis, Paris.

[Beakley and Ludlow, 1992] Beakley, B. and Ludlow, P., editors (1992). *The Philosophy of Mind*. MIT Press, Cambridge.

[Beckermann et al., 1992] Beckermann, A., Flohr, H., and Kim, J. (1992). *Emergence or Reduction ?* Walter de Gruyter, Berlin–New York.

[Bell, 1994] Bell, D. (1994). Logicism. In [Dancy and Sosa, 1992], page 265.

[Berckermann, 1992] Berckermann, A. (1992). *Supervenience, Emergence, and Reduction*, pages 94–118. In [Beckermann et al., 1992].

[Besnier, 1993] Besnier, J. M. (1993). *Histoire de la philosophie moderne et contemporaine*. Grasset, Paris.

[Blitz, 1990] Blitz, D. (1990). *Emergence evolution and the level struture of reality*, pages 153–169. In [Weingartner and Dorn, 1990].

[Block, 1978] Block, N. (1978). *Troubles with Functionalism*. Volume *I* of [Block, 1980a].

[Block, 1980a] Block, N., editor (1980a). *Readings in phisolophy of psychology*, volume I et II. Harvard University Press, Cambridge.

[Block, 1980b] Block, N. (1980b). *What is functionalism?* , pages 171–184. Volume *I* of [Block, 1980a]. Introduction à la troisième partie.

[Block and Fodor, 1972] Block, N. and Fodor, J. (1972). Wbat psychological states are Not. In [Block, 1980a], pages 237–250.

[Brentano, 1924] Brentano, F. (1924). *Psychologie vom Empirischem Standpunkt.* Felix Meiner Verlag, Leipzig. Trad. française de Gandillac, M. (1944) *Psychologie du point de vue empirique*, Paris, Aubier-Montaigne.

[Brightman, 1926] Brightman, E. S., editor (1926). *Proceedings of the sixth international congress of philosophy*, New York.

[Broad, 1925] Broad, C. (1925). *The Mind and Its Place in Nature.* Routledge and Kegan Paul, London.

[Bunge, 1959] Bunge, M. (1959). *Causality. The place of the causal principle in modern science.* Barvard University Press, Cambridge. Trad. espagnol de Bernán Rodiguez (1961) *Causolidad. El principio de causalidad en la ciencia moderna.* Buenos Aires. EUDEBA.

[Bunge, 1976] Bunge, M. (1976). Levels and reduction. *Am. J. Physiol.: Regul., Inter. and Compar. Physiol.*, 2(2): 75–82.

[Bunge, 1977] Bunge, M. (1977). Emergence and the mind. *Neuroscience*, 2: 501–509.

[Bunge, 1979] Bunge, M. (1979). *Treatrise on Basic Philosophy. Ontology II: a world of systems*, volume 4. D. Reidel Publishing Co.

[Bunge, 1994] Bunge, M. (1994). *L'écart entre les mathématiques et le réel*, pages 165–173. In [Porte, 1994].

[Burge, 1986] Burge, T. (1986). Individualisme and Psychology. *Philosophical Review*, 95(1): 3–45.

[Capitan and Merrill, 1967] Capitan, W. H. and Merrill, D. D. (1967). *Art, Mind and Religion.* University of Pittsburgh, Pittsburgh.

[Cariani, 1991] Cariani, P. (1991). *Emergence and Artificial Life*, pages 775–797. In [Langton et al., 1991].

[Carnap, 1928] Carnap, R. (1928). *Der logische Aufbau der Welt.* Weltkreis-Verlag, Berlin-Schlachtensee. Trad. anglaise Berkeley, R. A. (1967) *The logical construction of the world.* Cambridge, University Press, Berkeley.

[Carnap, 1932] Carnap, R. (1932). *An excerpt from Psychology in physical language*, pages 23–28. Volume I et II of [Block, 1980a]. Originaire parue en 1932 dans *Erkenntnis* Vol. 11.

[Causey, 1977] Causey, R. L. (1977). *Unity of science.* Reidel, Dordrecht.

[Cayla, 1991] Cayla, F. (1991). *Routes et déroutes de l'intentionnalité.* Ed. de l'èclat, Combas.

[Changeux and Connes, 1989] Changeux, J.-P. and Connes, A. (1989). *Matière à pensée.* Editions Odile Jacob.

[Chenciner, 1985] Chenciner, A. (1985). Singularité et fonctions différentiables. In *Encyclopædia Universalis*, volume Corpus 16, pages 933–943. Encyclopædia Universalis, Paris.

[Chisholm, 1972] Cbisholm, R. (1972). *Beyond Being and Nonbeing*, pages 53–67. In [Chisholm, 1982].

[Chisholm, 1973] Chisholm, R. (1973). *Homeless Objets*, pages 37–52. In [Chisholm, 1982].

[Chisholm, 1977] Chisholm, R. (1977). *Theory of Knowledge*. Prentice Ball, Englewood Cliff, N.J.

[Chisholm, 1982] Chisholm, R. (1982). *Brentano and Meinong Studies*. Editons Rodopi B. V., Amsterdam.

[Chomsky, 1957] Chomsky, N. (1957). *Syntactic Structures*. MIT Press, Cambridge. Trad. française Braudeou, M (1969) Paris, Ed. Seuil.

[Chomsky, 1959] Chomsky, N. (1959). *A review of B. F. Skinner's Verbal Behaviour*, pages 48–63. Volume 1 of [Block, 1980a]. Originairement *Langage*(1959) 35(1), 26-58.

[Chomsky, 1965] Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press., Cambridge.

[Churchland, 1981] Churchland, P. M. (1981). *Eliminative Materialisme and Propositional Attitudes*, pages 206–223. In [Lycan, 1990]. Aussi dans Churchland, Paul M. (1992a) pages 1-22.

[Churchland, 1992a] Churchland, P. M. (1992a). *A Neurocomputation approache*. MIT Press, Cambridge.

[Churchland, 1992b] Churchland, P. M. (1992b). *Reduction, Qualia, and the Direct Instrospection of Brain States*, pages 47–66. In [Churchland, 1992a].

[Churchland and Sejnowski, 1990] Churchland, P. S. and Sejnowski, T. (1990). *Neural Representation and Neural Computation*, pages 224–252. In [Lycan, 1990].

[Copeland, 1994] Copeland, B. J. (1994). Artificial intelligence. In [Guttenplan, 1994], pages 122–131.

[Cummins, 1980] Cummins, R. (1980). *Functional Analysis*, pages 185–190. Volume *I* of [Block, 1980a].

[Cummins, 1983] Cummins, R. (1983). *The nature of Psychological Explanation*. MIT Press, Cambridge, Mass.

[Dancy and Sosa, 1992] Dancy, J. and Sosa, E., editors (1992). *A Companion to the Epistemology*. Blackwell Companions to Philosophy. Basi Blackwell Ltd., Oxford, deuxième 1994 edition.

[Davidson, 1963a] Davidson, D. (1963a). Actions, Reason, and Causes. *Journal of Philosopy*, 60: 685–700. Aussi dans Davidson, Donald (1980a).

[Davidson, 1983b] Davidson, D. (1983b). Causal Relation. *Journal of Philosopy*, 64: 691–703. Aussi dans Davidson, Donald (1980a).

[Davidson, 1970] Davidson, D. (1970). *Experience and Theories*, chapter Mental Events, pages 79–101. University of Massachussetts Press, Amherst. Aussi dans Davidson(1980a) et Beakley, B. et Ludlow, P. (1992).

[Davidson, 1980a] Davidson, D. (1980a). *Essay on Actions and Events*. Oxford University Press, Oxford. Trad. française Engel, Pascal (1993) *Actions et événements*, Paris, PUF.

[Davidson, 1980b] Davidson, D. (1980b). *Psychology as Philosophy*, chapter 12, pages 230–259. In [Davidson, 1980a]. Trad. française Engel, Pascal (1993) *Actions et événements*, Paris, PUF.

[Davidson, 1984a] Davidson, D. (1984a). Belief and the Basis of Meaning. In [Davidson, 1984d], chapter 10.

[Davidson, 1984b] Davidson, D. (1984b). Radical Interpretation. In [Davidson, 1984d], chapter 9.

[Davidson, 1984c] Davidson, D. (1984c). Thought and Talk. In [Davidson, 1984d], chapter 11.

[Davidson, 1984d] Davidson, D. (1984d). *Truth and Interpretation*. Clarendon Press, Oxford.

[Davidson, 1993] Davidson, D. (1993). Thinking Causes. In [Beil and Mele, 1993], pages 3–18.

[Davidson, 1994] Davidson, D. (1994). *Davidson, Donald*, pages 231–236. In [Guttenplan, 1994].

[de Souabe Zyriane, 1985a] de Souabe Zyriane, P. (1985a). Berbart, J. F. In *Encyclopædia Universalis*, volume Thesurus Index II, pages 1372–1373. Encyclopædia Universalis, Paris.

[de Souabe Zyriane, 1985b] de Souabe Zyriane, P. (1985b). Lotze, R. H. In *Encyclopædia Universalis*, volume Thesurus Index II, page 1777. Encyclopædia Universalis, Paris.

[de Souabe Zyriane, 1985c] de Souabe Zyriane, P. (1985c). Wuodt, Wilhelm. In *Encyclopædia Universalis*, volume Thesurus Index III, page 3183. Encyclopædia Universalis, Paris.

[Dennett, 1990] Dennett, D. (1990). *Quining Qualia*, pages 519–547. In [Lycan, 1990].

[Döring, 1995] Döring, F. (1995). Contrafactuals and Laws: A methodological diversion. Seminaire au CREA–Ecole Polytechnique en julliet de 1995.

[Drestke, 1981] Drestke, F. (1981). *Knowledge and the Flow of Information*. MIT Press, Cambridge.

[Drestke, 1986] Drestke, F. (1986). Misrepresentation. In Bogdan, R. J., editor, *Belief*. Oxford Univ. Press, Oxford.

[Drestke, 1994] Drestke, F. (1994). *Drestke, Fred*, pages 259–265. In [Guttenplan, 1994].

[Dretske, 1971] Dretske, F. (1971). Conclusive reasons. *Australian Journal of philosophy*, 49: 1–22.

[Dreyfus, 1982a] Dreyfus, B. (1982a). The perceptual *Norma*. In [Dreyfus, 1982b], pages 97–123.

[Dreyfus, 1982b] Dreyfus, H. L., editor (1982b). *Husserl Intentionality and Cognitive Science*. MIT Press, Cambridge.

[Driesch, 1926] Driesch, H. (1926). Emergent evolution. In [Brightman, 1926], pages 1–9.

[Dubucs, 1992] Dubucs, J. P. (1992). Arguments gödeliens contre la psychologie computationnelle. *Centre de Recherche Sémilogiques - Travaux de logique*, 7: 73–90.

[Dummet, 1954] Dummet, M. (1954). Review of Greak and Black's Traslation from the Philosophical writings of Gottlob Frege. *Mind*, 66: 102–105.

[Dupuy, 1994] Dupuy, J.-P. (1994). *Aux origines des aciences cognitives*. Editions la dècouverte, Paris.

[Engel et al., 1992] Engel, A., König, P., Gray, C., and Singer, W. (1992). Temporal coding by coherent oscillations as a potential solution to the binding problem: Physiological evidence. In [Schuster, 1992].

[Engel, 1989] Engel, P. (1989). *La norme du vrai*. Gallimard, Paris.

[Engel, 1992] Engel, P. (1992). Actions, raison et causes mentales. *Révue de théologie et de philosphie*, 124: 305–322. Actes du Colloque *Philosophie de l'action: ontologie et intentionalité*.

[Engel, 1994a] Engel, P. (1994a). *Davidson et la philosophie du langage*. Presses Universitaires de France, Paris.

[Engel, 1994b] Engel, P. (1994b). *Introduction à la philosophie de l'esprit*. Serie Sciences Cognitives. Ed. la découvert, Paris.

[Espinoza, 1990] Espinoza, M. (1990). *The four causes*, pages 171–190. In [Weingartner and Dorn, 1990].

[Feigl and Scriven, 1956] Feigl, H. and Scriven, M., editors (1956). *Minesota Studies in the philosophy of science: the foundations of science and the Concept of Psychology and Psychoanalysis*, volume I–II. Uoiversity of Minnesota Press, Minneapolis.

[Feigl et al., 1958] Feigl, H., Scriven, M., and Maxwell, G., editors (1958). *Minnesota Studies in philosophy of sciences: Concept, Theories and Mind–Body Problem*, volume II. University of Minnesota, Minneapolis.

[Feyerabend, 1963] Feyerabeud, P. K. (1963). *Mental events and the brain*, pages 204–205. In [Lycan, 1990].

[Fodor, 1968] Fodor, J. (1968). *Psychological explanation*. Random House, New York.

[Fodor, 1975] Fodor, J. (1975). *The language of thought*. Harvard Uoiversity Press, Cambridge.

[Fodor, 1978] Fodor, J. (1978). Tom Swift and bis procedural grandmotber. *Cognition*, 6: 229–247.

[Fodor, 1980] Fodor, J. (1980). Methodological Solipsism considered as a research strategy in cognitive psychology. *The behavioral and brain science*, 3: 63–109.

[Fodor, 1981a] Fodor, J. (1981a). *Representation*. Harvester Press, Brington.

[Fodor, 1981b] Fodor, J. (1981b). *Something on the State of the Art*, pages 1–31. In [Fodor, 1981a].

[Fodor, 1986] Fodor, J. (1986). *La modularité de l'esprit*. Ed. Minuit, Paris.

[Fodor, 1987] Fodor, J. (1987). *Psychosemantics*. MIT Press, Cambridge.

[Fodor, 1989a] Fodor, J. (1989a). Making mind matter more. In [Fodor, 1990], pages 137–159.

[Fodor, 1989b] Fodor, J. (1989b). Substitution arguments and tbe iodividualisation of belief. In [Fodor, 1990], pages 161–176.

[Fodor, 1990] Fodor, J. (1990). *A theory of content and other essays*. MIT Press, Cambridge.

[Fodor, 1991] Fodor, J. (1991). You can fool some of the people all the time, everyting else being equal; Hedged Laws and Psychological Explanation. *Mind*, 100: 19–34.

[Fodor, 1994] Fodor, J. (1994). *The elm and the Expert*. MIT Press, Cambridge.

[Fodor and LePore, 1992] Fodor, J. and LePore, E. (1992). *Holism. A shopper's guide*. Blackwel, Cambridge.

[Fodor and Pylysbyn, 1988] Fodor, J. and Pylyshyn, Z. (1988). Connectionism and cognitive arcbitecture: A critical analysis. *Cognition*, 28: 3–71.

[Fodor, 1977] Fodor, J. D. (1977). *Semantics. Theories of Meaning in Generative Grammar*. The Harvester Press, Sussex.

[Føllesdal, 1982a] Føllesdal, D. (1982a). Brentano and Husserl. In [Dreyfus, 1982b], pages 31–41.

[Føllesdal, 1982b] Føllesdal, D. (1982b). The notion of *Noema*. In [Dreyfus, 1982b], pages 73–80.

[Fuks and Rokhlin, 1984] Fuks, D. B. and Rokhlin, V. A. (1984). *Beginner's Course in Topology*. Springer–Verlag, Berlin.

[Galifret, 1990] Galifret, Y. (1990). *Les mécanismes de la vision*. Pour la Science, Paris.

[Gibson, 1950] Gibson, J. J. (1950). *The perception of the visual world*. Greenwood Press, Connecticut.

[Gibson and Dibble, 1952] Gibson, J. J. and Dibble, F. (1952). Explanatory experiments on the stimulus conditions of a visual surface. *Journal of experimental psychology*, 43: 415–419.

[Girard, 1985] Girard, J.-Y. (1985). Théorie de la démostration. In *Encyclopædia Universalis*, volume Corpus 5, pages 1101–1105. Encyclopædia Universalis, Paris.

[Glansdorff and Prigogine, 1985] Glansdorff, P. and Prigogine, I. (1985). Thermodynamique. In *Encyclopædia Universalis*, volume Corpus 17, pages 1163–1166. Encyclopædia Universalis, Paris.

[Globus et al., 1976] Globus, G. G., Maxwell, G., and Sovodnik, I. (1976). *Consciousness and the brain. A Scientific and Philosophical Inquiry*. Plenum Press, New York.

[Gochet and Gribomont, 1990] Gochet, P. and Gribomont, P. (1990). *Logique*, volume I, II. Hermès, Paris.

[Goldman, 1967] Goldman, A. (1967). A Causal theory of knowing. *The journal of philosophy*, LXIV(12): 375–372.

[Guelat, 1980] Guelat, J. (1980). Théorie des catastrophes et applications. Master's thesis, Institut de mathèmatique et d'Informatique. Université de Neuchâtel. Sous la direction du Professeur François Sigrist.

[Gurwitsch, 1957] Gurwitsch, A. (1957). *Théorie du champ de la conscience*. Desclèe de Brouwer, Burge. Il existe une version anglaise posterieure à la française (1962) *The field of the consciousness*, Dusquesne University Press, Pittsburgh.

[Guttenplan, 1994] Guttenplan, S., editor (1994). *A Companion to the Philosophy of Mind*. Blackwell Companions to Philosophy. Basi Blackwell Ltd., Oxford.

[Hall, 1982] Hall, H. (1982). Was Husserl a Realist or an Idealist? In [Dreyfus, 1982b], pages 189–190.

[Hare, 1952] Hare, R. M. (1952). *The language of morals*. Oxford University Press, Oxford.

[Haugeland, 1981] Haugeland, J. (1981). *Mind Design: Philosophy, Psychology, Artificial Intelligence*. MIT Press, Cambridge.

[Heil and Mele, 1993] Heil, J. and Mele, A., editors (1993). *Mental Causation*. Clarendon Press, Oxford.

[Hempel, 1949] Hempel, C. (1949). *The logical analysis of Psychology*, pages 14–23. Volume I et II of [Block, 1980a].

[Hempel, 1988] Hempel, C. (1988). Provisos: A problem concerning the inferential function of scientific theories. In Grünbaun, A. and Salmon, W., editors, *The limitations of Deductivism*. U. of California Press, California.

[Hempel and Oppenheim, 1948] Hempel, C. G. and Oppenheim, P. (1948). Logic of explanation. *Philosophy of* science, 15: 135–175.

[Hintikka, 1996] Hintikka, J. (1995). Husserl: la dimension phénoménologique. *Les études philosophiques*, (1): 37–84.

[Hodges, 1988] Hodges, A. (1988). *Alain Turing ou l'énigme de l'intelligence artificielle*. Payot, Paris.

[Hook, 1960] Hook, S., editor (1960). *Dimension of mind*. New York University Press, New York.

[Hubel and Wiesel, 1990] Hubel, D. and Wiesel, T. (1990). Les mécanismes cérébraux de la vision. In [Galifret, 1990], pages 108–130.

[Husserl, 1950] Husserl, E. (1950). *Idées directrices pour une phénoménologie*. Gallimard, Paris. Traduction française Paul Ricœur.

[Husserl, 1963] Husserl, E. (1963). *Recherches Logiques*. PUF, Paris. Traduction française Élie, H., Kelkel, A. et Schérer, R.

[Husserl, 1985] Husserl, E. (1985). *L'Idée de la phénoménologie*. PUF, Paris. Traduction française Lowit.

[Jackson, 1982] Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quartely*, 32: 127–136. Aussi dans Lycan, William (Ed.) (1990) pages 469-477.

[Jacob, 1980] Jacob, P. (1980). *L'empirisme logique*. Ed. de Minuit, Paris.

[Jacob, 1992a] Jacob, P. (1992a). *Cerveau, esprit et representation mentale*, pages 313–351. In [Andler, 1992].

[Jacob, 1992b] Jacob, P. (1992b). Proprietés mentales et explication causale. *Revue de Métphysique et de Morale*, 2: 295–303.

[Jacob, 1993] Jacob, P. (1993). Fodor, la psychologie scientifique et les attributions de croyances. *Philosophiques*, XX(1): 131–158.

[Jacob, 1995] Jacob, P. (1995). Un réalisme intentionnel est–il condamné à l'atomisme sémantique? Technical Report 9513, Ecole Polytechnique–CREA.

[Janet, 1901] Janet, P. (1901). *Les Causes Finales*. Felix Alcan, Paris.

[Kant, 1975] Kant, E. (1975). *Critique à la raison pure*. PUF.

[Kim, 1990] Kim, J. (1990). Supervenience as a philosophical concept. *Metaphilosophy*, 21(1): 1–27. Aussi dans Kim, J.(1993) pages 131–160.

[Kim, 1991] Kim, J. (1991). Dretske on bow reasons explain behavior. In [McLaughlin, 1991], pages 52–73.

[Kim, 1992a] Kim, J. (1992a). *"Downward Causation" Emergentism and Nonreductive Physicalism*, pages 119–138. In [Beckermann et al., 1992].

[Kim, 1992b] Kim, J. (1992b). Multiple Realisation and tbe Metaphysics of Reduction. *Philosophy and Phenomenological Research*, 2(1): 1–26.

[Kim, 1992c] Kim, J. (1992c). *Supervience and Mind*. Cambridge Studies in Philosophy. Cambridge Universty Press, Cambridge.

[Kim, 1993a] Kim, J. (1993a). Can Supervenience save anomalous Monisme. In [Heil and Mele, 1993], pages 19–26.

[Kim, 1993b] Kim, J. (1993b). Tbe Non–reductivist's Troubles with Mental Causation. In [Heil and Mele, 1993], pages 189–210.

[Kim, 1993c] Kim, J. (1993c). "Strong" and "Global" supevenience revisited. In [Kim, 1992c], pages 79–91.

[Kim, 1993d] Kim, J. (1993d). "Supervenience " for multiple domains. In [Kim, 1992c], pages 109–130.

[Kim, 1994] Kim, J. (1994). Supervinience. In [Guttenplan, 1994], pages 575–583.

[Kim and Sosa, 1995] Kim, J. and Sosa, E., editors (1995). *A Companion to Metaphysics*. Blackwell Companions to Philosophy. Basi Blackwell Ltd., Oxford.

[Kistler, 1995a] Kistler, M. (1995a). *Causalité, loi, représentation*. PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.

[Kistler, 1995b] Kistler, M. (1995b). Toutes les lois sont-elles *certeris paribus?*

[Klibansky and Pears, 1993] Klibansky, R. and Pears, D. (1993). *La philosophie en Europe*. Gallimard, Paris.

[Kripke, 1971] Kripke, S. (1971). *Meaning and Necessity*, pages 43–45. In [Beakley and Ludlow, 1992].

[Kripke, 1980] Kripke, S. (1980). *Naming and Necessity*. Basil Blackwell, Oxford.

[Kubn, 1962] Kuhn, T. S. (1962). *International Encyclopedia of Unified Science*, chapter Tbe Structure of Scientific Revolutions. Chicago University Press, Chicago, 1970 deuxième edition.

[Lalande, 1985] Lalande, A. (1985). *Vocabulaire Technique et Critique de la Philosophie*. Editions Etudes Vivantes, Paris.

[Langton, 1986] Langton, C. (1986). Studying Artificial Life with Cellular Automata. *Physica*, D 22: 120–149.

[Langton, 1989] Langton, C. G., editor (1989). *Artificial Life*. Addison Wesley, Redwood City, California.

[Langton, 1995] Langton, C. G., editor (1995). *Artificial Life. A overwiew*. MIT Press, Cambridge, Massacbusetts.

[Langton et al., 1991] Langton, G. C., Taylor, C., Farmer, J. D., and Rasmussen, S. (1991). *Artificial Life III. Procedings of the workshop on artificie al intelligence held february, 1990 in Santa Fe, New Mexico*. Addison Wesley, Redwood City, California.

[Lashley, 1929] Lashley, K. (1929). *Brain, Mechanisms and Intelligence*. University of Chicago Press, Chicago.

[Leibniz, 1992] Leibniz, G. W. (1992). *The nature and Communication of Substances*, chapter 15, pages 119–121. In [Beakley and Ludlow, 1992]. Loemker, L.(ed.) (1956) Philosophical Papers and Letters of Leibniz, vol. 2. Chicago, University of Chicago Press.

[LePore and Loewer, 1989] LePore, E. and Loewer, B. (1989). More on Making Mind Matter. *Philosophical Topics*, 17: 175–191.

[LePore and McLaugbing, 1985] LePore, E. and McLaughing, B. P., editors (1985). *Actions and Events: Perspectives on the Philosophy of Donald Davidson*. Basil Blackwel, Oxford.

[Lewes, 1875] Lewes, G. H. (1875). *Problems of life and mind*. Kegan Paul, Trench, Turbner AND Co., London.

[Lewis, 1972] Lewis, D. (1972). Psychophysical and Tbeoretical Identifications. In [Block, 1980a], pages 249–258. Première publicatioo : (1972) *Australian Journal of Philosophy*, 50.

[Lewis, 1994] Lewis, D. (1994). Lewis, David: Reduction of Mind. In [Guttenplan, 1994], pages 412–431.

[Loewer and Rey, 1991] Loewer, B. and Rey, G., editors (1991). *Meaning in Mind: Fodor and his Critics*. Blackwell, Oxoford.

[Lubow, 1989] Lubow, R. E. (1989). *Latent inhibition and conditioned attention theory.* Cambridge University Press, Cambridge.

[Lycan, 1990] Lycan, W. G. (1990). *Mind and Cognition.* Ed. Blackell, Cambridge.

[Macdonald and Macdonald, 1994] Macdonald, C. and Macdonald, G., editors (1994). *The Philosophy of Psychology: Debates on Psychological Explanation.* Basil Blackwell, Oxford.

[Malebranche, 1963] Malebranche, N. (1963). *De la Recherche de la verité.* Librairie philosopbique J. Vrin, Paris. Rodis-Levis, G. (editor).

[Malebranche, 1980] Malebranche, N. (1980). *The Union of Soul and Body,* chapter 18, pages 115–118. Io [Beakley and Ludlow, 1992]. Repris de Lenon, Tbomas and Olscamp, Paul (1980) The Search after Trutb, Cambridge, Cambridge University Press.

[Marr, 1969] Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology,* 202: 437–470.

[Marr, 1978] Marr, D. (1978). Early processing of visual information. *Phil. Trans. R. Soc. Lond. B,* 275: 484–519.

[Marr, 1977] Marr, D. (1977). Artificial intelligence–A personal view. *Artificial Intelligence,* 9: 37–48.

[Marr, 1980] Marr, D. (1980). *Vision.* Freedman Co., New York.

[McIntyre, 1982] McIntyre, R. (1982). Intending and Refering. In [Dreyfus, 1982b], pages 215–233.

[McLaughlin, 1991] McLaughlin, B., editor (1991). *Dretske and his critics.* Basil Blackwell, Cambridge.

[McLaughlin, 1993] McLaughlin, B. L. (1993). On Davisons' response to the Charge of Epiphenomenalisme. In [Heil and Mele, 1993].

[McLaughlin, 1992] McLaughlin, B. P. (1992). *The rise and the fall of British Emergentism,* pages 49–91. In [Beckermann et al., 1992].

[McMullin, 1995] McMullin, E. (1995). Matter. In [Kim and Sosa, 1995], pages 299–302.

[Meehl and Sellars, 1956] Meehl, P. E. and Sellars, W. (1956). *The concept of Emergence,* pages 239–252. Volume I of [Feigl and Scriven, 1956].

[Miéville, 1991] Miéville, D. (1991). *Introduction à la théorie de systèmes formelles.* Centre de Recherches sémiologiques. Université de Neuchâtel, Neuchâtel.

[Mill, 1843] Mill, J. S. (1843). *System of logic.* Longmans, Green, Reader, London. 8 ème Edition: 1872.

[Milnor, 1985] Milnor, J. W. (1985). *From the differential viewpoint.* The Virginia Press of Virginia, Cbarlottesville.

[Minsky, 1968] Minsky, M. L., editor (1968). *Semantic Information Processing.* MIT Press, Cambridge.

[Mohanty, 1982] Mohanty, J. N. (1982). Busserl and Frege: A new look at their relationship. In [Dreyfus, 1982b], pages 43–59.

[Morgan, 1923] Morgan, C. L. (1923). *Emergent Evolution.* Williams and Norgate, Londo.

[Morlet, 1985] Morlet, C. (1985). Topologie differentielle. In *Encyclopædia Universalis,* volume Corpus 18, pages 79–86. Encyclopædia Universalis, Paris.

[Moya, 1990] Moya, C. (1990). *The philosophy of action.* Polity Press, Cambridge.

[Nagel, 1961] Nagel, E. (1961). *The structure of Sciences.* Ed. Harcourt Brace World, New York.

[Nagel, 1974] Nagel, T. (1974). *What is it like to be a bat?* , pages 159–188. Volume 1 of [Hlock, 1980a].

[Newell, 1980] Newell, A. (1980). Physical symbols systems. *Cognitive Science*, 4: 135–183.

[Newell, 1982] Newell, A. (1982). The knowledge level. *Artifical Intelligence*, 18: 87–127.

[Newell and Simon, 1976] Newell, A. and Simon, H. (1976). Computer science as empirical inquiry : Symbols and search. In [Haugeland, 1981], pages 35–68.

[Nicolis and Prigogine, 1985] Nicolis, G. and Prigogine, I. (1985). *Exploring complexity.* Piper, Muchen.

[Nussbaum, 1995] Nussbaum, M. (1995). Hylomorphism. In [Kim and Sosa, 1995], pages 221–222.

[Oguien, 1994] Oguien, R. (1994). L'action. Séminaire informel de recherche du Séminaire de Philosophie de l'Université de Neuchâtel.

[Oppenheim and Putnam, 1958] Oppenheim, P. and Putnam, H. (1958). *Unity of science as a Working hypothesis*, pages 3–36. Volume II of [Feigl et al., 1958].

[Pacherie, 1992] Pacherie, E. (1992). *Perspectives physicalistes sur l'intentionnalité.* PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris.

[Pacherie, 1993] Pacherie, E. (1993). *Naturaliser l'intentionnalité.* Presses Universitaires de France, Paris.

[Pap, 1951] Pap, A. (1951). The concept of absolute emergence. *The british journal of Philosophy of Sciences*, 2, mai 1951–février 1952: 302–311.

[Pattee, 1989] Pattee, H. (1989). *Simulation, realisation and theories of life*, pages 63–77. In [Langton, 1989].

[Pélissier and Tête, 1995] Pélissier, A. and Tête, A. (1995). *Sciences cognitives: Textes fondateurs.* PUF, Paris.

[Pepper, 1926] Pepper, S. C. (1926). Emergence. *Jornal of philosophy*, 23: 241–245.

[Petitot, 1992] Petitot, J. (1992). *Physique du sens.* Edition du CNRS, Paris.

[Petitot, 1993] Petitot, J. (1993). Phenomenology of perception, qualitative physics and sheaf mereology. In *16th. International Wittgenstein Symposium "Philosophy and the Cognitive Science"*, Kirchberg–Wechsel.

[Petitot, 1994a] Petitot, J. (1994a). *La sémiophysique : de la physique qualitative aux sciences cognitives*, pages 499–545. In [Porte, 1994].

[Petitot, 1994b] Petitot, J. (1994b). Sheaf mereology and space. Presenté au First International Summer Institute in Cognitive Sciences. 9–10 July 1994.

[Petitot and Smith, 1994] Petitot, J. and Smith, B. (1994). Physics and the phenomenal world. In Poli, R. and Simons, P. M., editors, *Formal ontology.* Kluwer, London.

[Philipse, 1995] Philipse, H. (1995). Transcendental Idealism. In [Smith and Smith, 1995], pages 239–322.

[Place, 1956] Place, U. T. (1956). *Is consciousness a Brian Process?* , pages 33–39. In [Beakley and Ludlow, 1992].

[Porte, 1994] Porte, M., editor (1994). *Passion des formes à René Thom.* ENS Editions, Fontenay–St Cloud.

[Poston and Stewart, 1978] Poston, T. and Stewart, I. (1978). *Catastraphe Theory and its Applications.* Pitman, London.

[Prigogine, 1980] Prigogine, I. (1980). *From Being to Becomming.* Freeman, San Francisco.

[Proust, 1982] Proust, J. (1982). *Preface de la traduction française*, pages 7–30. In [Searle, 1979]. Trad. française Proust, Joëlle (1982) Sens et expression. Paris, Ed. de Minuit.

[Proust, 1990] Proust, J. (1990). De la difficulté d'être naturaliste en matière d'Intentionnalité. *Revue de synthèse*, IV: 13–32.

[Proust, 1993] Proust, J. (1993). *Le fonctionnalisme et les limite de la multiréalisabilité: des interaction entre structure et fonction*, pages 641–670. In [Klibansky and Pears, 1993].

[Putnam, 1960] Putnam, H. (1960). *Minds and Machines*, pages 138–164. In [Hook, 1960]. Aussi dans Putnam (1975) pages 362–385.

[Putnam, 1965] Putnam, H. (1965). *Brains and Behavior*, chapter 2, pages 24–35. Volume I et II of [Hlock, 1980a].

[Putnam, 1967a] Putnam, H. (1967a). *The nature of Mental States*, pages 47–54. In [Lycan, 1990]. Aussi dans Putnam(1975) pages 429–440.

[Putnam, 1967b] Putnam, H. (1967b). *Psychological Predicates.* In [Capitan and Merrill, 1967].

[Putnam, 1973] Putnam, H. (1973). *Philosophy and Our Mental Life*, pages 291–303. Volume *II* of [Putnam, 1975b].

[Putnam, 1975a] Putnam, H. (1975a). *The Meaning of 'Meaning'*, pages 215–271. Volume *II* of [Putnam, 1975b].

[Putnam, 1975b] Putnam, H. (1975b). *Mind, Language and Reality, Philosophical Papers*, volume *II*. Cambridge University Press, Cambrigde.

[Putnam, 1985] Putnam, H. (1985). *After empiricism*, pages 20–30. In [Rajchman and Cornel, 1985].

[Putnam, 1988] Putnam, H. (1988). *Representation and Reality.* MIT Press., Cambridge. Trad. française Tiercelin, Claudine (1990). *Représentation et Réalité.* Paris, Gallimard.

[Putnam, 1994] Putnam, H. (1994). *Putnam, Hilary*, pages 507–513. In [Guttenplan, 1994].

[Pylyshyn, 1980] Pylyshyn, Z. (1980). Computer and cognition. *The behavioral and brain science*, 3: 111–169.

[Pylyshyn, 1984] Pylyshyn, Z. (1984). *Computer and Cognition.* MIT Press, Cambridge.

[Quine, 1981] Quine, W. (1981). *Theories and Things.* The belknap Press of Harvard University Press, Cambridge.

[Quine, 1955] Quine, W. V. (1955). On Frege's way out. *Mind*, 72(254): 145–159.

[Quine, 1960] Quine, W. V. (1960). *Word and Objet.* MIT Press, Cambridge. Trad. française de Joseph Dopp et Paul Gochet (1977). *Le mot et la chose.* Paris, Flammarion.

[Quioe, 1963] Quine, W. V. (1963). *From a logical point of view.* Harper and Row, New York.

[Rajchman and Cornel, 1985] Rajchman, J. and Cornel, W., editors (1985). *Post–analytic philosophy.* Columbia University Press, New York.

[Rasmussen, 1991] Rasmussen, S. (1991). *Aspect of Information, life, reality and physics*, pages 767–773. In [Langton et al., 1991].

[Récanati, 1992] Récanati, F. (1992). *Contenu sémantique et contenu cognitif des énoncés*, pages 238–270. In [Andler, 1992].

[Renaut, 1993] Renaut, A. (1993). *Sartre, le dernier philosophe*. Ed. Grasset, Paris.

[Rey, 1994] Rey, G. (1994). *Concepts*, pages 185–193. In [Guttenplan, 1994].

[Rivenc, 1995] Rivenc, F. (1995). Husserl avec et contre Frege. *Les études philosophiques*, 1: 13–38.

[Rodriguez, 1994] Rodriguez, M. (1994). *Approche Constructiviste de l'architecture de contrôle et de la représentation des connaissances*. PhD thesis, IIIA. Université de Neuchâtel.

[Russell, 1900] Russell, B. (1900). *A Critical Examination of the Philosophy of Leibniz*. Cambridge University Press, Cambridge.

[Ruesell, 1903] Russell, B. (1903). *The Principles of Mathematics*. Cambridge University Press, Cambridge.

[Russell, 1940] Russell, B. (1940). *An Inquiry into Meaning and Truth*. Penguin Books Ltd., Middlesex, septième (1973) edition.

[Russell, 1953] Russell, B. (1953). *Philosophica Investigations*, chapter Introduction aux *Tractatus*, pages 1–21. In [Wittgenstein, 1953]. Trad. française Klossowski, Pierre (1961) Paris, Ed. Gallimard.

[Ryle, 1949] Ryle, G. (1949). Descates' Myth. In [Beakley and Ludlow, 1992], pages 23–31.

[Scaglione, 1989] Scaglione, M. (1989). La controverse Fodor - Jobnson Laird. Ecoles des Hautes Etudes en Science Sociales.

[Scaglione, 1991] Scaglione, M. (1991). Théories de la vision. la dymanique perception–action en intelligence artificielle. Lausanne. EPFL–IREC.

[Schaffer, 1968] Schaffer, J. A. (1968). *Philosophy of Mind*. Prentice-Hall Ed., London.

[Schiffer, 1991] Schiffer, S. (1991). Certeris Paribus Laws. *Mind*, 100: 1–17.

[Schuster, 1992] Schuster, H., editor (1992). *Non linear dynamics and Neural Network*. Springer, Berlin.

[Searle, 1979] Searle, J. (1979). *Expression and Meaning*. Cambridge University Press, Cambridge. Trad. française Proust, Joëlle (1982) Sens et expression. Paris, Ed. de Minuit.

[Searle, 1983] Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, Cambridge. Trad. française Pichevin, Claude . *L'intentionalité*. Paris, Ed. de Minuit.

[Searle, 1994a] Searle, J. (1994a). *Intentionality*, pages 379–386. In [Guttenplan, 1994].

[Searle, 1994b] Searle, J. (1994b). *The rediscovery of the Mind*. MIT Press, Cambridge.

[Smart, 1959] Smart, J. J. C. (1959). Sensation and Brain Processes. *Philosophical Review*, 68: 141–156.

[Smart, 1981] Smart, J. J. C. (1981). Physicalism and Emergence. *Neuroscience*, 6: 109–113.

[Smith, 1994] Smith, B. (1994). Fiat objects. Presenté dans le cours de *Foundation of cognition* pendant le FISI–CS à l'Université de l'État de New York en Buffalo.

[Smith and Smith, 1995] Smith, B. and Smith, D. W. (1995). *The Cambrige Companion to Husserl.* Cambridge Universty Press, Cambridge.

[Smolensky, 1994] Smolensky, P. (1994). *Constituent structure and explanation in an integrated connectionist–Symbolic cognitive architecture.* In [Macdonald and Macdonald, 1994].

[Sober, 1991] Sober, E. (1991). *Learning from functionalism–Prospects for Strong Artificial Life,* pages 749–765. In [Langton et al., 1991].

[Sosa, 1993] Sosa, E. (1993). Davidson's Thinking Causes. In [Heil and Mele, 1993].

[Sperry, 1980] Sperry, R. W. (1980). Mind–brain interaction: mentalism, yes: dualisme, no. *Neurascience,* 5: 195–206.

[Sperry, 1986] Sperry, R. W. (1986). Macro– versus micro–determinisme. *Philosphy of Science,* 53: 285–270.

[Steels, 1995] Steels, L. (1995). *The Artificial Life Roots of Artificial Intelligence,* pages 75–110. In [Langton, 1995].

[Teller, 1992] Teller, P. (1992). *A contemporary look at Emergence,* pages 139–153. In [Beckermann et al., 1992].

[Thom, 1980] Thom, R. (1980). *Modéles mathématiques de la morphongenèse.* Christia Bourgois Ed., Paris.

[Thom, 1983] Thom, R. (1983). *Paroboles et catastrophes.* Flammarion, Paris.

[Thom, 1991] Thom, R. (1991). *Prédire n'est pas expliquer.* Flammarion, Paris.

[Toffoli, 1982] Toffoli, T. (1982). Physics and computation. *Int. J. Theor. Phys.,* 21(3–4): 165–175.

[Turing, 1950] Turing, A. (1950). Computating machinery and intelligence. *Mind,* 59: 433–80. Trad. franaise dans [Pélissier et Tête, 1995] pages 247–295.

[Ullman, 1979] Ullman, S. (1979). *The Interpretation of Visual Motion.* MIT Press, Cambridge.

[Varela and Bourgine, 1991] Varela, F. and Bourgine, P. (1991). *Toward a proctice of autonomous systems. Proceedings of the first European Conference on Artificial Intelligence.* MIT Press, Cambridge, MA.

[Varela et al., 1993] Varela, F., Thompson, E., and Rosch, E. (1993). *The embodied Mind.* MIT Press, Cambridge.

[Weingartner and Dorn, 1990] Weingartner, P. and Dorn, G. J. W., editors (1990). *Studies on Mario Bunge's Treatise.* Rodopi, Amsterdam.

[Wimsatt, 1976] Wimsatt, W. C. (1976). *Reduction, levels of organisation, and the Mind–Body Problem,* chapter 8, pages 205–267. In [Globus et al., 1976].

[Winograd and Flores, 1989] Winograd, T. and Flores, F. (1989). *L'intelligence artificielle en question.* PUF, Paris.

[Wittgenstein, 1921] Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus.* Edition allemande dans les *Annalen der Naturphilosophie* edition. Trad. française Klossowski, Pierre (1961) Paris, Ed. Gallimard.

[Wittgenstein, 1953] Wittgenstein, L. (1953). *Philosophica Investigations.* Trad. française Klossowski, Pierre (1961) Paris, Ed. Callimard.

[Woods, 1975] Woods, W. (1975). Foundations for semantics networks. pages 1–53. Academic Press. In Bobrow, D. and Collons, A. editor (1975) *Representation and undestanding.*