

Journées de Rochebrune
Rencontres interdisciplinaires sur les systèmes
complexes naturels et artificiels.

Edition 1998 : du 1 au 7 février

Avec la participation d'Isabelle Stengers

Quelles relations entretenons-nous avec nos modèles ?

Les modèles et leur élaboration sont au coeur de l'activité scientifique moderne. La physique, la biologie, l'économie, les sciences de la cognition, pour ne citer que quelques exemples, se structurent autour d'une grande variété de modèles, que les scientifiques contestent, interrogent, modifient, raffinent ou créent de toutes pièces. Nous proposons de nous interroger sur les relations que nous entretenons avec ces objets très particuliers que sont les modèles :

Pourquoi les créons-nous ? Quelles épreuves leur imposons-nous ?
Comment les faisons-nous évoluer ? Quelles limites intrinsèques peuvent-ils avoir ?

Comment leur accordons nous notre confiance ?

Dans quelle mesure croyons-nous en eux ?

Comment en changeons-nous ?

Comment les utilisons-nous ? Sont-ils des instruments de pouvoir ? Nous servent-ils à exclure d'autres pratiques ?

En quoi la croyance dans nos modèles nous transforme-t-elle ? En quoi ces pratiques d'élaboration de modèles nous distinguent-elles des non-modernes ?

Engager une réflexion approfondie sur ces questions est un enjeu important pour mieux comprendre les pratiques scientifiques et les améliorer, mais aussi pour rendre possible les interactions et les échanges avec d'autres types de pratiques.

Comité d'organisation

**Christian Brassac
Eric Dedieu
Guillaume Deffuant
Nils Ferrand
Herve Glotin
Christophe Parisse**

Comité de Programme

**Hugues Bersini
Jean-Bernard Billeter
Paul Bourgina
Christian Brassac
Michele Courant
Guillaume Deffuant
Jean-Louis Dessalles
Wolf Eberwein
Nils Ferrand
Thierry Fuhs
Jean Claude Heudin
Dominique Luzeaux
Anne Nicolle
Pierre Yves Raccah
Jean Sallantin
John Stewart
Martine Timsit**

Avec le soutien de :

**l'European Conference on Artificial Life (ECAL)
la Direction Générale pour l'Armement (DGA)**

Sous le patronage de :

l'Association pour la Recherche Cognitive (ARC)

Programme

Lundi

- Christian Brassac** p 7
Modèle(s) de l'intercompréhension
- Olivier Simonin et Jacques Ferber** p 29
Analyse des performances des systèmes collectifs
- Pierre Saurel** p 45
La modélisation ne se réduit ni à l'étude des modèles ni à l'exploration de leurs simulations
- Danièle Bourcier** p 61
Représenter quelques règles de droit dans une machine. Au delà de Faust et de Frankenstein
- Bernard Morand** p 77
Le diagramme, le modèle et le signe

Mardi

- Hugues Bersini** p 95
Conscience avec sciences
- Anne Nicolle**
Modélisation des phénomènes complexes par des logiciels d'étude
- Dominique Luzeaux** p 107
Modèle, théorie, langage : le point de vue de la logique mathématique
- Jacques Blanc-Talon** p 119
Modèles abstraits en perception artificielle

Herminia Peraita et José Luis Dobato p 131
Modèles neurophysiologiques sur la détérioration de la
mémoire conceptuelle chez les patients d'Alzheimer :
confrontation de deux modèles

Mercredi

Nicolas Grégori p 141
Modèles intermédiaires et concepts scientifiques

Gérard Collet p 161
Point de vue linguistique sur la modélisation

Catherine Fayada p 181
Modélisation et pratique en psychopathologie neurologique

Jeudi

Jean-Pierre Rossi p 197
La notion de modèle et son utilisation en Psychologie

Miriam Edith Scaglione p 219
Les modèles de la Vie Artificielle et le(s) concept(s)
d'émergence : de la métaphore à la réalisation

Fabienne de Gaulejac et Alain Gallo p 237
Le paradoxe du psychologue : de l'usage d'un référentiel à la
construction du référentiel du sujet d'étude

Pek Van Andel p 247
A model of serendipity

Martine Timsit, Christophe Parisse et Evelyne Andreewsky p 283
Psychopathologie, modèles et pouvoir

Vendredi

Guillaume Deffuant p 293
L'autonomie comme approche de nos interactions avec nos
modèles : la clôture selon F. Varela ou selon C. Castoriadis ?

Index des auteurs

Evelyne Andreewsky	p 281
Hugues Bersini	p 95
Jacques Blanc-Talon	p 119
Danièle Bourcier	p 61
Christian Brassac	p 7
Gérard Collet	p 161
Fabienne de Gaulejac	p 237
Guillaume Deffuant	p 291
José Luis Dobato	p 131
Catherine Fayada	p 181
Jacques Ferber	p 29
Alain Gallo	p 237
Nicolas Grégori	p 141
Dominique Luzeaux	p 107
Bernard Morand	p 77
Anne Nicolle	-
Christophe Parisse	p 281
Herminia Pereira	p 131
Jean-Pierre Rossi	p 197
Pierre Saurel	p 45
Miriam Edith Scaglione	p 219
Olivier Simonin	p 29
Martine Timsit	p 281
Pek Van Andel	p 247

Modèle(s) de l'intercompréhension

Christian BRASSAC

*Laboratoire de Psychologie de l'Interaction
Groupe de Recherches sur les Communications
Université Nancy 2 — BP 33-97
54015 Nancy cedex
brassac@clsh.u-nancy.fr*

Introduction

La modélisation des phénomènes conversationnels constitue un point qui divise la communauté des chercheurs participant au développement de l'analyse des conversations. Au-delà de cette question, nous développerons dans cette réflexion la position que nous adoptons face à l'élaboration d'un modèle des processus interlocutoires. Envisager cette position sous tous ses angles nous aura amené à rencontrer l'interrogation qui constitue le thème de ces journées, « Quelles relations entretenons-nous avec nos modèles ? ».

De fait nos travaux nous ont conduit à proposer un modèle de l'intercompréhension interhumaine. Fondé sur des catégories provenant d'une part de la philosophie du langage et, d'autre part, des logiques non monotones, il prétend constituer une description formalisée des processus cognitifs et sociaux qui se déploient dans l'interaction verbale de face à face. J'ai bien écrit : il prétend. La question bien sûr est celle de la validité de ce modèle. C'est autour de cette question que nous allons articuler l'exposé.

L'attitude scientifique qui est la nôtre est celle d'un psychologue de l'interaction. L'objet est principalement constitué par les mécanismes cognitifs engendrés dans et par l'interlocution. Ces mécanismes sont envisagés comme des activités humaines, ancrées dans la situation intersubjective. Ainsi donc, et tout naturellement, nous convoquerons la notion de modèle telle qu'elle est travaillée en psychologie cognitive. En examinant les différents types de modèles, leurs statuts respectifs, leurs efficiences méthodologiques, nous « projeterons » ces définitions sur l'interrogation qui nous occupe. Nous montrerons par exemple en quoi les modèles de simulation élaborés en psychologie cognitive peuvent nous venir en aide pour envisager la validité de notre modèle.

Pour ce faire nous commencerons par exposer notre vision de ce processus que nous tenons pour central dans toute relation intersubjective, l'intercompréhension (**1. L'intercompréhension : un processus "distribué"**). Dans un second temps, nous invoquerons les différents types de modèles que les psychologues mobilisent pour approcher les phénomènes

cognitifs humains (2. **Modéliser l'intercompréhension ; avec quels outils ?**). Nous montrerons en quoi le modèle de l'enchaînement conversationnel que nous prônons nous semble répondre aux critères que ces derniers imposent en la matière (3. **Modéliser l'enchaînement conversationnel**). Nous terminerons en interrogeant la validité de notre modèle à l'aide de la simulation informatique de conversations (4. **L'intercompréhension : l'objet de simulation ?**).

1. **L'intercompréhension : un processus "distribué"**

1.1. *La communicativité : un paradigme dominant*

Les travaux classiques (Trognon, 1986) consacrés à l'étude des échanges langagiers interhumains sont tous fondés sur un postulat de départ :

Lorsqu'un énoncé est produit par un des locuteurs, il est *traité* par l'auditeur.

Ce traitement est le fait du seul auditeur. Son objectif est de trouver, de découvrir, de détecter le contenu que le locuteur a pour intention de communiquer. Les modèles codiques et inférentiels de la communication ont pour ambition de rendre compte du mode de traitement mis en œuvre par l'auditeur. Qu'il soit de l'ordre du décodage ou du travail inférentiel, le traitement s'ancre sur l'énoncé produit. La mécanique est la suivante :

1. Le locuteur possède une certaine intention communicative.
2. Il produit un énoncé porteur de sens.
3. L'auditeur perçoit cet énoncé et le "traite".
4. A l'issue du traitement, il récupère, il trouve, il atteint le sens intentionné.

Le traitement est "bon" lorsque l'on observe une certaine relation entre le sens intentionné et le sens découvert. Remarquez qu'il n'est pas dit qu'il s'agit là du sens littéral ou non littéral de l'énoncé. Le travail peut concerner, sauf pour le modèle codique mais peu importe, les actes indirects, les implicatures conversationnelles, etc. L'implicite est évidemment dans la portée du traitement de ce sens préexistant. On dit alors que les interlocuteurs "se comprennent". Tout le monde est content : le locuteur a communiqué ostensiblement un sens, l'auditeur a découvert ce qu'il fallait : ils se sont compris.

Quel que soit le mode de traitement, cette façon de voir place au centre du dispositif d'intercompréhension l'intention de sens du locuteur, **préexistante au travail d'interprétation-décodage de l'auditeur**. Cette façon de voir ressort du paradigme de la communicativité que Francis Jacques (1985) oppose à celui de la communicabilité.

1.2. La communicabilité : un paradigme concurrent

On peut concevoir un paradigme de la communicabilité qui va à l'encontre de cette vision. Il place au centre de ses préoccupations une intercompréhension vue comme une co-construction processuelle d'un sens des énoncés successifs, produisant une **cognition distribuée**. Dans cette optique, on ne considère pas la conversation comme le lieu où s'organise l'adéquation de deux cognitions individuelles. Elle est conçue comme le lieu de l'**intercommunicabilité des cognitions** en présence (l'expression est d'Alain Trognon, 1991). L'intention de sens est une production conjointe, distribuée sur l'espace intersubjectif que constitue l'entre-deux conversationnel (Jacques, 1982, 1985).

Il n'y a pas un sens qui est communiqué, l'énoncé produit initialement est simplement porteur d'un potentiel de sens. L'échange se déroule autour de l'actualisation d'un élément de ce potentiel, actualisation qui est le produit de la dyade, qui est l'œuvre conjointe des interactants. Voyons sur un exemple (que nous avons déjà traité dans Brassac et Stewart, 1996).

Soit l'extrait conversationnel suivant (Dessalles, école de l'ARC de Bonas, 1995).

- t₀ E₁ Oh c'est marrant ! J'ai exactement ce tableau chez moi, il a la même taille, il représente la même chose
t₁ L₁ On te l'a peut-être volé
t₂ E₂ Non le mien, il est plus sombre, il est plus beau.

Voici clairement ce que nous aimerions montrer :

- a. E₁ est une forme linguistique, porteuse d'un potentiel de sens ;
- b. L₁ est une proposition d'actualisation d'un des éléments de ce potentiel ;
- c. le couple (E₂, (L₁→E₁)) est une prise de position, à deux, par rapport à cette proposition d'actualisation. Cette prise de position est négociable ;
- d. s'il y a un sens, il est en suspens, par exemple au temps t₃ de l'énonciation de L₂.

a. Il est clair que le locuteur E exprime sous forme langagière "quelque chose". L'énoncé E₁ est une forme linguistique, porteuse de sens. Nous ne nous permettrons pas de dire ici qu'il s'agit d'une affirmation, de l'expression d'un regret, d'une inquiétude, d'une question ou autre action langagière. Nous dirons simplement que cet énoncé est **susceptible de porter tout ceci**, et peut-être autre chose que nous, analystes, "ne voyons pas". La question n'est sûrement pas de privilégier l'un ou l'autre de ces possibles.

b. En énonçant L_1 , L propose d'actualiser un des sens du potentiel. Il s'agit d'un sens que l'on pourrait dire "au pied de la lettre". En effet en disant "Oh c'est marrant", "exactement" et "ce", E exprime son étonnement que son tableau soit là devant lui, dans cette vitrine. Pourquoi peut-on dire cela ? Parce que L apporte une explication à cet étonnement : "c'est le tien, on te l'a volé". Il satisfait les conditions préparatoires de l'acte de langage expressif d'étonnement.

c. En énonçant E_2 , E invalide cette proposition de sens. En effet, il donne un élément supplémentaire qui prouve que "ce" tableau n'est pas le sien : c'est la tonalité. Par conséquent, il dit à son interlocuteur : il ne faut pas prendre ce que j'ai dit au pied de la lettre. Le couple (E_2 , ($L_1 \rightarrow E_1$)) a pour fonction de suspendre, dans ce cas, le mécanisme d'ajustement des cognitions.

d. A cet instant, l'observateur ne peut en aucun cas dire quelque chose comme : E a voulu dire ceci, L ne l'a pas compris. Le processus est en cours, son produit fini (s'il existe) est encore indéterminé, il est négociable.

Remarquez que dire tout de go que L_1 est ironique, tombe sous la même critique : de quel droit l'observateur peut-il pénétrer dans le système cognitif de L pour y "voir" qu'il ne croit pas à ce qu'il dit en L_1 . La réponse sur la question du caractère ironique de ce "on te l'a peut-être volé" est à chercher dans la suite de l'interaction. Peut-être d'ailleurs ne pourra-t-on pas la résoudre.

L'extrait étant limité à trois tours de parole on ne peut "savoir" si ce phénomène d'intercompréhension se résoudra en un point de fixation sur le (?) sens en jeu ici. En tout état de cause, nous aimerions insister sur le point suivant. L'interlocution est un endroit où se déploie un flux de sens qui peut porter la mésentente, le malentendu et qui est marqué par une indétermination radicale (voir à ce sujet les pages très éclairantes de Pierre Livet (1994, chapitre 1 particulièrement)). Cette indétermination ne constitue en aucun cas un obstacle à l'intercompréhension. En effet, c'est précisément parce que le processus est producteur d'indétermination qu'il permet l'intercommunicabilité des cognitions. Le déroulement de la conversation se satisfait pleinement de la possibilité d'une simple entente provisoire. Il s'actualise même précisément sur ce provisoire. De ce point de vue, il est intéressant de se pencher sur les cas où les interlocuteurs se rendent mutuellement compte que jusqu'alors ils étaient en train de "surfer" sur un malentendu. Il se produit alors un hiatus dans le déroulement linéaire de la communication. Mais loin d'être le signe d'un échec, un tel événement est en fait la marque d'une communication réussie. En effet, se rendre compte que l'on croyait se comprendre, mais qu'on se trompait, signifie "en relief" que la communication peut être, sinon parfaite, néanmoins réelle.

Autrement dit, si E voulait, intentionnait de communiquer une cognition, son expression linguistique, E_e , est susceptible de le permettre ; la cognition en question est communicable, même si non communiquée. Mais E peut très bien se satisfaire de ne pas “avoir été entendu”, pour autant que son interlocuteur coopère à une construction conjointe. L’important c’est la communicabilité et non la communicativité.

1.3. Conclusion

Selon nous il y a un danger majeur à parler de l’activité de compréhension réalisée par les sujets en interlocution. En effet, on peut soutenir que lorsqu’un énoncé est produit par un des locuteurs, il n’est pas compris par l’auditeur. Ce dernier ne met pas en place une procédure de traitement qui le mènerait à découvrir ce que voulait dire son interlocuteur, qui le conduirait à exhiber un sens qui préexisterait à ce traitement.

Le travail de participation conjointe à l’émergence de sens ne porte pas sur un objet qu’il faut observer de l’extérieur. Ce travail est un façonnage, une sculpture, qui se réalise à deux, d’un objet en perpétuel devenir : le flux de sens des expressions langagières produites par les conversants. Le nom de ce travail ? L’intercompréhension.

2. Modéliser l’intercompréhension ; avec quels outils ?

2.1. Observer, modéliser, valider. Telle est la démarche systématique prônée dans les sciences qui se donnent pour objet les phénomènes naturels qui surviennent dans le monde. La psychologie n’échappe pas à ce principe méthodologique, en particulier la psychologie des activités cognitives produites par les humains. Ainsi Bisseret dit-il, en parlant des démarches de recherche en psychologie cognitive,

qu’elles « sont celles des sciences naturelles et expérimentales : après l’observation, la construction et la validation d’hypothèses et de modèles » (1988 : 128).

Mais quels compléments d’objet direct s’agit-il de placer derrière ces trois verbes d’action ? S’agissant d’étudier la cognition humaine, on trouve : la mémoire, la résolution de problème, la prise de décision, la production langagière, la reconnaissance de formes et autres. C’est dire que la chose à observer et modéliser n’est autre qu’une certaine forme de traitement de l’information en provenance de l’environnement dans lequel est plongé le sujet producteur des activités étudiées. En tout cas le postulat est souverain. On le trouve par exemple énoncé dès l’entame du préambule du *Traité de Psychologie Cognitive* :

« Le débat entre le behaviorisme et le mentalisme a ouvert une nouvelle problématique, celle du traitement de l'information, dont se réclame le cognitivisme » (Bonnet *et al.*, 1989 : V).

En tout état de cause, c'est le sujet humain dans son individualité qui est, le plus souvent, l'objet d'études. On s'intéresse à la mémoire qu'un individu possède, procède et/ou actualise. On rend compte de l'activité cognitive d'un sujet placé en position de résoudre un problème, de prendre une décision ou de reconnaître des formes. On analyse la production langagière d'un seul sujet parlant. Très généralement, la perspective est égologique, monologique ; et ce, que l'on se place au niveau de l'observation (même si l'observé prend part à une activité groupale), de la modélisation (la formalisation concerne les propriétés, les caractéristiques individuelles) et de la validation (par exemple, l'intelligence artificielle classique, pour ce qui concerne la validation par simulation informatique, est radicalement monologique – avant l'avènement de son prolongement distribué, nous y reviendrons –).

Il est pourtant des activités cognitives humaines qui prennent place dans l'interaction et qui sont le produit de l'interaction. L'intercompréhension, processus cognitif tout à fait central dans les activités humaines, est de celles-ci. Il n'est pas de lieu où ne s'exercent les talents du psychologue, de toutes sous-disciplines, qui puisse faire l'économie de ce processus intersubjectif. Son étude est cependant souvent oubliée... peut-être parce que ce processus s'ancre sur une relation et non sur des individus, peut-être parce qu'il s'agit plus d'un processus inter-cognitif que simplement cognitif. Ainsi donc, nous proposons pour complément d'objet direct des verbes précités :

- **observer** la conversation comme lieu de la *construction conjointe de cognitions distribuées* ;
- **modéliser** les mécanismes de perlaboration de ce produit "inter-cognitif" qu'est l'*intercompréhension* ;
- **valider** le *modèle des mécanismes de l'interaction langagière* qui génèrent cette intercompréhension.

2.2. *Quelle activité décrire et modéliser ?*

Nous allons nous attarder sur les rapports qu'entretiennent enchaînement conversationnel, intercompréhension et modélisation. Nous le ferons en nous appuyant sur les nombreuses contributions que des psychologues de la cognition ont apportées à l'important ouvrage en la matière intitulé *Psychologie cognitive, modèles et méthodes* (Caverni *et al.*, 1988).

Comme nous l'avons rapidement évoqué plus haut, l'objet d'études classiquement assigné à la psychologie cognitive est le sujet ; le sujet *via* les activités cognitives que l'on envisage comme autant de modes de traitement de l'information provenant d'un monde extérieur à ce sujet.

L'être vivant ("le sujet"), placé dans un environnement et soumis à des stimulations, produit en réponse une activité observable ("le comportement") (Caverni et Bastien, 1988 : 7).

En nous attachant aux mécanismes engendrant la conversation, nous sommes amenés à tenter de comprendre le fonctionnement d'une dyade de sujets en situation d'interaction langagière. L'"être vivant" est pour nous, le **couple de sujets parlant**, l'"activité observable" est la **séquence des dire**s proférés par les membres du couple.

Le temps de description ressort d'une attitude naturaliste. Il s'agit en effet de capter des conversations naturelles et authentiques. Quelles que soient les situations, les sujets parlant agissent spontanément, sans contrôle expérimental, dans un objectif propre à la situation donnée. Certains achètent un billet de train à un employé, d'autres réfléchissent à une alternative pour enfin prendre une décision, d'autres jouent ou encore résolvent un problème, etc. Capter pour décrire le plus finement possible "ce qui se passe". Capter pour rendre compte du comportement de la dyade porteuse du mécanisme d'intercompréhension ; rendre compte au sens que détaillent Caverni et Bastien :

« mettre en évidence les relations stables que les comportements manifestent avec des propriétés de l'environnement dans lequel ils sont produits [et] élucider non plus seulement les facteurs externes qui déterminent les comportements, mais principalement les phénomènes susceptibles de rendre compte de leur élaboration par le sujet » (1988 : 7).

Autrement dit, nous nous assignons comme objectif d'élucider les phénomènes qui engendrent l'élaboration, par le couple d'interactants, de la séquence des dire

2.3. User d'un modèle-cadre

Bien entendu, il y a là convocation de catégories d'analyse permettant un rendu adéquat des processus décrits ; il y a là convocation d'un modèle que Rouanet (1983) a qualifié de cadre. Reprenant à leur compte cette dénomination, Lecoutre, Rouanet et Denhière le définissent ainsi :

« le modèle-cadre correspond à un dispositif, une maquette, que l'on construit aux fins de rendre opérationnelles les questions que l'on se pose sur les phénomènes étudiés » (1988 : 384).

Autrement dit, si la description pénétrante des enchaînements conversationnels soulèvent des questions, c'est grâce à ce modèle qu'on pourra les opérationnaliser.

C'est à cet endroit qu'il nous a paru pertinent d'en appeler à la théorie des actes de langage. Cette utilisation fait l'objet de nombreuses controverses (par exemple les ethnométhodologues, conversationnalistes affirmés, rejettent cet usage comme moyen d'analyse (Levinson, 1983, chap. 6) ; voir aussi Trognon et Brassac, 1988 ; Brassac, 1992 ; 1994a ; 1994b ; Moeschler, 1990). En tout état de cause, l'axiomatisation de cette théorie, qui prend la forme de la logique illocutoire ainsi que son extension, la sémantique formelle générale (Vanderveken, 1988 ; 1990) jouent ce rôle de modèle-cadre (nous reviendrons sur ce soubassement théorique quand nous exposerons la teneur de la logique interlocutoire). Il nous a permis d'interroger très précisément l'interaction verbale en entrant par une porte essentielle à nos yeux : l'aspect pragmatique. Ce faisant, cet attentif examen des actes accomplis par les interactants nous a fait en retour interroger la catégorie de base et ses propriétés. L'aller et retour entre les données empiriques et l'appareil conceptuel a produit une remise en cause fondamentale des caractéristiques de la logique illocutoire des conditions de succès et de satisfaction (Brassac, 1992 ; Trognon et Brassac, 1992 ; Ghiglione et Trognon, 1993) qui constitue le modèle cadre.

3. Modéliser l'enchaînement conversationnel

3.1. La logique interlocutoire...

La logique interlocutoire est une **dialogisation** de la sémantique formelle générale proposée par Vanderveken (1988 ; 1990). Disons quelques mots de cette dernière avant de montrer, en utilisant un exemple, pourquoi un remaniement est tout à fait nécessaire pour qu'elle puisse faire l'objet d'une utilisation adéquate afin de rendre compte de l'enchaînement conversationnel.

a. Adopter une perspective pragmatico-linguistique relativement aux faits conversationnels, c'est d'une part s'intéresser plus particulièrement à l'aspect verbal des interactions et d'autre part postuler ceci : produire un énoncé dans une conversation, c'est réaliser une action. Cette action, dite acte de langage s'écrit toujours sous la forme F(P), avec F force illocutoire qui s'applique, au sens mathématique du terme, sur P, contenu propositionnel

La logique illocutoire, axiomatisation de la théorie des actes de langage propose un ensemble de conditions à l'accomplissement réussi d'un acte de langage (Searle et Vanderveken, 1985). Ces conditions sont dites conditions de succès. Ainsi lorsque qu'un locuteur réalise une promesse, il le fait avec succès s'il remplit un certain nombre de conditions (être dans la capacité de faire en sorte que le contenu propositionnel devienne vrai, être sincère, donner un contenu à la promesse qui soit au futur par rapport à l'instant de l'énonciation, etc.). Cela dit, le fait qu'il tienne cette promesse n'a rien à voir avec ces conditions. Il est très possible de réussir parfaitement la promesse portée par "*Je promets d'arriver tout à l'heure à huit heures précises*" sans tenir cette promesse. Ce sont les conditions de satisfaction qui régissent le fait que cette promesse soit tenue ou non (Vanderveken, 1988 ; 1990).

On dira que cette promesse est tenue (satisfaite au sens technique) si le locuteur rend vrai le contenu propositionnel (qu'il arrive tout à l'heure à huit précises) et ce, parce qu'il avait énoncé cette promesse (elle ne serait pas pleinement tenue si il était arrivé à huit précises par hasard). Les conditions de satisfaction permettent l'opérationnalisation de l'aspect perlocutoire de l'acte de langage. En unifiant les conditions de satisfaction à celles de succès, Vanderveken dépasse une limite de la logique illocutoire. Cette dernière ne traite pas de la perlocution, n'envisage pas les effets des actions langagières sur le monde (hormis la compréhension de l'auditeur). Cela dit la sémantique formelle générale (qui résulte de cette intégration des deux valeurs sémantiques de tout acte de langage, le succès et la satisfaction) reste largement insuffisante pour rendre compte de l'enchaînement conversationnel. En effet (i) elle reste une sémantique de la littéralité, (ii) elle garde un caractère statique et (iii) elle est aussi monologique que la logique illocutoire. Certes, elle permet le pas vers une dialogisation du fait de la liaison qu'un de ses théorèmes fournit entre la satisfaction et le succès, du fait de la mise en scène de l'effet sur le monde et par là, de la réaction de l'auditeur... mais elle ne le franchit pas.

La logique interlocutoire que nous proposons s'étaye sur cette sémantique formelle générale et la dépasse en intégrant les aspects non littéraux, en adoptant une perspective radicalement dynamique relative à la mise en séquence des actes de langage, bref comme nous l'annoncions plus haut, en la dialogisant. Nous avons décrit cette logique en plusieurs endroits (Brassac, 1992 ; Trognon, 1991 ; Trognon et Brassac, 1992 ; Ghiglione et Trognon, 1993). Voyons une application de son principe de base sur l'exemple suivant.

b. Une des quatre séquences suivantes, mettant en scène un homme (E) raccompagnant en voiture chez elle une jolie jeune femme (L) et arrivant au pied de l'immeuble, a été observée et retranscrite.

Séquence 1

E₁ Tu as le téléphone ici
L₁ Oui, c'est moderne
E₂ Ah... je n'aurais pas cru

Séquence 2

E₁ Tu as le téléphone ici
L₁ Oui, c'est le numéro 04-83-35-36-09
E₂ Ah ben je pourrais t'appeler comme ça

Séquence 3

E₁ Tu as le téléphone ici
L₁ Oui, c'est moderne
E₂ (rires) Ah ben j'aurais bien aimé que tu me donnes le numéro

Séquence 4

E₁ Tu as le téléphone ici
L₁ Oui, c'est le numéro 04-83-35-36-09
E₂ Euh, mais je ne te demandais pas le numéro

D'abord une question brutale : dans l'ensemble de ces quatre séquences, y en a-t-il une, ou plus, où les deux interactants se sont compris ? Si l'on est tenté de répondre "oui, il se sont compris à la séquence 2" c'est que l'on pense que E avait pour intention communicative de demander le numéro de téléphone (et non pas seulement de demander si L avait le téléphone). Et par conséquent, que la séquence 4 n'est pas possible. Mais de quel droit peut-on affirmer ceci ? Quel analyste a le pouvoir de décider que E demandait, implicitement, le numéro ? Remarquons qu'il n'est pas plus justifié de dire que E voulait, littéralement, savoir si L avait le téléphone ou non. Mais là n'est pas l'important. L'important c'est le sens que ces deux interactants donnent, dans l'ici et maintenant de la conversation, à E₁.

Plus techniquement, on peut dire que E₁ est une assertion formalisable par ASS(p) et paraphrasable performativement par "Je dis que tu as le téléphone ici". Au titre des actes non littéraux associés à cette assertion, on peut évoquer sans doute une *requête d'information* formalisable par DIR(p) et paraphrasable performativement par "Je te demande si tu as le téléphone" ou aussi une *requête d'action* formalisable par DIR'(p') et paraphrasable performativement par "Je te demande ton numéro de téléphone". Il est possible que cet énoncé porte d'autres actes illocutoires (pourquoi pas une expression d'étonnement, d'admiration ou autres... ?). Le fait décisif, c'est que L, en répliquant, satisfait, au sens technique du terme vu plus haut, soit

DIR(p), pour les séquences 1 et 3, soit DIR'(p'), pour les séquences 2 et 4. Ce faisant, selon une loi fondamentale de la sémantique générale, il donne à DIR(p) [ou à DIR'(p')] un statut d'acte de langage réussi. C'est-à-dire que c'est L qui, rétroactivement, propose à E d'actualiser son dire en une requête d'information [et respectivement en une requête d'action]. A cet instant de l'échange, le statut du dire de E est en balance. En effet, son locuteur peut soit ratifier la proposition (c'est le cas des séquences 1 et 2), soit l'invalider (c'est le cas des séquences 3 et 4). On a alors, au troisième tour de parole, une stabilisation du sens qu'a acquis le premier énoncé. Stabilisation et non pas fixation car ce statut est toujours négociable dans l'après conversationnel.

D'après nous, il n'est en aucun cas plus justifié de dire, en tant qu'analyste, que E et L se sont compris dans l'une ou l'autre de ces séquences. De plus c'est un leurre que de penser que E lui-même, interrogé par après, pourra donner la clé et dire si c'est ce qu'il "avait voulu dire" car on aura affaire là à une reconstruction *a posteriori*. Prendre au sérieux une option empreinte de communicabilité (vs communicativité) nous mène à conclure ceci : les deux interactants se sont "intercompris" dans les quatre séquences. Peu importe de savoir si L a bien "découvert" ce que "voulait dire" E. Ce qui est central c'est qu'ils ont, au cours du développement de l'interlocution, fait émerger un sens de E_i dont ils sont co-responsables. Parmi les quatre séquences, c'est la première qui a réellement été observée ; le sens de E_i, alors stabilisé, a été recyclé plusieurs semaines plus tard par les deux mêmes interactants.

C'est la logique interlocutoire, en tant qu'elle rend compte de façon formelle des enchaînements d'actions langagières accomplies au cours de l'interlocution, qui permet de justifier cette vision dynamique de la construction conjointe du sens, de justifier cette appréhension constructiviste de l'intercompréhension interhumaine.

3.2. ... comme modèle hypothétique

Progressivement s'est ainsi constitué un modèle de l'enchaînement conversationnel et par là de l'intercompréhension interhumaine.

Ce que nous proposons semble en effet répondre aux différentes caractérisations que donnent les contributeurs évoqués plus haut de ce que doit être un modèle. Pour Gonzalez (1988), par exemple, il s'agit d'un dispositif qui doit être :

un « instrument de conceptualisation » (p. 462), une « description non ambiguë d'un phénomène » (p. 446), une « interprétation acceptable, à un moment donné, d'une classe de phénomènes »

(p. 447) ou encore une « construction hypothétique proposant une certaine intelligibilité d'une activité cognitive » (p. 458).

Pour Tiberghien (1988), il doit s'agir :
d'un « discours logico-mathématique sur un ensemble limité de phénomènes dans des conditions soigneusement définies » (p. 14)
c'est-à-dire d'une « micro-théorie » (p. 15).

Expliquons pourquoi notre dispositif théorique peut prétendre à endosser ces réquisits méthodologiques.

Nous soutenons que la logique interlocutoire est un discours logique permettant de rendre compte de la génération processuelle de l'intercompréhension. Plus, elle constitue un modèle formel de ce mécanisme interlocutoire au sens où Tiberghien le définit :

« système relationnel, de forme, de structure et de fonction entre deux objets dont l'un est un système de forme et l'autre, l'ensemble limité des objets que l'on souhaite décrire et/ou interpréter » (1988 : 17).

En effet, la logique interlocutoire possède comme le précise le même auteur :

« une syntaxe (qui est celle du système formel utilisé) et une sémantique (résultant de la projection du système formel dans une réalité signifiante) » (1988 : 15).

Sa syntaxe se fonde évidemment sur les postulats, définitions, théorèmes, règles de la logique des conditions de succès et de satisfaction des actes illocutoires (dont l'exposé constitue l'intégralité du deuxième tome de Vanderveken, 1990) et y adjoint les formules qui rendent compte de la compositionnalité des propriétés des actes de langage produits par les deux interactants. Elle inclut la formalisation du langage et de la démonstration. Sa sémantique constitue une interprétation de la séquenciation des actes enchaînés verbalement par les conversants.

La logique interlocutoire est donc l'expression formalisée d'un **modèle hypothétique**, que Lecoutre, Rouanet et Denhière définissent comme un :

« mécanisme, plus ou moins précis, qui exprime le fait qu'un processus cognitif se passe d'une certaine façon » (1988 : 384) ;

modèle pourvoyant une description et une interprétation d'un secteur de la réalité intersubjective.

Ainsi donc nous affirmons à cet endroit que :

1. La sémantique formelle générale (logique des conditions de succès et de satisfaction des actes illocutoires) sert de modèle-cadre à la description explicative de l'enchaînement conversationnel.

2. La logique interlocutoire est l'expression formalisée d'un modèle hypothétique de l'intercompréhension.

Mais aussi que :

3. Tout ceci laisse entière la question de la validité d'une telle modélisation des mécanismes de l'interaction langagière... que nous allons traiter dans la partie suivante.

4. L'intercompréhension : l'objet de simulation ?

4.1. Simuler vs expérimenter

Nous sommes là à l'exact endroit de la mise à l'épreuve des faits, de ce modèle.

« La psychologie doit [...] valider ses modèles : elle doit montrer qu'au niveau de description choisi, le modèle est une bonne approximation du fonctionnement humain qu'il est censé représenter » nous dit Bisseret (1988 : 136).

Il est en effet clair que l'on ne peut se suffire de la « belle cohérence interne » de cette logique interlocutoire (pour autant que l'on puisse la qualifier ainsi). Nous suivons là sans hésitation Tiberghien qui affirme que :

« de toute façon, et de toute façon irrévocable, la valeur d'un modèle ne peut s'apprécier uniquement à partir de critères de cohérence et de validité internes » (1988 : 22).

L'expérimentation nous semble inadéquate pour deux raisons. La première est qu'il nous semble très délicat, au sujet du processus d'intercompréhension, de répondre aux questions suivantes : quelles variables indépendantes instancier ? Quelle variable dépendante mesurer ? Quelles conditions et situations expérimentales contrôler ? La seconde est plus fondamentale. Nous nous intéressons, comme mentionné ci-devant, tout à fait prioritairement à des interlocutions naturelles et authentiques. Ce faisant, nous faisons œuvre de psychologue au sens où nous faisons nôtre la formule de Bisseret :

« Le sujet dont la psychologie cherche à décrire et à expliquer le fonctionnement est l'homme *dans la vie courante* » (1988 : 130, c'est nous qui soulignons).

En appeler à une parabole, bien connue autant que plaisante, comme le fait cet auteur n'est pas décisif pour écarter ce type de démarche :

« Il n'est certes pas possible de "comprendre le fonctionnement humain" par la seule observation, mais chercher à le faire par le seul moyen des situations de laboratoire, parce que là est le "contrôle rigoureux", me fait penser à la parabole du quidam qui, la nuit, avait perdu sa clé là-bas, mais qui la cherchait ici, sous le réverbère, parce qu'il y faisait clair ! » (1988 : 132)...

Reste la simulation. Nous adopterons, par un choix délibéré qui ne méconnaît pas les difficultés sous-jacentes, le recours à une simulation informatique.

4.2. Estimer l'intercompréhension interhumaine...

Ce que nous avons présenté jusqu'à présent, le modèle formel hypothétique, a pour objectif de rendre compte de lois, de principes, de règles, de la génération de l'intercompréhension. Il s'agit là d'une tentative de reconstruction, à partir de phénomènes directement observables (les formes langagières successivement produites), de processus de l'activité des systèmes nerveux centraux concernés qui eux ne sont pas observables. C'est en ce sens que Tiberghien parle de modèles d'estimation :

« de façon très générale, il est possible d'identifier deux classes différentes de modèles : des *modèles d'estimation* et des *modèles de simulation*. Les premiers tentent uniquement d'expliquer les relations empiriques, fonctionnelles ou corrélationnelles — ce sont les modèles au sens habituel du terme —, les seconds tentent, en outre, de reproduire les relations mises à jour par les modèles d'estimations » (1993 : 43, c'est l'auteur qui souligne).

Autrement, dit, valider signifie aller au-delà d'une reconstruction formelle de l'inobservable pour atteindre une reconstruction qui, elle, est plus de la *reproduction*. Cette reproduction est réalisée à l'aide de systèmes calculatoires artefactuels, en l'occurrence ici à la machinerie informatique.

4.3. ... pour mieux l'émuler

4.3.1. Interaction sociale et intelligence artificielle distribuée

Ainsi donc, nous sommes amenés à emprunter les chemins de l'intelligence artificielle. Il est clair que notre projet interactionniste nous conduit plus précisément encore vers une sous-partie de ce domaine, nous voulons parler de l'IAD (pour Intelligence Artificielle Distribuée) (Ferber, 1995). L'intelligence artificielle distribuée est par nature confrontée à la problématique de l'interaction. Un système multi-agent, parangon de l'objet d'étude de cette branche de l'intelligence artificielle, est constitué d'une

population d'entités artefactuelles rassemblées autour d'une tâche commune. Ces entités sont des agents qui ont un certain nombre de caractéristiques leur permettant de communiquer, de percevoir l'environnement, d'avoir des comportements autonomes et de poursuivre des objectifs individuels. Entités réelles ou abstraites, ces agents sont parties prenantes de phénomènes interactionnels qui organisent et structurent la population qu'ils constituent.

Omniprésence de l'interaction, travail sur les modes de communication intentionnelle, mise en avant de l'action des agents sur leur environnement..., on comprend bien pourquoi les chercheurs informaticiens en ce domaine auront tenté de trouver dans la théorie des actes de langage un outil leur permettant d'instrumentaliser les interactions inter-agents. Nous avons donné à plusieurs reprises notre position à l'égard de l'importation de cette théorie en IAD (Brassac, 1993 ; Brassac *et al.*, 1996 ; Brassac et Chevrier, 1996). En 1993, par exemple, nous écrivions :

« Il s'avère que cette théorie n'a pu répondre aux espoirs qu'on avait place en elle. (...) Convoquée pour rendre compte, pour tenter d'implémenter les problèmes d'interaction communicative entre agents cognitifs, [Elle] est maintenant tout à fait dépassée. [Elle] a beaucoup évolué ces dernières années ; plusieurs avancées lui ont fait franchir des pas importants dans la compréhension de l'enchaînement des actions que constitue toute communication intersubjective. Nous prétendons que les développements récents (de 1988 à 1993) de la théorie des actes de langage doivent inciter à réexaminer la pertinence de la version récente de cette dernière pour la problématique qui nous intéresse ici » (Brassac, 1993).

Et bien sûr de poursuivre en mettant en avant la perspective dialogique que nous avons ci-devant développée.

Il est clair, par le fait même du projet de ce versant distributionnaliste de l'IA, que la convocation de cette théorie des actes de langage s'impose quand il s'agit d'essayer de simuler l'interaction interhumaine. Mais la position que nous occupons ici est "à rebours".

Nous parlions alors d'une extériorité de notre position à l'égard de ce champ :

« Psychologue social dont les recherches ont pour objet l'interaction sociale, et plus particulièrement cette forme prototypique de l'échange langagier de face à face qu'est la conversation, notre point de vue sera extérieur au champ propre de l'intelligence artificielle distribuée. Une extériorité toute relative eu égard à un domaine qui ne cesse d'en appeler aux recherches menées en sciences humaines et sociales concernant les

mécanismes et les protocoles de communication. Extériorité relative mais aussi déclarée pour revendiquer la pertinence de recherches menées sur des conversations dites naturelles, authentiques ou ordinaires dans le dessein de réfléchir à une possible extension de leurs résultats à des dialogues finalisés et à implémenter » (Brassac, 1993).

De fait, nous voulions user des résultats obtenus en sciences humaines de l'interaction pour le profit des sciences de l'artificiel. Après l'aller, il s'agit maintenant d'effectuer le retour en posant la question suivante : en quoi les dispositifs informatiques pourront-ils permettre la simulation envisagée ?

En résumé, et pour conclure cette partie, nous voulons avancer les idées suivantes :

- l'emprunt du versant distributionnaliste de l'intelligence artificielle est motivé par le fait que notre objet n'est pas le sujet individué, monadique, mais le couple de sujets, la dyade ;
- la simulation d'une activité "naturelle" qui procède de l'intersubjectivité nous conduit à user d'artefacts relevant du monde des multi-agents informatiques.

4.3.2. Une tentative de simulation de prise de rendez-vous

Ainsi donc nous avons pour ambition de reproduire artificiellement, en utilisant le savoir-faire multi-agent, le fonctionnement d'une dyade humaine en train d'interagir verbalement, d'un couple de sujets en train de s'intercomprendre. Les difficultés qui nous accueillent à l'entame de ce projet sont au minimum nombreuses, qu'elles soient d'ordre technologique, méthodologique et/ou épistémologique. Certes ! Mais gageons que ce projet nous amènera à les expliciter et ainsi, soit à les mieux circonscrire, soit à les déclarer insurmontables.

Comme on l'a dit plus haut, une des spécificités de l'approche multi-agents vient du fait qu'y est considéré pour point focal non pas l'agent comme entité isolée, mais l'unité globale constituée par l'interdépendance des agents dans leur rapport à un milieu commun. Il s'agit en d'autres termes d'étudier le comportement global du collectif d'agents ; ce comportement est émergent et provient de l'ensemble des phénomènes locaux. Une telle étude s'envisage donc par une approche ascendante selon laquelle seules les règles de comportement de chaque agent sont spécifiées ; le phénomène d'intérêt devant émerger de l'interaction. Si les systèmes multi-agents sont de plus en plus utilisés dans des aspects "résolution distribuée de problèmes", une nouvelle orientation se dessine actuellement :

celle de l'utilisation d'une approche multi-agents dans les domaines de la simulation, ceux-ci étant classiquement abordés par des approches mathématiques (équations différentielles, modélisation stochastique...). En effet les systèmes multi-agents apportent une vision radicalement nouvelle au problème de la simulation et semblent particulièrement adaptés pour simuler des comportements de groupe (par exemple une "société" cellulaire, animale ou humaine) dont les structures, comportements et autres activités émergent des interactions entre les individus. On parle dès lors de simulation comportementale (Pesty, 1996). Voyons comment nous envisageons d'aborder techniquement la question.

Nous avons déjà dit en quoi, selon nous, les langages d'interaction se fondant sur la théorie des actes de langage en IAD, KQML (Finin *et al.*, 1994), IL (Demazeau, 1995), nous semblent inadéquats pour notre projet (Brassac *et al.*, 1996). L'argument principal est qu'ils ne permettent pas d'atteindre un aspect tout à fait essentiel de l'interaction conversationnelle que nous voulons simuler : la négociation des statuts interlocutoires des énoncés (dits improprement 'messages' dans ces approches (Brassac et Pesty, 1997)). Ce que nous visons ici, c'est la spécification d'un protocole de communication inter-agents qui laisse place à l'indétermination inhérente aux actes langagiers que les interactants réalisent en conversation. Plus précisément, il s'agit de permettre à l'agent-auditeur d'interpréter un 'message' de façon non contrainte. Autrement dit, lorsque cet agent reçoit la production langagière « *jeudi, huit heures* » [1], il devrait pouvoir actualiser les sens suivants : « *je suis libre jeudi à huit heures* », « *je te demande si tu es libre jeudi à huit heures* », « *je serais là jeudi à huit heures* », etc. Evidemment tout cela dépend du lieu conversationnel où est proféré [1]. Il s'agit aussi de permettre à l'agent-locuteur de ratifier ou non l'une de ses actualisations. Cela requiert une certaine autonomie pour les agents qui devra s'appuyer sur une plate-forme qui le permet. Cette plate-forme de simulation sera en quelque sorte apte à engendrer une dynamique conversationnelle.

Une ébauche de cette plate-forme a d'ores et déjà été élaborée et implantée. On la trouve dans un travail de DEA de Sciences Cognitives de l'université de Grenoble (Alidra, 1997), travail encadré par Sylvie Pesty et auquel nous avons participé. Y est présenté un simulateur de prise de rendez-vous, réalisé en langage JAVA sur système Unix. La gestion artificielle de dialogues, dont la finalité est de trouver une plage horaire que les deux agents peuvent libérer pour une rencontre, s'étaye sur la perspective dialogique que nous défendons. De plus, les 'profils' des agents, définis par des paramètres configurables, sont intégrés dans les déterminants de la plate-forme. Plusieurs dialogues simulés sont présentés dans leur déroulement effectif et commentés. Bien évidemment, il ne s'agit que de l'ouverture d'une brèche dans un mur de difficultés à venir. Cela dit, l'expérimentation

proposée, reposant sur la manipulation de paramètres tels l'hésitation, l'insistance, les relations de préséance, semble promettre des résultats intéressants.

Conclusion

Les recherches menées au GRC depuis une bonne dizaine d'années concernent toutes la question qui peut se formuler ainsi : interagissant en utilisant le langage dans une situation donnée, comment deux personnes humaines se comprennent-elles ? Nous avons montré que pour donner des éléments de réponse à cette question, point n'était besoin de postuler l'existence d'un sens préexistant et à l'expression et à la compréhension de la forme linguistique proférée en contexte. Il suffit d'accepter l'idée simple selon laquelle les interactants, immergés dans un potentiel subtilement impermanent de sens, le façonnent conjointement et processuellement et ainsi en font émerger un sens provisoire et toujours négociable. L'intercompréhension est cette co-opération intersubjective.

Le modèle formel qui nous amène à adopter cette perspective est la logique interlocutoire. Cette dernière, qui rend compte de la dynamique du processus conversationnel et de son caractère fortement indéterminé, a le statut de modèle d'estimation (Tiberghien, 1993). A l'issue de ce travail de description-modélisation de l'intercompréhension, la question posée ci-dessus reçoit une réponse que l'on peut (très succinctement) formuler comme suit : Les interactants humains se comprennent en ce sens où ils sont co-responsables du sens des actions langagières accomplies par la relation qu'ils contribuent à former.

Notre travail atteint l'étape de la validation. En projetant d'émuler le modèle (Tiberghien, 1993), nous envisageons une simulation informatique relevant du domaine des multi-agents ; évidemment difficile, cette dernière est actuellement en cours de réalisation (Alidra, 1997).

Cela dit, indépendamment des résultats qu'elle pourra apporter, nous soutenons que :

- l'intercompréhension doit occuper une place centrale dans l'ensemble des objets d'études de la psychologie cognitive ;
- l'intercompréhension doit être (et peut être) étudiée de façon classique (observer, modéliser, valider) ;
- la démarche qui doit être mise en œuvre pour mener à bien cette étude scientifique de l'intercompréhension, doit s'inscrire dans une perspective radicalement interactionniste. Etudier de la sorte ce processus essentiel ouvre la voie à la constitution d'un champ de recherches en psychologie cognitive interactionniste.

Références bibliographiques

- Alidra, L. (1997). Les langages d'interaction dans les systèmes multi-agents. *Rapport de DEA Sciences Cognitives*. Institut National Polytechnique de Grenoble. Soutenu le 24 juin 1997.
- Bisseret, A. – Modèles pour comprendre et réussir. In J.P. Caverni, C. Bastien, P. Mendelsohn et G. Tiberghien (Eds.), *Psychologie cognitive, modèles et méthodes*, PUG, Grenoble, pp. 127- 140, 1988.
- Bonnet, C., Ghiglione, R., Richard, J.F. (1989). *Traité de psychologie cognitive 1*. Dunod, Paris.
- Bovet, P. – Introduction. In J.P. Caverni, C. Bastien, P. Mendelsohn et G. Tiberghien (Eds.), *Psychologie cognitive, modèles et méthodes*, PUG, Grenoble, pp. 377-382, 1988.
- Brassac, Ch. (1992). Analyse de conversations et théorie des actes de langage. *Cahiers de Linguistique Française*, **13**, 62-76.
- Brassac, Ch. (1993). Théorie des actes de langage et Intelligence Artificielle Distribuée. *Actes des Journées Systèmes Multi-Agents du PRC-IA*, Montpellier, 17 décembre 1993.
- Brassac, Ch. (1994). Speech acts and conversational sequencing. *Pragmatics and Cognition*, **Vol 2(1)**, 191-205. [1994a]
- Brassac, Ch. (1994). Actes de langage et enchaînement conversationnel. *Cahiers d'épistémologie n°9401*. Université du Québec à Montréal. [1994b]
- Brassac, Ch. (1994). L'interaction inter-agents : non littéralité et processualité, *Actes des Deuxièmes Journées Francophones Intelligence Artificielle Distribuée - Systèmes Multi-Agents*, Voiron, 9-10-11 mai 1994. [1994c]
- Brassac, Ch., Pesty, S. (1997). Ancrer les agents dans l'environnement. *Cinquièmes Journées Francophones d'Intelligence Artificielle Distribuée et Systèmes Multi-Agents, JFIADSMA'97*, La Colle-sur-Loup, 2-4 avril 1997.
- Brassac, Ch., Stewart, J. (1996). Le sens dans les processus interlocutoires : un observé ou un co-construit ? *Actes des Cinquièmes Journées de Rochebrune "Du social au collectif"*, 29 janvier-3 février 1996, 85-94.
- Brassac, Ch., Chevrier, V. (1996). Vers un réexamen du statut de l'interaction dans les systèmes multi-agents. *Interaction et cognitions*, **Vol. 1(1)**, 3-22.
- Brassac, Ch., de Almeida, J., Grégori, N., Saint-Dizier, V. – La théorie des actes de langage en intelligence artificielle distribuée : utilisations et

- limites. In Müller, Jean-Pierre et Joël Quinqueton (Eds.), *IA distribuée et systèmes multi-agents, JFIADSMA '96*, Hermès, Paris pp. 229-249, 1996.
- Caverni, J.P., Bastien C., Mendelsohn P., Tiberghien, G., (1988). *Psychologie cognitive, modèles et méthodes*. PUG, Grenoble.
- Caverni, J.P., Bastien, C. – La psychologie cognitive face à ses enjeux. In J.P. Caverni, C. Bastien, P. Mendelsohn et G. Tiberghien (Eds.), *Psychologie cognitive, modèles et méthodes*, PUG, Grenoble, pp. 7-10, 1988.
- Demazeau, Y. (1995). From cognitive interactions to collective behaviour in agent-based systems. *Proceedings of the First European Conference on Cognitive Science*, Saint Malo, France, 4-7 April 1995.
- Dessalles, J.L. (1995). Les rôles du contexte et de la situation dans la cognition. *Actes de l'école d'été de l'ARC*, Bonas, volume C, 6-25.
- Ferber, J. (1995). *Les systèmes multi-agents : vers une intelligence collective*. InterEditions, Paris.
- Finin, T., Fritzson R., Mac Kay, D. and Mac Entire, R. (1994). KQML as an agent communication language. *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*. ACM Press, New York.
- Ghiglione, R. (1986). *L'homme communicant*. Colin, Paris.
- Ghiglione, R. (1983). Comprendre l'homme, construire des modèles. *Comportements*, 4. Editions du CNRS, Paris.
- Ghiglione, R. et Trognon, A. (1993). *Où va la pragmatique ? De la pragmatique à la psychologie sociale*. PUG, Grenoble.
- Gonzalez, M. – Sur les pratiques de validation de modèles en psychologie cognitive. In J.P. Caverni, C. Bastien, P. Mendelsohn et G. Tiberghien (Eds.), *Psychologie cognitive, modèles et méthodes*, PUG, Grenoble, pp. 443-464, 1988.
- Grice, P. (1979). Logique et conversation. *Communications*, 30, 57-72.
- Jacques, F. (1995). Dialogue, dialogisme, interlocution. *Séminaire interdisciplinaire de pragmatique, Institute for Advanced Studies*, The Hebrew University of Jerusalem, 24-26 mai 1995.
- Jacques, F. (1982). *Différence et subjectivité*. Aubier, Paris.
- Jacques, F. (1985). *L'espace logique de l'interlocution*. PUF, Paris.
- Le Ny, J.F. (1993). *Intelligence naturelle et intelligence artificielle*. PUF, Paris.
- Lecoutre, B., Rouanet, H., Denhière, G., – L'inférence statistique comme instrument de validation de modèles. In J.P. Caverni, C. Bastien, P. Mendelsohn et G. Tiberghien (Eds.), *Psychologie cognitive, modèles et méthodes*, PUG, Grenoble, pp. 383-394, 1988.

- Levinson, S. C. (1983). *Pragmatics*. CUP, Cambridge.
- Livet, P. (1994). *La communauté virtuelle. Action et communication*. Editions de l'éclat, Combas.
- Moeschler, J. — Théorie des actes de langage et analyse de conversation. In M. Charolles, S. Fischer et J. Jayez (Eds.), *Le discours. Représentations et interprétation*. PUN, Nancy, 1990
- Pesty, S. (1996). Systèmes multi-agents et simulation comportementale. *Rapport interne équipe MAGMA*.
- Rouanet, H. (1983). Modèles en tous genres et pratiques statisticiennes. In R. Ghiglione (Ed.), *Comprendre l'homme, construire des modèles. Comportements*, 4, 113-124, Editions du CNRS, Paris.
- Shannon, C. E., Weaver, W., (1949). *The mathematical theory of communication*. University of Illinois Press, Urbana.
- Searle, J. R. et Vanderveken, D. (1985). *Foundations of illocutionary logic*. CUP, Cambridge.
- Sperber, D. et Wilson, D. (1985). *La pertinence. Communication et cognition*. Editions de Minuit, Paris.
- Stewart, J. (1995). Les rôles du contexte et de la situation dans la cognition. *Actes de l'école d'été de l'ARC, volume C*, Bonas, 103-115.
- Tiberghien, G. — Modèles de l'activité cognitive. In J.P. Caverni, C. Bastien, P. Mendelsohn et G. Tiberghien (Eds.), *Psychologie cognitive, modèles et méthodes*, PUG, Grenoble, pp. 13-26, 1988.
- Tiberghien, G. — Questions de modélisation et de simulation cognitives. In J.-F. Le Ny (Ed.), *Intelligence naturelle et intelligence artificielle*. PUF, Paris, 1993.
- Trognon, A. — Les modèles linguistiques de la communication. In R. Ghiglione (Ed.), *L'homme communicant*, Colin, Paris, 1986.
- Trognon, A. (1991). L'interaction en général : sujets, groupes, cognitions, représentations sociales. *Connexions*, 57, 9-27.
- Trognon, A. et Brassac, Ch. (1988). Actes de langage et conversation. *Intellectica* 6, 211-232.
- Trognon, A. et Brassac, Ch. (1992). L'enchaînement conversationnel. *Cahiers de Linguistique Française*, 13, 76-107.
- Trognon, A. et Brassac, Ch. (1995). Formalizing the theory of intentionality. *Journal of Pragmatics*, 23, 555-562.
- Vanderveken, D. (1988). *Les actes de discours*. Mardaga, Bruxelles .
- Vanderveken, D. (1990). *Meaning and speech acts, formal semantics of success and satisfaction*. (2 tomes). CUP, Cambridge.

Analyse des Performances des Systèmes Collectifs

Olivier SIMONIN & Jacques FERBER

LIRMM - Université de Montpellier II,
161, rue Ada, 34392 Montpellier Cedex 5, FRANCE

Novembre 1997

Résumé

Dans ce papier nous proposons un modèle pour évaluer les performances quantitatives des systèmes multi-agents réactifs simples. Cette étude concerne les problèmes pouvant être décomposés en une série de tâches répétitives. Notre approche est basée sur une description analytique du système et sur le calcul de probabilité des événements.

Cette méthode est appliquée au problème des robots collecteurs d'échantillons. En particulier, nous calculons le gain apporté par l'emploi d'agents travaillant en équipes. L'obtention de formules générales nous permet d'analyser l'influence des différents paramètres. Puis nous comparons ces résultats théoriques avec des mesures de simulation. Ainsi, nous pouvons juger de la qualité du modèle et proposer des solutions pour l'étendre. Enfin, nous discutons des difficultés liées à la modélisation des systèmes collectifs complexes.

1 Introduction

Depuis quelques années, l'approche multi-agents s'est montrée intéressante dans le cadre de l'étude des systèmes complexes et de la recherche de nouvelles techniques de résolution. En vie artificielle, de nombreux travaux ont permis de simuler et d'analyser des sociétés d'insectes (cf. [2] [6] [7] [9]). Parallèlement, en robotique, des systèmes multi-robots ont montré leur efficacité pour effectuer des tâches telles que la collecte ou le tri d'éléments dans un environnement dynamique (cf. [1] [3] [7] [8]). Ces systèmes collectifs sont constitués d'agents réactifs, c'est-à-dire d'entités autonomes munies d'une architecture simple basée sur l'action-réaction (avec souvent la capacité de se spécialiser).

Toutefois, la plupart de ces études souffrent d'un manque de formalisme, en particulier, dans l'évaluation des performances. En effet, l'analyse des systèmes collectifs s'effectue généralement par expérimentation. Cette approche empirique, basée sur l'observation directe (de systèmes réels ou simulés) ne permet pas d'obtenir une analyse quantitative. L'objectif de cette étude est de proposer un modèle permettant d'analyser un système multi-agents dans son ensemble et d'en déterminer ses performances.

Dans un premier temps, il est intéressant de se pencher sur les différents travaux effectués sur le sujet. Des modèles mathématiques ont été proposés pour définir les comportements de sélection d'action (Maes [12], Brooks [3]), de spécialisation des agents (Drogoul et Ferber [9]) et du travail collectif chez

les insectes (Theraulaz [2], Deneubourg [6] [7]). Ces différentes études donnent des résultats qualitatifs sur le comportement des agents mais plus rarement sur les performances globales. Toutefois, J.L. Deneubourg et R. Beckers proposent un ensemble d'études sur les performances des systèmes collectifs [1] [7], mais celles-ci sont basées essentiellement sur l'expérimentation.

Notre modèle doit considérer l'ensemble du système, c'est-à-dire se situer au niveau de l'observateur. Mais de nombreuses difficultés liées à l'aspect aléatoire des événements apparaissent. En effet, comme le précisent Kolen et Pollack [10], si l'on cherche à décrire avec précision le comportement d'un système, il apparaît de nombreuses interactions chaotiques et imprévisibles. Tout le problème consiste donc à modéliser le système au bon niveau de description. Mais peut-on trouver un niveau de description simple sans perdre l'exactitude des comportements? Selon M. Mataric [13], les difficultés de cette analyse sont liées à deux propriétés intrinsèques aux SMA :

- L'action d'un agent dépend des états-actions des autres agents.
- Le comportement global du système est plus déterminé par les interactions entre les agents que par les comportements individuels.

Dans le cas des systèmes réactifs, ces deux propriétés sont moins fortes car les interactions entre agents sont élémentaires et locales. En particulier, les communications entre agents réactifs sont limitées. *C'est l'environnement qui se charge de la transmission des informations* (diffusion d'odeurs, signaux, création de chemins...). On peut donc ajouter une troisième propriété :

- Le comportement des agents dépend de l'état de l'environnement et de sa complexité.

Dans la section suivante, nous posons un ensemble d'hypothèses simplificatrices permettant de modéliser des SMA réactifs. Notre approche consiste à décrire les comportements collectifs en termes de probabilités. Cette démarche est possible avec des agents réactifs car leur comportement est simple (répétitif) et les communications limitées. La principale difficulté est de trouver un modèle suffisamment riche pour décrire le système et en déduire des probabilités exactes. Toutefois, cette démarche rencontre plusieurs difficultés liées à l'aspect dynamique des SMA.

Un problème fondamental apparaît : la modélisation de ces comportements nous impose-t-elle d'avoir une totale description du système? En effet, est-il nécessaire de connaître l'ensemble des caractéristiques du système (environnement, agents, processus émergents) pour le modéliser? Mais dans ce cas, l'analyse devient impossible ou sans intérêt. A l'inverse, la connaissance seulement d'une partie des paramètres peut-elle suffire à l'analyse? L'étude qui suit tente d'apporter une réponse à ces délicates questions.

Notre approche consiste à tenir compte de l'ensemble des paramètres mais de manière simplifiée. Cependant, comme le souligne M. Mataric [13], il n'est pas possible de déterminer les phénomènes émergents sans réaliser et tester le système. Un travail préalable d'observation est donc nécessaire pour connaître et comprendre ces phénomènes. De ce fait, nous analysons des systèmes ne reposant pas sur une résolution purement émergente. (cf [11] pour une présentation et une définition de l'émergence).

2 Méthode utilisée

2.1 Objectif

Le modèle que nous présentons a pour but d'évaluer la performance d'un système collectif et d'établir des comparaisons entre différentes techniques de travail. Pour mesurer cette performance, le temps est pris comme unique paramètre de référence. En effet, la durée du processus est une caractéristique essentielle pour évaluer l'efficacité d'un système artificiel chargé d'accomplir des tâches définies par le constructeur.

Cependant, pour comparer correctement l'efficacité de plusieurs techniques de travail, il faut nécessairement se placer dans les mêmes conditions. En particulier:

- L'environnement (noté ξ) et l'état initial sont strictement identiques;
- Le nombre d'agents est identique;
- Les agents possèdent des capacités équivalentes (type de communication et de tâche);
- Les hypothèses de modélisation sont les mêmes (cf. 2.2).

Dans ces conditions, nous pouvons calculer la différence de performance entre 2 techniques de travail appliquées au problème P (T_{m_1} et T_{m_2} étant les durées moyennes pour les deux techniques employées):

$$\text{gain}(P) = T_{m_1} - T_{m_2} \quad (1)$$

Pour étudier le signe du gain, il faut tout d'abord déterminer les temps de travail pour chaque technique. Le déroulement des événements étant non déterministe, nous pouvons seulement calculer statistiquement des *durées moyennes*.

2.2 Hypothèses

Les difficultés exposées plus haut sur la modélisation des SMA nous amènent à préciser le type de problème que nous étudions. Ce sont des systèmes collectifs à base d'agents réactifs (ils n'ont pas de carte "mentale" de l'environnement). Ces robots sont développés dans le but d'accomplir une tâche définie par le constructeur. Nous verrons à travers divers exemples, que la durée des processus est liée essentiellement aux tâches de recherche et de transport dans l'environnement. Les temps de traitement sur les objets immobiles sont négligés.

Les systèmes étudiés supposent les hypothèses suivantes vérifiées :

- Les agents sont indépendants (évitements négligés) et se déplacent à vitesse constante;
- Le temps est considéré discret. En une itération, tous les agents effectuent un déplacement élémentaire;
- Le comportement des agents est connu (algorithmes);

- P_0 le processus élémentaire (ou de référence), exemple : une sortie en exploration.
- T_0 la durée moyenne de P_0 (calculée suivant les algorithmes de déplacement)
- p_0 la probabilité que l'objet recherché soit trouvé durant le processus P_0 .
- T_{be} la durée moyenne du processus P_0 lorsque l'élément est détecté (parcours base-élément en phase d'exploration).

Le temps moyen de la tâche i est calculé en déterminant le nombre moyen de processus P_0 qu'un agent doit accomplir pour découvrir l'élément. Le processus P_0 suit une loi binomiale, chaque exécution est une réussite ou un échec. Si n_s est le rang du succès, pour découvrir l'élément, un agent isolé explore l'environnement pendant une durée moyenne approximative $n_s T_0$.

Cependant, lors du processus P_0 où l'élément est découvert, le chemin parcouru a nécessairement la durée moyenne T_{be} . D'où, un temps d'exploration égal à : $(n_s - 1)T_0 + T_{be}$

La probabilité de réussite au $n^{ième}$ processus vaut $p_0(1 - p_0)^{(n-1)}$. Nous avons montré que la durée moyenne d'exploration est égale à (cf. [15])

$$\left(\frac{1}{p_0} - 1\right)T_0 + T_{be} \quad (3)$$

Déterminons maintenant la probabilité $p_{0(N)}$ de découverte durant P_0 pour un ensemble de N agents. Les agents sont indépendants et cherchent à réaliser la même tâche. La probabilité pour que les N agents échouent en une sortie vaut $(1 - p_0)^N$. Donc, la probabilité que l'un d'eux réussisse est

$$p_{0(N)} = 1 - (1 - p_0)^N \quad (4)$$

Pour connaître la durée moyenne de recherche avec N agents, on applique la formule (3) en remplaçant p_0 par $p_{0(N)}$.

Remarque : si p_0 est une petite valeur, le développement limité de (4) au premier ordre nous permet d'obtenir $p_{0(N)} = N.p_0$.

Notons f_T la fonction temps, calculée suivant la probabilité p :

$$f_T(p) = \left(\frac{1}{p} - 1\right).T_0 + T_{be} \quad (5)$$

La valeur de p calculée suivant la formule (4) est comprise entre 0 et 1, donc $\frac{1}{p} - 1 \geq 0$. Ainsi $f_T \geq T_{be}$, ce qui est logique puisqu'il faut au moins une durée T_{be} pour découvrir la roche.

Bien entendu, la formule (5) n'est applicable que si la probabilité p est calculable. Nous verrons que la principale difficulté d'application de cette méthode réside dans l'obtention de probabilités précises.

Le processus de référence P_0 doit être choisi avec précaution. En effet, s'il est commun à de nombreuses tâches, les formules établies sont plus simples à étudier (car T_0 est une constante). La durée d'une tâche i exprimée avec la

formule (5) est de la forme $\chi_i \cdot T_0 + \psi_i \cdot T_{be}$. Les fonctions χ_i et ψ_i dépendent des paramètres du système et des *probabilités*, alors

$$T_P = T_0 \cdot \sum_{i=1}^n \chi_i + T_{be} \cdot \sum_{i=1}^n \psi_i$$

La durée globale du processus dépend ainsi des fonctions χ_i et ψ_i , c'est-à-dire des probabilités.

Peut-on appliquer ces formules si l'environnement contient plusieurs éléments à récupérer ?

Les éléments recherchés sont généralement répartis aléatoirement dans l'environnement. Considérons, dans un premier temps, la présence d'un seul élément. La probabilité, pour un agent, de le détecter vaut $\frac{S_0}{S}$, où S est la surface de ξ et S_0 la surface perçue lors du processus P_0 . Pour un chemin ch donné, la probabilité vaut

$$p_1(ch) = \frac{Surface(ch)}{S} \quad (6)$$

S'il y a m éléments dans ξ , $(1 - p_1(ch))^m$ est la probabilité que l'agent ne détecte aucun bloc en effectuant le parcours ch . Donc, la probabilité pour un agent de découvrir au moins un élément est égale à

$$p_m(ch) = 1 - (1 - p_1(ch))^m \quad (7)$$

Pour obtenir la valeur de p_m sur tous les chemins, il faut exprimer l'espace des chemins possibles et calculer la moyenne des $p_m(ch)$ sur tous ces chemins.

Une fois défini p_m , on peut calculer $p_{m(N)}$ la probabilité pour N agents (en appliquant la formule (4)), et en déduire la durée moyenne d'exploration avec la formule (5).

Nous présentons dans la section suivante, un problème type qui nous sert d'application à la méthode d'analyse présentée ci-dessus.

3 Des robots collecteurs

3.1 Description

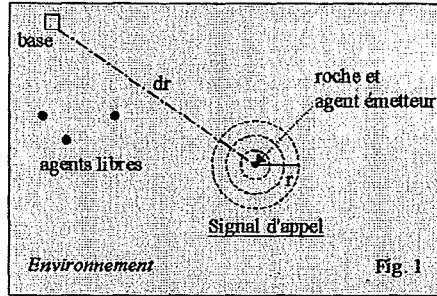
Ce problème classique de la robotique collective peut être énoncé comme suit : on suppose qu'il existe une base fixe à partir de laquelle plusieurs robots mobiles doivent explorer un espace *a priori* inconnu pour trouver, récupérer et transporter le minerai découvert jusqu'à la base. L'énoncé peut être étendu en remplaçant la phase de récupération par des traitements plus généraux. On peut résumer le processus général par l'algorithme suivant

1. **Exploration** : si l'un des agents détecte un bloc de roche alors il passe en phase 2.
2. L'agent doit **recruter** N agents (pour exécuter une à n actions). Puis passer en phase 3.
3. **Transport** de la roche vers la base.

Le problème peut s'exprimer en tâches :

$$T_P = T_{explo} + T_{recrut} + T_{trait} + T_{retour} \quad (8)$$

Pour obtenir une modélisation simple, nous considérons pour l'instant, qu'il n'y a qu'un élément de roche dans l'environnement (cf. fig. 1). Dans ces conditions, un agent est soit libre, soit engagé sur cet unique élément. Ainsi, le nombre d'agents libres dans l'environnement est calculable facilement.



Le niveau de spécialisation des agents doit être précisé car il influe grandement sur les comportements et les performances. Notons F l'ensemble des spécialités et $\Theta : \text{agent} \rightarrow P(F)$ la fonction donnant l'ensemble des spécialités F_i exécutables par un agent. $P(F)$ représente une partie de F .

Nous étudions 2 cas extrêmes :

1. les agents sont **totipotents** (ils peuvent tous accomplir l'ensemble des spécialités).
 $\forall \text{ agent}, \Theta(\text{agent}) = F$.
2. les agents sont **spécialisés** (chaque agent est capable d'accomplir seulement une fonction).
 $\forall \text{ agent}, \exists ! F_i \in F \text{ tel que } F_i \in \Theta(\text{agent})$.

3.2 Intérêt du problème

Plusieurs études ont été menées autour de la récupération de minerai par un ensemble de robots. En particulier, Brooks [4], Deneubourg [7] et Steels [16] ont proposé des robots réactifs utilisant les techniques des insectes. Ce problème présente l'avantage d'être clairement défini et de posséder de nombreuses méthodes de résolution (dépos de marques pour tracer des chemins [6] [16], formation de chaînes d'agents [8]). De plus, ce problème caractérise bien les processus nécessitant la présence de plusieurs agents pour effectuer une tâche (par exemple le transport).

Enfin, la plupart de ces solutions ne reposent pas sur des phénomènes émergents complexes. Ce sont les qualités de redondance et de robustesse des SMA réactifs qui sont employées, le système s'adapte suivant les configurations de l'environnement et les éventuelles pannes d'agents.

Finalement, ce problème représente bien la plupart des processus de la robotique collective. Cela nous permet de modéliser des tâches d'exploration, de regroupement (par signaux) et de traitements en milieu hostile.

3.3 Emploi des équipes

Dans les travaux cités ci-dessus, les agents sont des entités autonomes et individuelles. De ce fait, l'apparition d'un comportement en équipe reste un phénomène émergent. Or, nous supposons que la formation et le maintien d'un groupe avec des agents réactifs est partiellement implémentable dans l'architecture (cf. travaux [15]). Une des méthodes que nous étudions consiste à utiliser des équipes d'agents. Toutefois, si le constructeur dispose d'un nombre limité de robots, il peut hésiter à les grouper. En effet, *avec un nombre identique d'agents*, les équipes explorent une surface moins étendue qu'un ensemble d'agents libres. La surface de perception d'une équipe est plus faible que celle des agents libres et son déplacement plus "lent" (à cause du maintien du groupe). Donc, la durée totale du processus n'est pas forcément plus courte !

Nous appliquons le modèle probabiliste exposé plus haut pour évaluer le gain entre un système d'agents en équipes et un système d'agents libres.

4 Modélisation Analytique

4.1 L'environnement

L'environnement est ramené à un espace plan borné. Dans un premier temps, nous supposons l'existence *d'un seul bloc de roche dans l'environnement*. Ainsi, les agents ne peuvent travailler que sur le même élément de roche. Le nombre d'agents libres est égal à la différence entre le nombre total d'agents et celui des agents travaillant sur l'élément de roche. L'environnement est dépourvu d'obstacles et la zone d'exploration est bornée.

4.2 Les agents

Soit E l'ensemble des agents de cardinal N . Chaque agent peut se déplacer au plus d'une unité de mouvement par itération. Il peut percevoir la présence d'un élément (de roche) ou d'un agent dans un disque de rayon φ centré sur lui-même. Ainsi, les agents sont capables de s'éviter en modifiant leur trajectoire. On suppose que chaque agent utilise un système de contournement basé sur la perception de la direction et de la distance des autres agents proches (cf. solution proposée par K. Zeghal [17]). Ces capacités d'observation servent aussi aux agents qui se déplacent en groupe. A la manière des meutes d'animaux, ils suivent le groupe en se maintenant à une distance minimale des autres agents (cf. travaux de Reynolds [14]).

Les agents ne communiquent pas entre eux directement, par contre, ils peuvent émettre un signal d'appel pour attirer d'autres agents. Ce signal correspond à la diffusion d'une onde, détectable par les autres agents jusqu'à une certaine distance r (cf. fig. 1). L'intérieur du disque correspond à la zone de détection de l'appel.

Le processus P_0 de référence choisi est une sortie en exploration. Celle-ci est limitée à un éloignement maximum d_{max} de la base pour des raisons de ravitaillement des agents.

Exploration avec des agents libres:

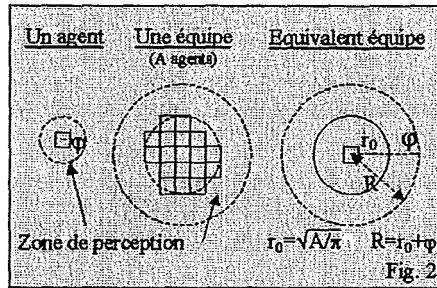
Muni d'un rayon de perception φ , l'agent a une probabilité p_φ (supérieure à p_0) de détecter la roche durant le processus P_0 . Soit T_φ le temps moyen pour qu'un seul agent libre découvre le bloc de roche :

$$T_\varphi = f_T(p_\varphi) \quad (9)$$

Exploration avec des agents groupés:

Dans un espace discret (cas des simulations), les agents sont localisés sur des cases adjacentes. Par conséquent, la surface de perception d'un groupe est inférieure à celle d'un nombre identique d'agents libres (cf. fig. 2). C'est évident, car les agents groupés perçoivent des zones communes.

Hypothèse : nous assimilons un groupe de A agents à un unique agent dont la surface de perception correspond à celle des agents groupés (cf. fig. 2). Soit R le rayon de perception du groupe.



$$R = r_0 + \varphi \text{ et } r_0 = \sqrt{\frac{A}{\pi}} \quad (10)$$

On suppose que les équipes effectuent des déplacements d'exploration semblables à ceux des agents libres. Cependant, les mouvements consacrés au maintien du groupe peuvent gêner le déplacement global. Ainsi, on pose α la probabilité pour une équipe d'effectuer un déplacement d'exploration par itération. Soit p_R , la probabilité qu'une équipe découvre la roche durant P_0 , calculée en fonction de son algorithme de recherche et de sa zone de perception. Finalement, le temps moyen d'exploration T_e pour qu'elle détecte la roche est :

$$T_e = \frac{1}{\alpha} f_T(p_R) \quad (11)$$

4.3 Agents totipotents

Suivant la modélisation que nous venons d'établir, nous pouvons déterminer la durée moyenne des phases de recherche, de regroupement (pour les

agents libres) et de transport. On peut éliminer la dernière phase car, pour les deux techniques, la durée et le nombre d'agents nécessaires sont identiques. Enfin, dans le cas d'agents totipotents, les temps de traitement sur la roche ne sont pas comptabilisés car ceux-ci sont négligeables face aux temps de recherche.

4.3.1 Durée pour un système d'agents libres

Phase 1 : détecter le bloc de roche

Il y a N agents lancés à la recherche du bloc. Etant donné que les agents sont indépendants, il faut un temps moyen $T_{explo.lib}$ pour que l'un d'entre eux détecte la roche.

$$T_{explo.lib} = f_T(p_\varphi(N)) \quad (12)$$

Phase 2 : Formation d'un groupe de A agents

L'agent qui a découvert le bloc doit recruter $A - 1$ autres robots. Pour cela, il émet un signal d'appel rapide. On représente ce signal par un disque de rayon r centré sur l'agent découvreur (cf. fig. 1). La probabilité p_r qu'un agent détecte le signal en une sortie est supérieure à p_0 . Cette probabilité est fonction des paramètres r , dr et de l'algorithme d'exploration (la durée de diffusion du signal est nulle). Notons T_{recrut} le temps moyen pour que les $A - 1$ agents soient recrutés sachant que le nombre d'agents libres diminue à chaque recrutement.

$$T_{recrut} = \left(\frac{\log \left(1 - \frac{A-1}{N-1} \right)}{\log(1 - p_r)} - 1 \right) \cdot T_0 + T_{be} \quad (13)$$

Le calcul de T_{recrut} est présenté en détail dans [15].

Notons que l'on trouve une formule de la forme $\chi_i \cdot T_0 + \psi_i \cdot T_{be}$.

Durée totale : $T_1 = T_{explo.lib} + T_{recrut}$

4.3.2 Durée pour un système d'agents en équipes

Phase 1 : détecter le bloc de roche

Les N agents lancés à la recherche du bloc sont groupés par A éléments, donc il y a $\frac{N}{A}$ équipes. Connaissant le temps moyen T_e pour qu'une équipe détecte le bloc, le temps d'exploration pour $\frac{N}{A}$ équipes vaut :

$$T_{explo.eq} = \frac{1}{\alpha} f_T(p_R(\frac{N}{A})) \quad (14)$$

L'équipe qui a découvert le bloc n'a pas besoin de "recruter" d'autres robots, donc la durée totale T_2 est égale à $T_{explo.eq}$.

4.3.3 Calcul du gain et Analyse

$$gain(P) = T_1 - T_2 \quad (15)$$

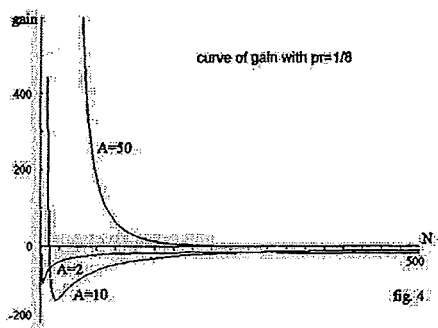
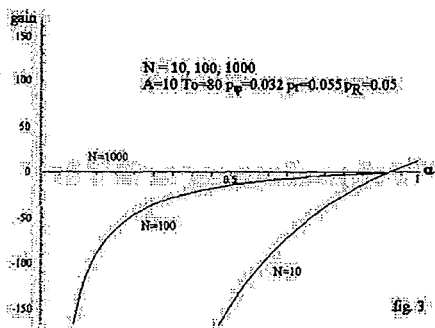
$$gain(P) = \left(2 - \frac{1}{\alpha} \right) \cdot T_{be} + T_0 \left[\frac{\log \left(1 - \frac{A-1}{N-1} \right)}{\log(1 - p_r)} + \frac{1}{p_\varphi(N)} + \frac{1}{\alpha} - \frac{1}{\alpha \cdot p_R(\frac{N}{A})} - 2 \right]$$

La lecture de l'équation (15) montre que le signe du gain dépend des probabilités p_r , p_φ , p_R , et du nombre d'agents. C'est à dire des paramètres φ , N , A , r , dr et α . Il n'est pas surprenant de constater que le gain dépend de nombreux paramètres car, malgré la simplicité de l'environnement, la description reste complexe.

Analyse par variation des paramètres :

Tout d'abord, notons que si le nombre d'agents est important, le gain est nul. En effet, pour $N \rightarrow \infty$, la formule (15) tend vers $(2 - \frac{1}{\alpha})T_{be} - T_0$. Ce résultat est logique, car la présence de nombreux agents rend les temps de détection et de formation très courts.

La valeur du coefficient α peut avoir un impact important sur le signe du gain. En effet, celui-ci représente la "vitesse" de déplacement des équipes. Le tracé du gain en fonction de $\alpha \in [0,1]$ est réalisé avec des valeurs arbitraires pour les autres paramètres (cf. fig. 3). Les courbes obtenues font nettement apparaître l'aspect décroissant du gain lorsque la valeur de α chute. Donc, pour qu'une équipe puisse rivaliser avec des agents libres, son déplacement doit être optimal (proche de 1).



L'utilisation des équipes apparaît intuitivement plus efficace qu'un système d'agents libres. Comparons donc les deux méthodes lorsque le système de recrutement est efficace. C'est à dire, lorsque le rayon du signal d'appel des agents libres est grand. Dans ce cas, la valeur de la probabilité p_r est grande. La courbe du gain est tracée pour $p_r = \frac{1}{8}$, $p_\varphi = 0.032$ et $p_R = 0.05$ (cf. fig 4). Le gain est sensiblement négatif pour les petites équipes et tend vers 0 quand $N \rightarrow \infty$. Les courbes débutent par une forte chute du gain, ceci montre que l'efficacité du système par recrutement croît "plus vite" que le système par équipes, quand N augmente. Finalement, les agents libres s'avèrent souvent plus efficaces que le système des équipes.

4.4 Agents spécialisés

Nous avons repris l'étude précédente avec des agents capables d'accomplir seulement une spécialité et de détecter la roche. Le traitement d'une roche consiste à effectuer séquentiellement toutes les spécialités.

Hypothèses : il y a n spécialités équitablement réparties parmi les agents. Donc, il y a $k = \frac{N}{n}$ agents de spécialité identique.

Nous avons étudié deux techniques de recrutement pour les agents libres :

- par appels successifs à chaque spécialité
- par un unique signal d'appel destiné à l'ensemble des spécialités.

Les formules obtenues sont de la forme $\chi_i.T_0 + \psi_i.T_{be}$ et dépendent des mêmes paramètres que l'étude réalisée avec des agents totipotents (excepté le nombre de spécialités).

Par exemple, le gain pour la 2^{ème} méthode :

$$\begin{aligned} gain(P) = & (n - \frac{1}{\alpha}).T_{be} + T_0 \left[(n - 1).(\frac{1}{p_r(N/n)} - 1) \right. \\ & \left. + \frac{1}{p_\varphi(N)} - \frac{1}{\alpha}(\frac{1}{p_R(N/n)} - 1) \right] \end{aligned} \quad (16)$$

Les tracés de courbes pour les deux méthodes montrent une plus grande efficacité du système des équipes, en particulier lorsque les agents libres utilisent une série d'appels à chaque spécialité.

4.5 Influence des paramètres

La modélisation effectuée nous permet de mieux saisir l'influence de chaque paramètre. Ils peuvent être classés en 3 catégories :

- les paramètres liés aux *capacités des agents* : φ, r, α . Leur variation modifie de façon importante les performances, car ils déterminent les probabilités de réussite.

- les paramètres de *cardinalité* : N, A . L'impact sur les performances est important pour une forte variation de leur valeur.

- les paramètres de l'environnement : d_{max}, dr . Ils peuvent apparaître dans le calcul des probabilités. Dans ce cas, leur influence peut être grande.

5 Résultats empiriques et comparaison

La modélisation du problème des robots collecteurs de minerai nous a permis d'analyser le gain apporté par l'emploi des équipes. Malgré la présence d'un seul bloc de roche, les formules obtenues dépendent de nombreux paramètres, en particulier des probabilités.

Il est intéressant de comparer les valeurs théoriques obtenues avec des mesures de simulations et ainsi de juger de la qualité du modèle. Nous examinons le problème présenté précédemment avec des agents totipotents. Les simulations sont réalisées à l'aide de la plate-forme JavaNetAgents.

5.1 Des agents indépendants

L'algorithme d'exploration se résume comme suit : les robots débutent leur exploration dans une direction aléatoire. Puis, ils effectuent régulièrement un

pas d'éloignement radial de la base, suivi d'un déplacement latéral circulaire de longueur dl . La distance dl est fonction de l'éloignement: plus l'agent est loin, plus cette distance est grande. On trouvera l'algorithme et les valeurs des paramètres dans [15]. L'espace exploré par les agents est un disque de rayon d_{max} centré sur la base. Cet espace ne contient qu'un seul bloc de roche se trouvant à une distance dr de la base.

Le temps T_0 de référence correspond à la durée moyenne d'une sortie. Cette valeur est facilement calculable à partir de l'algorithme d'un agent, on obtient $T_0 = 70.45s$.

Durant ce processus de sortie, la probabilité pour un agent de trouver la roche est p_φ (rayon de perception φ) La roche étant considérée ponctuelle, p_φ est le rapport : $\frac{d_m}{C}$ où C est la circonférence du cercle passant par la roche et d_m la distance moyenne parcourue sur ce cercle (par sortie). La probabilité de franchir ce cercle à une quelconque position est uniforme, pour un agent ou une équipe (cf. algorithmes).

$$p_\varphi = \frac{1}{2\pi \cdot dr} \cdot \left(\frac{dr}{2a} + 2\varphi - 1 \right) \quad (17)$$

Le pas d'éloignement vaut $2 \cdot \varphi$, ainsi un agent ne revient pas sur une zone déjà explorée.

Il reste à calculer p_r , la probabilité pour un agent de détecter un signal de recrutement.

1er cas: le rayon du disque est faible comparé à dr

Nous pouvons considérer que le signal est ponctuel et appliquer le même calcul que celui de la détection de la roche. On remplace φ par r .

2eme cas: le rayon du disque n'est pas négligeable devant dr

Dans ce cas, on pose $p_r = \varepsilon(r, dr) \cdot p_0$, où ε est fonction du rayon de recrutement et de dr . Le calcul de ε est particulièrement complexe car l'évaluation de p_r nécessite des calculs de probabilité et de géométrie difficiles qui sortent du contexte de cette étude.

La simulation est réalisée avec des agents ne pouvant émettre qu'un court signal de recrutement. La valeur de p_r est donc calculée comme p_φ (cf. formule (21)). Or, cette approximation implique que la véritable valeur de p_r est légèrement supérieure.

Le système comprend 27 agents, lorsque l'un d'eux découvre la roche, il doit former un groupe de 9 éléments.

Les mesures sont réalisées en comptabilisant le nombre de déplacements élémentaires (itération) effectués par les agents. De nombreuses simulations sont effectuées pour obtenir des moyennes significatives.

Tableau des résultats:

Process.	essais	simulations	théorique
T_0	400	69	70.45
$T_{explo.lib}$	500	102	105.9
T_1	100	586	749

Les mesures obtenues pour T_0 et $T_{explo.lib}$ sont proches des valeurs théoriques, toutefois les approximations effectuées dans le calcul de T_1 expliquent

l'infériorité des mesures de T_1 (la probabilité devrait être plus grande, donc le temps théorique plus court).

5.2 Des agents en équipes

Les agents forment 3 équipes de 9 éléments, on en déduit le rayon R de perception d'une équipe. Le pas de déplacement de l'équipe qui vaut $2R$ est forcément supérieur à 2φ . Pour conserver un temps T_0 de sortie, les équipes effectuent un déplacement latéral plus long. On suppose que le déplacement des équipes est optimal: $\alpha = 1$.

Valeur théorique et résultats des simulations :

Process.	essais	simulations	theorique
$T_{explo.team}$	100	404.3s	469.5s

L'écart entre les deux valeurs est faible et s'explique logiquement par l'approximation sur p_R .

Valeurs du gain :

	libres	equipes	gain(P)
théorique	749	469.5	279.5
simulations	586	404.3	181.7

Sur l'exemple précis que nous avons décrit, le gain est positif (calcul théorique et mesures). Toutefois, l'écart entre le gain théorique et le gain mesuré est important. Les différences entre les 2 méthodes sont dues

1. aux approximations dans les calculs théoriques,
2. à la nécessité de réaliser de nombreuses simulations pour obtenir des moyennes significatives,
3. aux écarts inévitables entre un modèle continu et une simulation dans un espace discret.

6 Discussion

Le problème que nous avons modélisé et simulé met en évidence l'application de notre méthode. Mais les difficultés mathématiques liées aux calculs des probabilités sont un réel obstacle pour appliquer cette méthode à des systèmes plus complexes. Malgré tout, nous pouvons examiner des situations plus réalistes pour déterminer si les résultats précédents sont généralisables.

- Présence d'obstacles dans l'environnement

De façon générale, les petits obstacles ne posent pas de problèmes aux deux techniques, les agents libres et les équipes les "traversent" aisément. Ce sont les reliefs importants qui sont plus complexes à contourner. On peut envisager de munir les agents (et les équipes) d'algorithmes de suivi de contours (cf. [5]). Mais cette solution peut être néfaste aux agents libres. En effet, un agent libre ayant détecté un signal de recrutement peut le perdre en contournant un obstacle. Pour adapter ces changements au modèle, il est nécessaire de

réduire le paramètre α et de revoir le calcul des probabilités p_φ et p_R . Mais il semble difficile d'évaluer cette variation.

- Pertes d'agents (pannes, blocages)

Les systèmes réactifs sont très redondants, ce qui assure un bon fonctionnement de la première technique. Par contre, les équipes peuvent se trouver incapables d'accomplir une tâche commune lorsque l'un des agents tombe en panne, dans ce cas, les agents sont obligés de rentrer à la base. La présence d'agents libres prêts à remplacer un agent défectueux dans l'environnement peut être une solution intéressante. On peut aussi envisager de munir les équipes de robots supplémentaires. Dans ces conditions, les performances établies précédemment restent valables.

- Présence de plusieurs blocs de roche

Une première réflexion fait apparaître la faiblesse du système des agents libres lorsque ceux-ci sont peu nombreux comparés au nombre de blocs. En effet, le système peut se bloquer (deadlock) si plusieurs agents souhaitent former des groupes simultanément alors qu'il n'y a pas assez d'agents libres pour les constituer. Dans le cas inverse, où les agents sont plus nombreux que les blocs de roche, la technique est très coûteuse. Toutefois, si l'on souhaite évaluer la durée des tâches pour N agents et m roches, il faut calculer la probabilité générale de détection d'une des roches (cf. formule (7)) et appliquer la formule (4). Cependant, le développement mathématique est trop important pour être présenté dans ce papier.

Jusqu'à présent, nous avons étudié des systèmes simples où les agents ne communiquaient pas entre eux. Mais, dans la plupart des systèmes réactifs, les agents peuvent échanger des informations. Cette capacité est particulièrement importante lorsque les éléments à collecter sont groupés. Ainsi, le processus suit le schéma général suivant :

P = Détecter elt. + Communiquer position + Traitement

Nous savons maintenant évaluer la durée de la première tâche, les deux autres sont simples à déterminer car elles sont fonction de la durée base-élément T_{be} . Cette technique est une amélioration de celle des agents libres, donc sa durée est nécessairement plus courte. Finalement, les résultats de la première technique nous donnent une borne supérieure pour la durée globale du processus général.

En appliquant le modèle au problème des robots collecteurs de minerai, nous avons établi les performances de différentes techniques de travail (agents libres, équipes). L'obtention de formules générales a rendu possible l'étude des divers paramètres du système (tracés de courbes). Cependant, le modèle n'est pas applicable à tous les systèmes. En effet, dans notre démarche, nous avons pris en compte l'ensemble des événements et limité l'étude à des processus séquentiels. Il apparaît donc difficile d'appliquer notre méthode à un ensemble de processus s'exécutant en parallèle et dont les agents peuvent entrer en interaction. Néanmoins, les résultats obtenus montrent que le modèle peut analyser correctement certains processus des systèmes collectifs.

7 Conclusion et Perspectives

La complexité des systèmes étudiés nous a conduit à limiter l'étude à des situations simplifiées (environnement sans obstacles, indépendance des agents). Ceci nous a permis de faire une étude analytique du problème des robots collecteur de minerai. Cependant, les performances sont exprimées par des durées moyennes. Pour un état initial donné, nous déduisons une simple valeur, temps moyen supposé du processus général. Mais, cette durée moyenne est-elle une information suffisante pour évaluer un système? A l'évidence, elle ne contient aucune information qualitative sur le système et ne permet pas de connaître les variations sur différentes exécutions. En effet, si l'on examine plusieurs exécutions d'un système, il résulte des comportements et des durées pouvant être très différents. Il faut donc relativiser les résultats obtenus. Toutefois, pour les systèmes réactifs composés d'agents au comportement très répétitif, les valeurs calculées sont significatives.

L'étude montre qu'il est difficile d'intégrer au modèle des événements de type pertes d'agents ou échanges d'informations entre agents. Or, la modélisation de situations simples s'est déjà montrée complexe, en nécessitant la connaissance de l'ensemble des caractéristiques du système. Donc, la recherche de formules générales décrivant des systèmes plus riches apparaît une entreprise particulièrement ardue.

Malgré les difficultés évidentes d'application du modèle à des processus parallèles interagissant, l'approche probabiliste, couplée avec des hypothèses simplificatrices, semble une première étape prometteuse dans la modélisation des systèmes collectifs.

Remerciement

Nous remercions Vincent Decugis pour sa participation à l'élaboration du modèle probabiliste exposé dans l'article.

Références

- [1] Beckers R., Holland O.E. and Deneubourg J.L. (94): "From local actions to global tasks: Stigmergy and collective robotics" In *Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, Rodney A. Brooks and Pattie Maes, The MIT Press.
- [2] Bonabeau E., Theraulaz G. (94): "Intelligence collective", Hermes (Ed.).
- [3] Brooks R. (86): "a Robust Layered Control System for a Mobile Robot" In *IEEE Journal of Robotics and Automation*, RA-2, April 1986.
- [4] Brooks R. (90): "Elephant Don't Play Chess". In *Robotics and Autonomous Systems*. 6.
- [5] Decugis V. et Ferber J. (96): "Architecture Multi-agents de Robots Mobiles Autonomes" In *Les journées de Rochebrune 97*, France.
- [6] Deneubourg J.-L. et Goss S. (89): "Collective Patterns and Decision-Making". In *Ecology, Ethology and Evolution*. 1, p. 295-311.

- [7] Deneubourg J.-L., Goss S., Sendova-franks A., Dtrain C. et Chretien L. (91): "The Dynamics of Collective Sorting Robot-like Ants and Ant-like Robots." In *From Animals to Animats*, Paris, J.-A. Meyer et S. W. Wilson (Ed.), p. 356-363, MIT Press.
- [8] Drogoul A. et Ferber J. (92): "From Tom Thumb to the Dockers: Some Experiments with Foraging Robots." In *From Animals to Animats: Second Conference on Simulation of Adaptive Behavior (SAB 92)*, Hawaii, J.-A. Meyer, H. Roitblat et S. Wilson (Ed.), MIT Press.
- [9] Drogoul A. et Ferber J. (94): "Multi-Agent Simulation as a Tool for Modeling Societies: Application to Social Differentiation in Ant colonies." In *Artificial Social Systems*, vol. 830, C. Castelfranchi et E. Werner (Ed.), p. 3-23, Berlin, Springer-Verlag.
- [10] Kolen, J.F. & Pollack, J.B. (93): "Apparent Computational Complexity in Physical Systems", in *Proceedings of the fifteenth Annual Conference of the Cognitive Science Society*, Boulder, Colorado, p617-622.
- [11] JFIADSMA'97 Groupe de travail "Collectif" IAD/SMA de AF-CET/AFIA: "Emergence et SMA", La Colle sur Loup, 2-4 avril 1997.
- [12] Maes P. (90): "Situated Agents Can Have Goals" In *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*, Edited by P. Maes, MIT Press, Cambridge.
- [13] Maja J Mataric (94): "Interaction and Intelligent Behavior" These, MIT Department of Electrical Engineering and Computer Science.
- [14] Reynolds C. (87): "Flocks, Herds and Schools: A Distributed Behavioral Model" In *Computer Graphics*. 21(4), p. 25-34.
- [15] Simonin O. (97): "Regroupement et travail en équipe dans les systèmes multi-agents réactifs", Rapport de DEA, LIRMM, Montpellier, Juin 1997.
- [16] Steels L. (89): "Cooperation between Distributed Agents through Self-Organization." In *Decentralized A.I.*, Y. Demazeau et J.-P. Müller (Ed.), Elsevier/North-Holland.
- [17] Zeghal K. et Ferber J. (93): "A Reactive Approach for Distributed Air Traffic Control" In *International Conference on Artificial Intelligence & Expert Systems*, Avignon, France, 1993.

La modélisation ne se réduit ni à l'étude des modèles ni à l'exploration de leurs simulations

Pierre Saurel,
saurel@iut.univ-evry.fr

Université d'Évry et Laboratoire du CRÉA, École Polytechnique

Introduction

Dans cet article nous souhaitons esquisser dans un ensemble cohérent, différents éléments que rencontre tout praticien de la modélisation. Dans notre démarche nous nous appuyerons essentiellement sur trois figures. Ces schémas mettent en relation des domaines, comme la théorie ou les expériences. La définition de ces domaines et la nature des relations entre ces domaines ont fait et font encore l'objet de vifs débats en épistémologie¹. Notre propos n'est ni de prendre part à ces débats et de se positionner en statuant par exemple sur la nature de ces relations, ni même de préciser ou de présenter ces débats. *Ces figures sont donc pour nous un simple guide, un aide-mémoire, qui permet de garder en tête et de visualiser globalement les difficultés de la modélisation.* Ceci n'est que peu de chose en comparaison avec ce qui pourrait être présenté en épistémologie, mais ces quelques schémas nous ont permis de garder une ligne de conduite cohérente lorsque nous avons cherché à modéliser avec Wolf-Dieter Eberwein certains aspects de la mobilisation de masse en RDA en 1989 [Eberwein et Saurel, 1995]. De même que nous ne présentons pas explicitement les enjeux épistémologiques, nous avons choisi de ne pas citer les auteurs qui participent ou ont participé à ces débats. On pourra reconnaître parfois leur influence dans ce texte. **Notre objectif principal dans ce texte est de montrer que la modélisation est une activité qui ne peut pas être réduite à l'analyse d'un modèle² ou d'une simulation.** Le modèle à proprement parler est au centre de la modélisation, mais il n'en constitue qu'un des très nombreux aspects.

¹ On pourrait citer Nancy Cartwright, Paul Feyerabend, Ian Hacking, Thomas Kuhn, Imre Lakatos, Hilary Putnam, etc.

² Nous appelons modèle un système axiomatique muni d'une logique. Le système axiomatique de même que la logique peuvent éventuellement être implicites ou non totalement explicites.

I Une vision classique ou « popperienne » de la démarcation entre les théories scientifiques et les autres théories

La conception qui permet selon Popper d'affirmer le caractère scientifique de certaines théories [Popper, 1935], même si elle est désormais un peu ancienne, peut être considérée comme le cadre classique dans lequel se placent implicitement les scientifiques et en particulier les modélisateurs. Cette conception s'articule autour de différents aspects de la justification du caractère scientifique des théories et en particulier autour de la *détermination d'un critère de démarcation entre la connaissance scientifique et d'autres formes de connaissance*. Ce qui nous intéresse ici n'est pas le travail particulier de Popper mais le cadre conceptuel classique dans lequel il conçoit la possibilité de remettre en cause les théories. Nous ne retiendrons de la conception popperienne, pour notre propos, que les éléments qui apparaissent sur la figure 1.

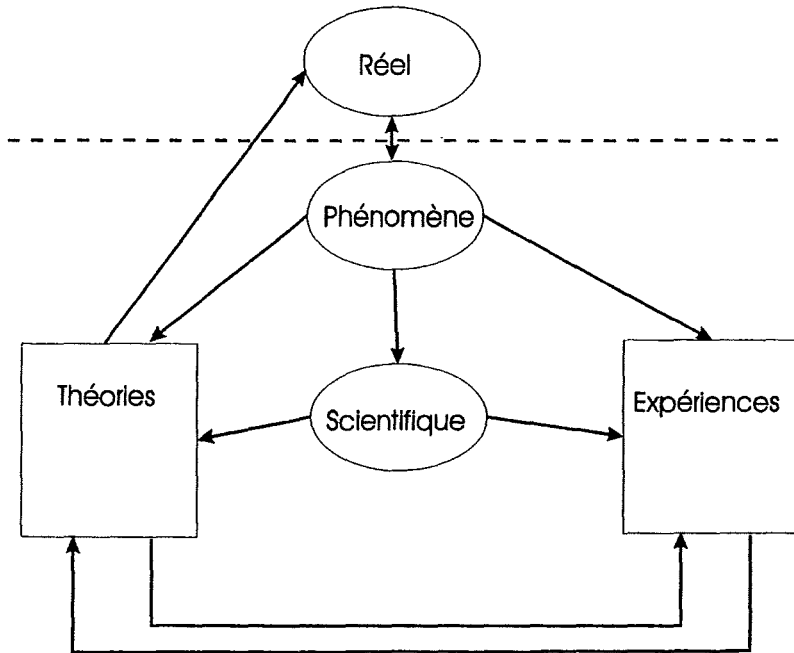


FIGURE 1 : *La conception classique de la confrontation des théories aux expériences.*

On peut interpréter grossièrement la figure 1 en disant que la connaissance scientifique est issue de la **confrontation de deux domaines différents** : d'une part les expériences³ et d'autre part les théories⁴. Des hypothèses sont formulées dans le cadre de théories. Des déductions sont réalisées à partir de ces hypothèses. Ces déductions sont alors confrontées aux faits expérimentaux. Pour des raisons de commensurabilité, la confrontation passe éventuellement par une mise au même format. Ce format est généralement obtenu par l'intermédiaire de la mesure qui permet d'associer un élément numérique, à une situation expérimentale. Il suffit alors de confronter cette mesure obtenue dans le cadre expérimental aux différentes valeurs qui ont pu être déduites à partir de l'hypothèse exprimée dans le cadre de la théorie. Une adéquation entre ces valeurs numériques permet de conforter une théorie, alors que des valeurs incompatibles entraînent la falsification de la théorie [Popper, 1935].

II Les quatre domaines de la modélisation

1 Les nouveaux domaines de la modélisation

Dans le contexte présenté précédemment, la notion de modèle au sens propre, c'est-à-dire de système axiomatique muni d'une logique, n'est pas évoquée. Ainsi dans la pratique actuelle de la modélisation, **le modèle est un nouveau domaine** qui s'ajoute aux deux domaines mis en évidence précédemment, *la théorie et l'expérience*.

Dans son livre [Popper, 1935], Popper n'utilise jamais le terme de modèle. Popper fait référence aux aspects linguistiques de l'argumentation scientifique, mais il ne s'appuie jamais sur les systèmes axiomatiques au sens

³ Dans ce texte le terme d'expérience est réservé pour parler de situations qui mettent en œuvre des phénomènes « naturels » dont le comportement n'est pas intégralement prédéfini par un observateur humain. Le terme ne saurait renvoyer aux « expériences computationnelles » qui ne partagent que leur nom avec les expériences à proprement parler. De fait, les expériences computationnelles n'ont rien d'expérimental puisqu'elles sont complètement déterminées *a priori* par le programmeur. Elles ne sauraient permettre de découvrir quoi que ce soit sur le « monde des phénomènes ». Ceci ne retire pas leur intérêt aux simulations, bien au contraire. Nous voulons simplement insister sur le fait qu'il faut clairement distinguer ces deux types d'expériences³.

⁴ Nous parlerons ici de théorie à propos d'un ensemble de concepts présentant certaines relations de dépendance mutuelle et de cohérence relative. Ces concepts devront de plus avoir été élaborés en corrélation avec un corpus important d'éléments expérimentaux. Nous ne faisons que suivre ici la définition d'une théorie scientifique selon Cavallès, pour qui « les concepts ne sont pas des catégories *a priori*, hors du temps de l'activité ; ils vivent et font système, et ces systèmes se transforment. L'histoire nous montre ces transformations » [Sinaceur, 1994] [p. 89]. Nous conservons le sens de théorie comme système particulier de concepts, mais nous ne préjugeons pas, en utilisant ce terme, du fait que ces concepts sont ou non en correspondance avec des catégories *a priori* ou qu'ils se constituent et se modifient par les relations qu'ils entretiennent.

fort⁵. Ceci transparait en particulier dans son utilisation systématique du terme *énoncé*, à propos des produits de la science, alors qu'il ne parle jamais d'énoncés symboliques. Toujours dans cette optique, il faut noter l'utilisation faite par Popper de la notion d'axiome. Cette notion chez Popper reste très proche d'un énoncé de la langue naturelle qui serait fixé comme une connaissance *a priori*. On reste ainsi dans une conception classique de la notion d'axiome.

La conception classique de la modélisation ne tient pas compte non plus de l'utilisation nouvelle et massive de l'ordinateur. En particulier l'ordinateur a imposé, dans la pratique de la modélisation, l'introduction de la **simulation comme quatrième domaine** qui est venu s'ajouter à la conception classique. Nous entendons par simulation la donnée d'un couple comprenant une machine abstraite et un modèle compatible avec cette machine⁶. Désormais la simulation est devenue un domaine important contribuant à la connaissance scientifique. Elle peut apporter différents éléments nouveaux. En particulier, elle impose de rendre véritablement effectif les systèmes axiomatiques étudiés. En effet, si la machine est donnée, le propos pour obtenir une simulation consiste à transformer le système axiomatique initial en un nouveau système axiomatique compatible avec la machine. On peut également choisir une autre stratégie pour obtenir une simulation : elle consiste cette fois à refuser de modifier le système axiomatique. Il convient alors, le système axiomatique étant donné, de construire, si c'est possible, une machine telle que le système axiomatique soit compatible avec cette machine. Cette méthode permet également d'obtenir un couple sous la forme recherchée, c'est-à-dire une simulation. Il est tout à fait clair que dans la pratique les modélisateurs préfèrent utiliser une machine de référence et transformer le système axiomatique initial si nécessaire.

Dans les conceptions classiques, un système d'axiomes, ou plus généralement d'énoncés, est porteur de tout le sens qu'il est susceptible de

⁵ Il est habituel de distinguer deux conceptions de l'axiomatique. La première qualifiée souvent de « naïve » [Blanché, 1955] que l'on trouve par exemple chez le premier Hilbert [Hilbert, 1899] ou chez Cantor est celle pour laquelle tout ou partie des axiomes n'est pas exprimé en termes purement symboliques. La seconde au contraire, postérieure aux travaux de Zermelo, Fraenkel ou von Neumann et que l'on retrouve également dans les travaux de Hilbert des années 1930, considère les axiomes comme des assertions purement symboliques qui explicitent les règles du jeu symbolique. Nous réservons le terme de système axiomatique au sens fort à propos de cette seconde conception de l'axiomatique.

⁶ La simulation est donc à mettre en parallèle avec la notion de système formel selon Gödel. Le terme de système formel tel qu'il a été défini par Gödel, suite aux travaux de Turing consiste très précisément à se donner un couple comprenant un modèle (c'est-à-dire un système axiomatique muni d'une logique) et une machine abstraite. Le modèle devant être compatible avec cette machine, c'est-à-dire descriptible selon les opérations élémentaires de cette machine.

généraler. Les axiomes sont alors le résumé suffisant de leur propre sens. Aussi, il peut sembler « inutile » d'analyser le fonctionnement effectif d'un système d'axiomes. Il suffit d'étudier les axiomes eux-mêmes. La question générale est de savoir, étant donné un énoncé particulier exprimé dans le cadre d'un système axiomatique, s'il est ou non conséquence du système d'axiomes. Répondre à cette question et déterminer ainsi les conséquences d'un système axiomatique n'est pas toujours chose aisée à réaliser à partir des axiomes eux-mêmes. La question générale est tellement ardue qu'on peut parfois proposer des énoncés écrits dans les termes du système axiomatique et dont on ne peut pas dire intrinsèquement, c'est-à-dire dans le cadre de cette axiomatique, s'ils sont ou non des conséquences possibles du système d'axiomes. La simulation en imposant la compatibilité du modèle avec une certaine machine permet de préciser les comportements qui sont conséquence effective, sur cette machine, du système d'axiomes. La simulation permet donc parfois d'analyser les conséquences possibles, sur une machine explicite, d'un système d'axiomes. En reprenant les termes de Popper, la simulation permet parfois, par son caractère opératoire lié à une machine, de connaître certaines conséquences d'une hypothèse, si l'on considère une hypothèse qui a été explicitée comme un axiome supplémentaire dans un système d'axiomes déjà existant.

2 Quelques caractéristiques de la modélisation

La conception classique comportait deux domaines et *la falsification d'une théorie avait lieu par la constatation ou la démonstration qu'une des hypothèses de la théorie avait des conséquences en contradiction avec les données*. Désormais, selon nous, la notion de modélisation ou du moins l'utilisation de ses différents domaines s'est substituée à la conception classique. **La modélisation comporte quatre domaines : la théorie, l'expérience, le modèle et la simulation** (figure 2).

De même que la conception classique n'était pas une simple juxtaposition ni une succession de la théorie et de l'expérience, la modélisation ne peut pas être conçue comme une simple juxtaposition ni une succession de ses quatre domaines. *La modélisation nécessite en fait un va-et-vient incessant entre ses quatre domaines*. La modélisation n'a donc rien de figé comme peut l'être un modèle. **La modélisation est une activité**⁷. *La*

⁷ Il ne s'agit pas là de l'activité du scientifique qui modélise, mais d'une interaction entre les domaines, intrinsèque à la modélisation. On peut ainsi comprendre pourquoi dans la pratique les modélisateurs ne font que rarement des modélisations complètes. Ils travaillent principalement sur certains domaines d'une modélisation complète potentielle et précisent ainsi certaines conditions de possibilité d'une telle modélisation.

modélisation nécessite des interactions fortes et permanentes entre ses quatre domaines. Les domaines de la modélisation sont dans une relation d'interdépendance. Ils doivent être élaborés, construits et étudiés dans une interaction réciproque. La modélisation n'est donc pas la simple écriture d'une équation à partir de données ou d'expériences facilement interprétables, comme une conception « naïve » aurait tendance à nous le laisser croire.

Une caractéristique importante de la modélisation est la nécessité de définir un objectif précis et de construire ensuite les différents domaines en fonction de cet objectif. Dans ce contexte, **modéliser c'est choisir**. La modélisation a toujours lieu à propos de quelque chose. On parle de modèles, mais il faut plutôt insister sur le fait qu'on a toujours un **modèle de quelque chose**. On ne peut alors pas considérer un modèle indépendamment de la modélisation qui l'a permis sans le dénaturer et lui enlever une partie importante des caractéristiques qui le définissent et en particulier de sa « sémantique ». Hors de son contexte, le modèle peut être interprété de différentes manières éventuellement contradictoires. *Seule la modélisation dans son ensemble permet d'élaborer la sémantique générale selon laquelle, en retour, on pourra interpréter les quatre domaines qui la constituent.*

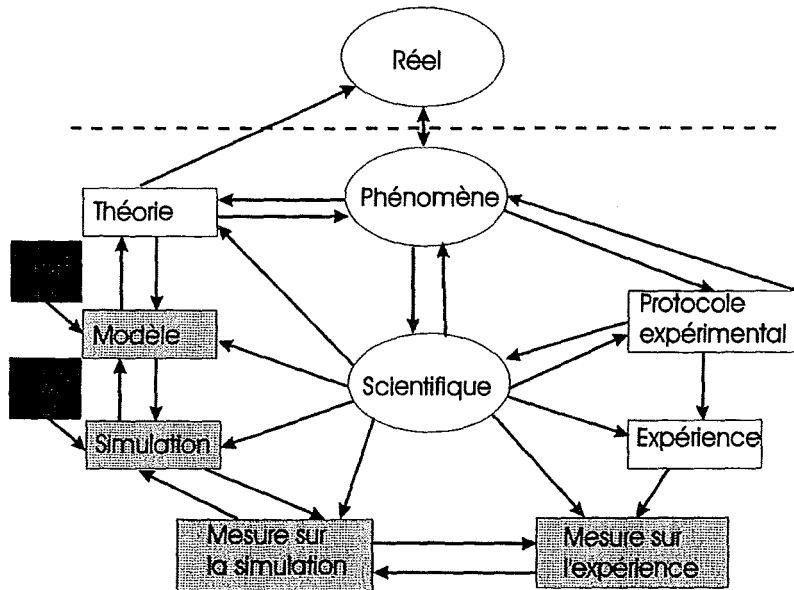


FIGURE 2 : Pratique actuelle de la confrontation des théories aux expériences : les quatre nouveaux domaines ainsi que l'analyse du modèle et l'analyse de la simulation.

3 Un cadre d'interprétation

La figure 2 permet d'interpréter facilement les différentes méthodologies que l'on rencontre habituellement chez ceux qui contribuent à la modélisation d'un phénomène.

On peut par exemple *interpréter la technique des gaz sur réseau dans ce cadre*⁸. Il apparaît clairement que tous les domaines de la modélisation ne rentrent pas en jeu. En effet, la constitution des équations de Navier-Stokes en fonction des phénomènes de mécanique des fluides en jeu n'intervient pas dans cette technique. La technique des gaz sur réseau consiste à proposer un système simulable à partir d'équations formelles. Les deux domaines en jeu dans ce cadre sont donc le modèle et la simulation. La théorie n'intervient pas car, dans le cadre des gaz sur réseau, les équations sont déjà connues et n'ont pas à être établies. Une simulation est ensuite possible qui peut éventuellement être confrontée au comportement du fluide particulier qui est étudié. Cependant la confrontation avec l'expérience n'entre pas dans le cadre de la technique des gaz sur réseau à proprement parler. Ainsi cette technique met seulement en jeu deux domaines de la modélisation. **De ce fait, il serait tout à fait abusif de parler de modélisation dans le cadre de la technique des gaz sur réseau, même si elle implique le modèle lui-même.**

Le modèle du tas de sable [Bak, Tang et Wiesenfeld, 1988] est un peu différent. Avec ce modèle on est en présence d'équations de modifications discrètes compatibles avec les machines abstraites que nous manipulons habituellement. Ce modèle peut ainsi être simulé et cette simulation permet de mettre en évidence un concept théorique, la notion d'état critique auto-organisé. Ainsi, deux domaines de la modélisation ont été mis en jeu, le modèle et la simulation. On peut même y ajouter la théorie si on considère qu'un concept théorique unique peut faire office de théorie. Cependant, aucun lien concluant n'a été établi avec des expériences. Même si on parle de modèle, il ne renvoie à aucune expérience. De fait, les quatre domaines de la modélisation n'interviennent pas ici. **On ne saurait alors**

⁸ Grossièrement cette technique consiste à proposer des versions discrètes d'équations aux dérivées partielles classiques en physique dont l'expression proche de celle d'automates permet aisément une étude et des simulations sur ordinateur [Frisch, Hasslacher et Pomeau, 1986], [Doolen *et al.*, 1990].

parler de modélisation au sens propre dans le cadre du modèle du tas de sable, même si un modèle est en jeu.

III Différentes sophistications actuelles

On trouve actuellement, dans la pratique, différentes sophistications aux aspects de la modélisation présentés précédemment. Nous les avons rassemblées sur la figure 2 et sur la figure 3.

1 *L'analyse du modèle et l'analyse de la simulation*

Aux quatre domaines précédents, on constate en analysant les travaux de certains modélisateurs, qu'il convient d'en adjoindre deux nouveaux que l'on retrouve sur la figure 2. Ceux-ci font leur apparition depuis peu et nous ne les considérerons pas directement comme partie intégrante de la modélisation à l'heure actuelle. *Il s'agit d'une part de l'analyse du modèle et d'autre part de l'analyse de la simulation.*

Jusqu'à présent, les modèles étudiés étaient suffisamment « simples » pour que l'on ne soit pas obligé de distinguer l'analyse du modèle du modèle lui-même. Les modèles étudiés sont bien souvent tellement sophistiqués que les méthodes mathématiques existantes ne permettent pas de les résoudre directement. Ce type de situation a déjà été rencontré historiquement, et cela amenait généralement la constitution d'outils mathématiques complètement nouveaux pour étudier ces modèles. Ce fut le cas par exemple dans les travaux de Volterra [Volterra, 1931].

La stratégie choisie à l'heure actuelle est par contre tout à fait différente. Elle consiste, lorsque le modèle est trop sophistiqué, à se ramener à un autre modèle, plus facile à étudier mathématiquement. On définit un nouveau modèle à partir du modèle initial, en espérant que leur comportement sera identique. L'analyse en champ moyen, par exemple, consiste par le biais d'une approximation à définir en fait un nouveau modèle. On sait démontrer dans certains cas, comme pour le modèle d'Ising en dimension une, que le comportement du modèle « approché » est différent de celui du modèle initial [Diu *et al.*, 1989]. Même qualitativement, les deux modèles n'ont pas le même comportement. On parle de modèle approché, mais ces modèles n'ont en fait rien de proche et il vaudrait mieux parler d'un nouveau modèle ou de perturbation du modèle initial.

L'analyse en champ moyen est un outil qui est aujourd'hui utilisé quasi-systématiquement pour étudier les comportements des gros systèmes fortement interconnectés. En fait cette analyse en champ moyen consiste à étudier non pas le modèle initial mais un nouveau modèle dont on peut

prédire le comportement. C'est en ce sens que nous parlons d'analyse du modèle. On peut citer comme exemple l'analyse du modèle d'évolution de Kauffman et celle des sophistications proposées par Bak [Bak, Flyvbjerg et Lautrup, 1992]. Partant du modèle initial, l'analyse en champ moyen permet de définir un nouveau modèle. Grâce à une analyse mathématique on peut obtenir les valeurs « théoriques » des plus petites barrières pour ce nouveau modèle. On peut également effectuer une simulation du modèle initial. Il est alors possible de comparer les valeurs théoriques obtenues sur le modèle approché et les valeurs obtenues pour le modèle initial à partir de la simulation. Dans ce contexte on ne trouve pas de concept particulier ni de théorie spécifique. L'essentiel des travaux consiste à conforter le modèle et le justifier par l'analyse mathématique d'un modèle proche et la simulation du modèle initial. Ces travaux ne font alors référence à aucune expérience particulière et il devient impossible de concevoir, dans ce cadre, des expériences qui pourraient invalider le modèle choisi. On se retrouve de nouveau en présence d'un modèle, sans qu'il soit possible pour autant de parler de modélisation. La situation est ici plus sophistiquée car on trouve plusieurs modèles apparemment proches et des simulations de ces modèles.

2 La multiréalisabilité

Chacun des différents domaines précédents constitue également un cadre contraignant. *L'expression, dans ce cadre, de la situation à modéliser, est d'un type prédéterminé.* Pour les modèles, par exemple, les expressions devraient se présenter sous la forme de systèmes axiomatiques, même si dans la pratique elles se font en termes d'expressions mathématiques, les axiomes en jeu, pas plus que la logique employée n'étant alors explicitées. De même, pour les simulations, le cadre est celui du programme informatique écrit dans un langage particulier, sur une machine déterminée.

On considère spontanément que la spécification du domaine dans lequel on travaille et de la situation à modéliser déterminent complètement l'aspect que la modélisation de ce problème prendra dans ce domaine. On considère par exemple que la modélisation d'une situation de diffusion dans le domaine du modèle *doit* prendre la forme d'une équation aux dérivées partielles. En fait il n'en est rien. Toujours pour cet exemple, on pourrait très bien choisir, dans le domaine du modèle, de prendre comme forme une équation aux différences, un automate booléen, voire les aspects mathématiques des techniques de gaz sur réseau.

Ainsi la spécification de la situation à modéliser et du domaine de la modélisation ne déterminent pas complètement un format précis. On peut

ainsi parler de **multiplicité des modèles**, dans le sens où chaque situation particulière dans le domaine du modèle peut prendre un format très différent.

Cette multiplicité des modèles permet de comprendre comment il est possible que des outils identiques soient utilisés pour n'importe quelle situation à modéliser. On peut ainsi employer systématiquement un unique outil mathématique pour transcrire dans le domaine des modèles toute situation à modéliser. Ceci permet de comprendre pourquoi certains scientifiques qui maîtrisent principalement un outil peuvent l'adapter à toutes les situations à étudier, c'est-à-dire traiter la physique, la biologie, l'éthologie, la psychologie, les sciences politiques, etc., avec leur outil. Il serait évidemment absurde, dans un tel contexte, d'attribuer une valeur ontologique à ces outils ou à ces modèles. Leur intérêt est méthodologique, il consiste à permettre d'exprimer, sous la forme d'un modèle, c'est-à-dire dans un des domaines de la modélisation, la situation que l'on cherche à étudier. Ainsi l'intérêt du modèle ou de l'outil est de participer à l'édifice qu'est la modélisation.

En fait la multiplicité n'est pas une spécificité propre au domaine des modèles, même si ses manifestations y sont les plus frappantes. Cette multiplicité se retrouve également dans les autres domaines de la modélisation. Nous n'évoquerons ici que la multiplicité des modèles, des simulations et des mesures sur ces simulations.

Nous avons pour l'instant insisté sur les différents domaines qui composent une modélisation, en les considérant relativement indépendamment les uns des autres. Nous avons présenté ces différents domaines, puis nous avons précisé une de leurs propriétés, à savoir leur multiplicité. Il faut noter de plus que ces différents domaines sont évidemment dépendants les uns des autres. Dans l'optique classique, les modèles se déduisent des théories, les simulations des modèles et les mesures sont faites à partir des modèles.

Pourtant, lorsque l'on se fixe un cadre théorique précis, on n'a pas encore déterminé les différents modèles susceptibles d'être exprimés dans ce cadre. La multiplicité des modèles nous amène même à constater que des modèles très différents dans leur format ou dans leur principe peuvent être construits en partant d'un même cadre théorique. On parle alors de **multiréalisabilité des théories**. La notion de multiréalisabilité est représentée sur la figure 3. On peut citer par exemple le cadre théorique proposé par Farmer, qui est assez précis, mais dont on peut « déduire » de nombreux modèles, affines sans doute, mais très différents par les types d'espaces mathématiques qu'ils mettent en œuvre et par les comportements qu'ils peuvent présenter.

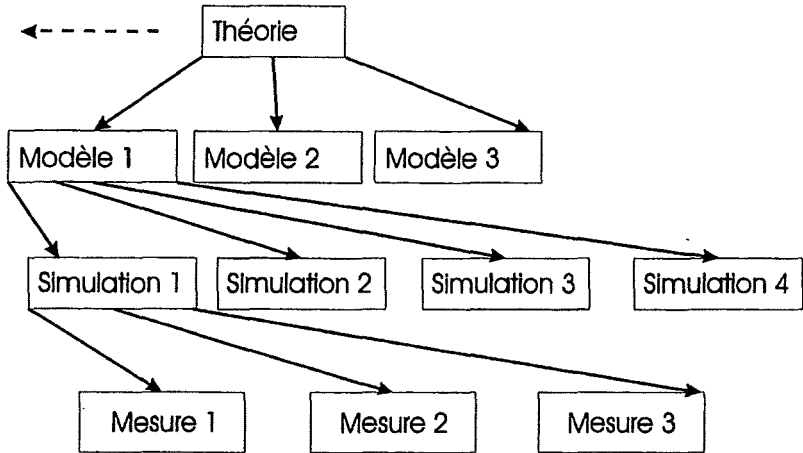


FIGURE 3 : La multiréalisation des théories et des modèles, ainsi que la multiplicité des mesures globales possibles sur une simulation. On notera également l'évolution des théories par déplacements de concepts.

On observe également une **multiréalisation des modèles**. Les scientifiques qui ont réalisé eux-mêmes des simulations de modèles qu'ils connaissaient bien savent que l'on peut concevoir, à partir d'un unique modèle, des simulations très différentes. Cette situation est frappante et même choquante. Ceci constitue une expérience déroutante. On imagine avoir bien spécifié son modèle et on constate que l'on est encore beaucoup trop loin d'un système qui pourrait être effectivement manipulé automatiquement par une machine. De nombreuses contraintes interviennent, liées globalement à la mise au format de la machine d'un modèle qui n'avait pas nécessairement été conçu pour cela. Le caractère discret des éléments manipulés intervient mais n'est pas le seul facteur de cette mise au format. Les difficultés peuvent provenir par exemple de la gestion des « erreurs » d'approximation. Par exemple certaines équations peuvent être discrétisées de différentes manières, certaines étant stables et d'autres pas. Des effets de bord peuvent intervenir, lorsque l'on souhaite manipuler un paramètre qui devrait être infiniment grand pour assurer la stabilité ou la convergence du modèle. Cette multiréalisation n'est pas une affaire de détail et n'est pas liée à une quelconque « incompétence » du scientifique qui réalise le passage du modèle à la simulation. Elle est généralement intrinsèque au système axiomatique considéré. De fait le passage du modèle à la simulation

nécessite une mise au format. Le modèle peut être parfaitement cohérent intrinsèquement et complet. Il n'empêche que c'est la mise au format de la simulation elle-même qui n'est quasiment jamais univoque. Ainsi la multiréalisabilité ne provient pas du modèle lui-même ou de la multiplicité de la simulation, mais de l'interaction entre le format du modèle et celui de la simulation.

Cette multiréalisabilité se retrouve également pour la simulation. Il est possible à partir d'une même simulation d'obtenir des mesures très différentes sur ces simulations. Ces mesures sont définies par le scientifique et ont rarement un caractère intrinsèque lié à la simulation.

En fait, comme nous l'avons vu plus en détail pour le passage des modèles aux simulations, c'est le changement de format⁹, à tous les niveaux, qui explique que l'on obtienne une multiréalisabilité des différents domaines, que ce soit la théorie, le modèle ou la simulation. Cette multiréalisabilité d'un domaine ne provient donc pas de lacunes qu'il présenterait et en particulier pas d'un « manque de spécification ». Si l'on est amené à préciser certains points, ce n'est pas obligatoirement parce qu'ils manquaient au domaine initial, mais plutôt parce qu'ils deviennent nécessaires dans le format du nouveau domaine. Leur absence n'entachait pas la cohérence interne du domaine initial. La nécessité de spécifier n'a donc rien à voir avec une quelconque carence du domaine précédent, ni une lacune de ceux qui ont pu proposer sa construction.

3 Le glissement de théorie

Il faut noter également que les théories elles-mêmes sont susceptibles de se modifier au cours du temps. On observe alors une forme de « dérive » des théories. Cette modification des concepts ne remet pas en cause les résultats obtenus antérieurement par la théorie. La communauté ne met simplement plus en avant les mêmes aspects de la théorie précédente. Les concepts conçus comme fondateurs et primitifs ne sont plus les mêmes. Ils peuvent porter le même nom et renvoyer désormais à de nouveaux contenus. Des changements conceptuels ont lieu qui modifient la façon dont les scientifiques appréhendent, voire pratiquent leur discipline. On parle souvent de « changement de paradigme » [Kuhn, 1962]. Ils peuvent intervenir pour tous les domaines. La multiplicité observée pour chaque domaine explique *a posteriori* que l'on puisse privilégier un format puis un autre. Le changement de paradigme pour un domaine consiste souvent à privilégier un nouveau

⁹ Dans le cas du passage du modèle à la simulation, ce qui est en jeu c'est le passage d'un format imposé par un modèle sans autre spécification à un nouveau format imposé cette fois par un couple composé d'un système axiomatique compatible avec une machine abstraite.

format en rejetant le format ancien. On peut alors reformuler, pour toutes les modélisations, dans toutes les disciplines, ces formats anciens dans ce nouveau format. Il s'agit là bien souvent d'un travail qui va au-delà de la simple transcription. En effet, comme nous l'avons souligné, le passage d'un format à un autre est une opération de traduction qui n'est généralement pas univoque et qui nécessite en conséquence un travail de compréhension approfondi de la situation à étudier. Ce changement de paradigme, c'est-à-dire ce changement de format au sein d'un même domaine, permet souvent d'établir des liens que l'on n'avait pas soupçonnés avec d'autres modélisations qui doivent rendre compte de situations apparemment très différentes. Il faut noter de plus que le concept de paradigme peut avoir des acceptions multiples et variées [Kuhn, 1962].

En fait ces schémas intègrent d'autres domaines, comme les protocoles expérimentaux, les expériences, etc., qu'il ne nous a pas semblé opportun de discuter dans l'optique du présent travail. Une analyse plus approfondie de ces schémas nécessiterait de préciser comment les débats et les différentes positions épistémologiques peuvent y être interprétés. On pourrait ainsi comparer entre eux les différents domaines et s'interroger sur le fait de savoir si les relations mises en évidence sont toutes de même nature.

Conclusion

Le lecteur se sentira sans doute frustré à la lecture de cet article sur la modélisation. Il pourra en particulier estimer, non sans raison, que l'ensemble est décrit trop rapidement, que les débats épistémologiques et leurs enjeux devraient être précisés, que les relations entre domaines ne sont pas aussi simples que nous voulons bien le dire, que l'on ne précise pas ici la forme de l'intervention du scientifique, que l'on ne voit pas bien ce que l'on peut attendre de la modélisation dans ces conditions, etc. En résumé on peut considérer cette présentation comme trop superficielle.

Il convient dans ces conditions de rappeler une des références actuelles pour la modélisation en sciences cognitives et dans plusieurs champs scientifiques proches des « Computer Sciences ». Il s'agit du schéma proposé par Rosen [Rosen, 1987] (figure 4). Ce schéma est la référence méthodologique sur laquelle s'appuient des spécialistes britanniques de la robotique située comme Hallam et Malcolm [Hallam et Malcolm, 1994], mais aussi John Casti [Casti, 1996], un théoricien du Santa Fe Institute. Selon ce schéma la modélisation est principalement un problème de mise en correspondance, d'encodage et de décodage, pour passer des systèmes physiques aux systèmes formels.

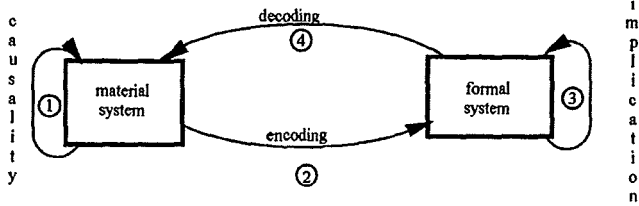


Figure 4a : La figure originale de Rosen. La modélisation selon Rosen [Rosen, 1987] est un problème d'encodage et de décodage des systèmes matériels vers les systèmes formels.

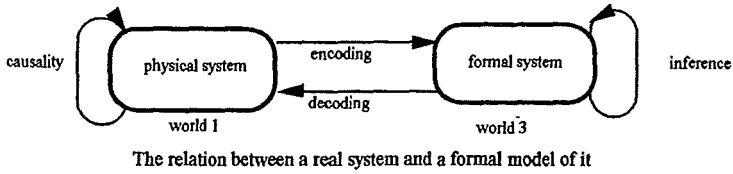


Figure 4b : La figure de Rosen reprise par Hallam et Malcolm [Hallam et Malcolm, 1994].

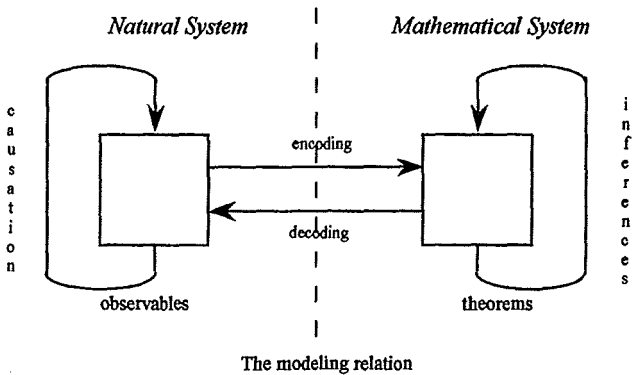


Figure 4c : La figure de Rosen reprise par Casti [Casti, 1996].

FIGURE 4 : La figure de Rosen sur la modélisation et ses différentes interprétations.

Nous avons voulu esquisser dans cet article un autre point de vue sur la modélisation. Dans ce cadre, la modélisation est une affaire bien plus sophistiquée que ne le laisse présager la figure proposée par Rosen.

Bibliographie

- [Bak, Tang et Wiesenfeld, 1988] Bak P., Tang C. et Wiesenfeld K. *Self-organized criticality*, Physical review A, vol. 38, n° 1, p. 364-374, 1988.
- [Bak, Flyvbjerg et Lautrup, 1992] Bak P., Flyvbjerg H. et Lautrup B. *Coevolution in a rugged fitness landscape*, Physical review A, vol. 46, n° 10, p. 6724-6730, 1992.
- [Blanché, 1955] Blanché R. *L'axiomatique*. Presses Universitaires de France, Paris, 1990. Première édition 1955.
- [Casti, 1996] Casti J. *The outer limits: in search of the « unknowable » in science*, Rapport technique 96-01-001, Santa Fe Institute, Santa Fe, 1996.
- [Diu *et al.*, 1989] Diu B., Guthmann C., Lederer D. et Roulet B. *Physique statistique*. Hermann, Paris, 1989. 1018 p.
- [Doolen *et al.*, 1990] Doolen G., Frisch U., Hasslacher B., Orszag S. et Wolfram S. *Lattice gas methods for partial differential equations*, Addison-Wesley, Redwood City California, 1990.
- [Eberwein et Saurel, 1995] Eberwein W.-D. et Saurel P., *The dynamics of collective behavior: modeling mass mobilization in the GDR*, 13-16 septembre 1995, Panel « Formal theory » at the Second pan-european conference in international relations, Fondation des sciences politiques, Paris, 1995.
- [Frisch, Hasslacher et Pomeau, 1986] Frisch U., Hasslacher B. et Pomeau Y. *Lattice-gas automata for the Navier-Stokes equation*, Physical review letters, vol. 56, n°14, p. 1505-1508, 1986.
- [Farmer, 1990] Farmer J. D. *A Rosetta stone for connectionism*, Physica D, vol. 42, p. 153-187, 1990.
- [Hallam et Malcolm, 1994] Hallam J. et Malcolm C., *Behaviour: perception, action and intelligence - the view from situated robotics*, Phil. Trans. R. Soc. Lond. A, p. 29-42, vol 349, 1994.

- [Hilbert, 1899] Hilbert D. *Les fondements de la géométrie*. Dunod, Paris, 1971. Traduction de *Grundlagen der Geometrie*. Première édition allemande Teubner, 1899.
- [Kuhn, 1962] Kuhn T. S. *La structure des révolutions scientifiques*. Flammarion, Paris, 1983. Traduit de *The structure of scientific revolutions*, 1970. Première édition The University of Chicago Press, Chicago, USA, 1962.
- [Popper, 1935] Popper K. *La logique de la découverte scientifique*. Payot, Paris, 1978. Traduit de *The logic of scientific discovery*. Première édition anglaise Hutchinson, 1959. Première édition allemande 1934.
- [Rosen, 1987] Rosen R. *On the scope of syntactics in mathematics and science: the machine metaphor*, Casti, J. et Karlqvist A., éditeurs, *Real brains, artificial minds*, p. 1-23, North-Holland, New York, 1987.
- [Sinaceur, 1994] Sinaceur H. *Jean Cavaillès : philosophie mathématique*, Presses Universitaires de France, Paris, 1994.
- [Volterra, 1931] Volterra V. *Leçons sur la théorie mathématique de la lutte pour la vie*, Gabay, Paris, 1994. Première édition Gauthier-Villars, Paris, 1931.

**Représenter quelques règles de droit
dans une machine
Au delà de Faust et de Frankenstein**

**Danièle Bourcier
Currap-idl**

Mes modèles : un essai faustien de clarification

S'interroger sur les modèles que l'on utilise dans ma discipline de base, les sciences du droit - voire se demander pourquoi on utilise la modélisation et non une démarche herméneutique par exemple- me paraît être une question sérieuse à laquelle je tenterai de répondre sérieusement. Cela signifie que j'en parlerai *autrement* que s'il me fallait décrire l'utilisation de ce modèle dans une conférence en Droit et intelligence artificielle par exemple.

Si on travaille dans un domaine interdisciplinaire et qu'en plus on veut de situer à un niveau méta-théorique¹ l'affaire devient compliquée. Si je dois reformuler *autrement* que ce que j'ai fait dans ce domaine, je donnerai quelques éléments supplémentaires de base mais sans renoncer à répondre à ce qui paraissait être le *starting point* de cette rencontre, à savoir non pas *quels modèles* utilisez-vous mais *que pensez-vous* de vos modèles?

C'est donc d'une expérience particulière qu'il sera question ici. Celle qui a consisté à chercher, dans un champ du savoir pratique qui se définit *a priori* comme modèle - à savoir le droit - comment créer d'autres *modèles* qui en rendraient compte. Prétention iconoclaste. Vouloir concevoir des modèles du modèle, surtout dans un domaine où tout est *déjà là*, plein de sens, peut apparaître parfaitement dérisoire. Cependant, cette quête d'autres modèles possibles, appliqués même fragmentairement à mon objet, n'a pas cessé depuis le moment où j'ai eu connaissance d'une certaine approche de la modélisation: celle que l'on situe autour de l'intelligence artificielle et des sciences cognitives.

¹Quine parlait de "montée sémantique" et Cavaillé opposait le "paradigme" qui neutralise les contenus superflus dans lesquels s'empêtraient les formes désirées et le "thème" qui mobilise les formes elles mêmes pour s'élever à un niveau où elles deviennent le signe d'opérations d'un degré supérieur (voir sur ces questions l'ouvrage de G.-G. Granger, *Formes, opérations, objets*, Mathésis, Vrin, 1994).

Je verrai donc d'abord brièvement comment cette démarche de modélisation s'est inscrite historiquement et en quoi le droit est un objet modélisable. Puis j'expliquerai ma propre démarche et en quoi elle a tenu compte des limites d'autres courants modélisateurs. J'essaierai ainsi de montrer en quoi la modélisation version cognitive a ouvert de nouvelles possibilités de représenter, de comprendre et surtout de valider nos hypothèses et nos croyances dans les sciences du droit.

1- Le droit est-il un objet modélisable? Etat des lieux

1.1- La recherche formelle en droit: *lex et calculus*

La recherche de formalismes et de modèles pour comprendre le droit n'est pas nouvelle. Les premiers essais philosophiques sont certainement liés à l'aristotélisme. Mais Aristote envisageait surtout l'éloquence judiciaire et peu le raisonnement normatif à partir des règles. Pour Platon, les lois plus que le jugement l'intéressent. Ainsi les lois, ne pouvant tout décrire, doivent être appliquées par des "fonctionnaires" valeureux. C'était déjà affirmer les limites de la règle comme élément unique de la décision et du comportement.

Il faut chercher du côté de Pierre de la Ramée pour comprendre ce que les commentateurs et les glossateurs en pleine période scolastique ont tenté de faire en liant représentation des concepts et structuration du texte (de l'écrit dans la page). Mais Leibniz est sans doute le premier qui ait voulu faire du droit un objet à modéliser. Deux démarches complémentaires ont orienté son travail: la recherche des *éléments fondamentaux* du droit - travail qui a été repris par Aurel David il y a quelques décennies- et l'analyse des *raisonnement juridique*.

Les travaux sur la cybernétique du droit ont suivi de près l'effervescence des travaux de Neumann après la première guerre mondiale. En France, Lucien Mehl a élaboré une théorie générale de l'administration qui est encore citée dans la plupart des ouvrages de méthodologie juridique. Parallèlement il a développé une réflexion sur la machine à juger (law machine) susceptible de faire des calculs simples à partir d'éléments fondamentaux dans le domaine fiscal par exemple. Mais c'est avec le début de l'informatique juridique que la question de la représentation de la connaissance juridique est devenue cruciale. Posons ainsi le débat: d'un côté une connaissance faite de textes (et d'éléments dans ce texte), de l'autre une machine capable de reconnaître des caractères et des séparateurs dans un corpus très vaste de textes. La notion d'implicite dans le langage juridique est devenue "visible".

Ce fut à ce moment que se développèrent les besoins de connaissances profondes (deep model) que l'intelligence artificielle appliquée au droit a mis à l'ordre du jour.

1.2- Le droit comme objet à modéliser

La plupart des juristes qui s'intéressent à la modélisation de leur domaine n'ont pas seulement pour but de remplacer l'homme par la machine. Le projet est moins ambitieux.

Car le droit n'est pas vraiment le modèle qu'il prétend être. Certes il décrit un état du monde. Et il propose des moyens pour *atteindre* ou *éviter* un autre état du monde (créer une société par actions, prévenir une atteinte à l'ordre public, définir un comportement délictueux...). En ce sens il est un *modèle* car il ne décrit jamais une réalité mais un de ses aspects possibles: la description est fragmentaire, incomplète et les choix indéterminés. D'un autre point de vue, il n'est pas non plus un *non-modèle* : il a ses chances de figurer parmi des objets modélisables du point de vue de la simulation. Il a à faire, en tant que texte en langue naturelle, avec d'autres champs du savoir comme la logique, la psychologie et les archipels des sciences cognitives. Mais ce ne fut ni langue ni le texte - heureusement- qu'il parut nécessaire de traiter. En ce qui me concerne, je me limitai à la description que l'on pouvait faire de ce texte, afin de le rendre *actif, utile., "performable"*. Ambition modeste mais surtout pratique car l'ordinateur ne connaît que les descriptions. De cet exercice, naissait une autre approche du droit comme pratique de la décision guidée par des règles. Mais de quelles règles? Si, lorsqu'on veut faire une machine à décider, une *law machine*, on s'aperçoit que les règles positives du droit ne suffisent pas, surgira une question incontournable: où sont les règles dont on ne parle jamais et *d'où*, pourtant, parle la langue du droit?

1.3- Un objet qui se propose comme modèle déjà là

La loi se présente comme un modèle à suivre : modèle de description du monde, modèle d'action d'une société dont elle est la transparente ossature. Modèle public puisque nul n'est censé l'ignorer: elle est écrite, codifiée, votée, promulguée, publiée...

Sa glorification, sa mise en majesté excèdent souvent la simple représentation matérielle de la règle, particulièrement dans notre pays, amoureux de la loi en tant qu'expression de la volonté générale. Elle est le symbole d'un temps éternel et d'un espace hiérarchisé : le Parlement est le temple des lois, la Justice a son Palais et ses mises en scène.

La loi a aussi ses modèles d'écriture: les codes, d'Hammourabi à Napoléon, symboles de pouvoir, qu'il s'agisse du conquérant ou du réformateur. Codifier des lois renforce la clarté du projet politique de recomposer un modèle. Hammourabi, le babylonien, le "roi des lois", avait inscrit dans la pierre les règles qui combinaient celles d'Akkad et de Sumer qu'il voulait unifier. Justinien voulait réécrire le droit romain, devenu illisible, à la gloire du Prince. A l'époque des lumières, le système juridique devait avoir un mécanisme aussi parfait qu'une *horloge*, l'horloge étant aussi le symbole d'un univers créé par un autre Grand Mécanicien. Enfin, pour Napoléon, le code civil avait constitué une de ses plus brillantes victoires sur la société féodale.

La loi est donc un modèle, et la représentation codifiée de ce modèle doit en renforcer la légitimité et la validité.

Cette image agit de deux façons différentes. Oeuvre d'art pour être contemplée et reflétée mais aussi modèle de rationalité pour être compris et suivi. Art et science.

1.4- Une particularité: ce modèle est son propre modèle

En outre, le droit est censé *contenir* son propre modèle opératoire. En montrant comment il se comporte, il désigne ce qu'il faut faire. Il est un corpus de règles, celles-ci pouvant être des règles substantielles (sur des objets, des concepts...) ou des règles procédurales (sur des actions, des processus...), des règles primaires ou des règles secondaires (qui indiquent comment gérer les règles primaires), des précédents voire des cas d'espèce. Il est censé donner les clés de son fonctionnement pour que nul ne l'ignore. Ce corps de règles représente *tout le droit*, et le *tout du droit*. Les règles d'interprétation sont aussi des règles. Le droit s'auto-administre. C'est ainsi que travaille la *fiction* d'un droit complet et que le juge commet un *déni de justice* s'il refuse de juger sur le fondement qu'il n'existe pas de disposition qui règle son espèce (article 4 du Code civil). Cette fiction renforce l'hypothèse que la méthode est congruente avec l'objet.

Cette approche est caractérisée dans les sciences comme celle du positivisme. Etre positiviste en droit signifie que l'on considère que le droit est immanent à son objet. *La norme est le fait*. Rien n'existe que ce qui est.

Les études de droit transposent cette approche dans ses modes d'apprentissage et d'analyse. Etudier le droit signifie étudier le droit tel qu'il est, la règle telle qu'elle est appliquée, applicable, interprétable. La glose fait même partie du droit. Les commentaires autorisés constituent une de ses sources.

Dans ces conditions, face à cette architecture monumentale, que reste-t-il au chercheur qui se pose la question du droit comme objet "géométral" ²?

1.5 - Les solutions alternatives : théorie, empirie, post modernisme etc.

Après avoir *appris consciencieusement*- objet et mode d'emploi - il a (ou elle) voulu *comprendre*. Se représenter le droit d'un autre point de vue. Pourquoi cette connaissance ne devait être acquise que dans le cadre des règles qui en autorisent l'interprétation? Pourquoi faut-il voir le droit du point de vue du droit fait et non du droit "in the making"? Pourquoi est-il si dangereux de vouloir ouvrir la porte de l'horloge?

Il, ce chercheur presque solitaire, n'était pas le premier à vouloir comprendre. De grands courants philosophiques le précédaient. Le droit a aussi ses formalistes. Y a-t-il une norme fondamentale qui peut "expliquer" la pyramide des règles?³ Ou bien, si le droit contient toutes ses règles d'interprétation, comment n'est-il pas menacé d'entropie? Le droit a aussi ses empiristes, ceux qui ne pensent pas que le droit suit un modèle. On les appelle les réalistes⁴. Pour eux, le droit ne représente rien, même pas lui-même. Il n'est que la somme de ses expériences. Il existera une autre tendance, ceux qui vont au delà du miroir, les forcenés de l'instrumentalisation de la règle: le droit n'est qu'un habillage de rapports de force quels qu'en soient la cause ou la raison. Le modèle du droit ne peut être en soi un objet d'étude: c'est un exemple particulier d'interactions dans un univers de règles et d'institutions.

Il y en fin ceux qui sont dans l'urgence de la décision, les plus nombreux: ceux qui légifèrent, administrent et jugent et se demandent pragmatiquement comment on peut utiliser efficacement un système si imparfait. Pour eux le modèle du droit est un épiphénomène.

Il ne s'agit là que des modèles à l'oeuvre en théorie du droit: il existe aussi un courant de logiciens du droit dont l'objectif a été de construire un

²Un objet peut être étudié de différents points de vue qui ne l'épuisent jamais . Granger parle aussi des formes qui ont pour but d'exténuier des contenus.

³Hans Kelsen, *Théorie pure du droit*, trad. Ch. Eisenman, Paris, Dalloz, 1962 : le droit est un ensemble hiérarchisé de normes, issues d'une norme *fondamentale*

⁴K.N. Llewellyn, *Jurisprudence: realism in Theory and Practice*, Chicago, The University of Chicago Press, 1962 : le droit c'est ce qu'en font les autorités, ce n'est pas un corpus des lois mais un ensemble de décisions. La décision à prendre n'est donc qu'une affaire d'intuition, de "feel of the law". L'énoncé de la règle n'est qu'une "coquille" qui ne prendra son sens que dans le contexte de l'énonciation. Certains, plus extrêmes, comme J. Frank dans l'article "Say it in music", fait application de l'idée de Gestalt à la décision judiciaire

système de logique des normes, fondés sur des termes primitifs, des règles de démonstration, des axiomes et des théorèmes⁵. Ce mouvement d'arithmétisation des inférences normatives n'a pas influencé la modélisation dont nous parlons ici: en revanche, l'ontologie est en train de devenir une branche active de l'intelligence artificielle appliquée au droit. Revenons au modèle cognitif.

1.6- Partir d'un modèle déceptif

Ce chercheur, prêt à accepter les règles du jeu voire tous les systèmes d'explication n'en finit pas d'être déçu. C'est l'image de l'horloge qui l'intéresse.

Il s'aperçoit que le droit n'est pas cette image glorieuse. D'abord les principes ne sont jamais uniques: ils s'équilibrent et se discutent. Les règles peuvent aider à rendre deux décisions antithétiques. Toute décision peut être contestée. Les lois, ces fameuses lois, sont devenues pléthoriques. La déclaration des droits de l'homme se déclinait en 500 mots; la législation européenne sur les oeufs de canard a besoin de 100 000 mots. Le droit est devenu jetable (dit le Conseil d'Etat). Les lois ne sont pas parfaites, ni simplement bien écrites. Elles sont toujours le résultat d'un compromis politique. Les codes? Une utopie. Nous sommes dans un monde quantique: la loi est devenue si précaire que l'on n'a même pas le temps de la "consolider". A peine celui de réécrire une loi dans un code, que la loi elle-même est modifiée. Le code des Impôts a été modifié 1650 fois en 1993. A peine un texte cite un autre texte que celui-ci est abrogé. Tout juste une compilation sommaire qui peut satisfaire l'avocat, non le citoyen.

Prenons les jugements et les arrêts. Certes, on a dit longtemps que l'arrêt était un syllogisme: cette image sublimée de la transparence et de la rationalité. A-t-on jamais cru qu'un juge raisonnât au moyen d'un syllogisme? Le sentiment de la justice comme le sens de la décision siège-t-il dans le cortex? Les prémisses étant ce qu'elle sont, la conclusion s'impose. Mais si les conclusions sont toujours obligatoires, en revanche les prémisses ne le sont pas.

⁵Le fondateur de ce courant est G.H. von Wright, *Norm and actions*, London, Routledge and Kegan Paul, 1963. Les travaux sur la logique déontique ont été développés entre autres par G. Kalinowski, *La logique des normes*, Paris, P.U.F., 1972.

1.7- A la recherche de la pièce perdue du puzzle?

Le chercheur qui aura regardé ce modèle sera un jour surpris de voir que cette rationalité là n'est pas ni descriptive ni explicative, elle est justificative... Et il manque une pièce au puzzle du modèle.

Il commencera, parce qu'il a toujours regardé par dessus son épaule, à prospecter dans le bureau d'à côté. Il se sera intéressé à l'arrivée de la modélisation dans les sciences humaines. Il se souvient que c'était dans les années 80 et qu'à ce moment on pensait au modèle comme une autre forme de l'analogie. Il s'agissait bien de modèle, comme celui du droit, mais ce modèle là, on le construisait et on le validait. Les systèmes à base de connaissances devenaient des instruments de mesure, des microscopes, des équerres, des règles pour nos opérations cognitives simples: l'intelligence artificielle permettait enfin de confronter la rectitude et la rationalité de l'objet-modèle à d'autres modèles et surtout à des modes de fonctionnement réels.

Sacrilège? Les premières recherches sur la modélisation du droit n'ont trouvé qu'une sollicitude courtoise de la part des milieux juridiques. Les modèles étaient acceptables s'ils étaient directement utilisables et extérieurs aux processus cognitifs individuels. Les outils ont alors été collés au droit de façon brutale. Certains étaient même efficaces parce que le texte lui même était une sorte d'algorithme (en matière fiscale ou douanière par exemple). Ou parce que les modèles d'interrogation de l'information étaient si frustrés (mots clés) que la réponse devenait aléatoire.

Mais comment peut-on penser que l'on peut retrouver un raisonnement ou un concept à l'aide d'un texte de loi? Et on retrouvait si peu de réponses à une question compliquée que ce piètre résultat ne pouvait que renforcer l'idée que le droit était si complexe que jamais une machine ne pourrait le représenter.

2 - Un modèle alternatif: comment décider avec des règles ?

Peu à peu, ce chercheur avait été lassé des interrogations booléennes et des check lists. Ces résultats hâtifs lui avaient montré que le droit c'était aussi une certaine image de la langue, une langue instrumentalisée, quasi-formalisée mais aussi une langue portant en elle-même une dynamique pragmatique. Et que cette langue était aussi une langue de pouvoir, gérant à la fois de la coercition et de l'autorité (Weber). Mais le texte de surface ne montrait pas directement ces liens intimes et complexes qui accompagnent le processus de compréhension et de lecture

des règles comme des réseaux contraignants de règles de pouvoir et des marges de discrétion..

Qu'il s'agisse des théories du droit ou de ces systèmes de recherche de l'information juridique, la question à résoudre restait la même: les énoncés étaient confrontés avec eux-mêmes sans qu'une quelconque interaction soit opérée avec ce qu'ils désignent et les acteurs qui doivent les appliquer ("les dire"?).

Pourtant cette langue donnait du pouvoir non seulement par ses procédures *externes* en contexte (la situation légitimant la performance normative) mais par son fonctionnement interne (le discours des conduites et des disciplines au coeur de la problématique foucauldienne). Il fallait rechercher ce que signifiait décider avec des règles au moyen du langage. Revisiter le langage comme procédure institutionnelle d'organisation de la décision et en représenter le fonctionnement dans une machine symbolique.

L'objectif de la recherche devenait celui-là: faire un système d'aide à la décision pour un décideur public dont l'action est encadrée par des règles, justifiée par des règles et qui doit produire une décision juridique face à des faits, à des événements, à une situation quelconque du monde réel.

2.1- Représenter la loi dans un modèle décisionnel

Un prototype de système à base de connaissances⁶ fut l'occasion de représenter et de valider une hypothèse sur le fonctionnement d'un système de règles en droit. Rappelons que ces travaux ont été élaborés sur dix ans tout au long des années 80. Certes on a depuis beaucoup critiqué, après une phase de discours visionnaires souvent irresponsables les modèles "systèmes experts" dans les laboratoires (pas assez formels?) alors que dans les expériences de terrain là où la connaissance était élaboré et les prototypes testés, l'assimilation et l'évaluation se sont déroulées plus calmement : les modèles sont-ils aussi facilement jetables?

Il existait un fort courant en science du droit pour ce type de représentation mais les modélisations portaient sur des dimensions cognitives différentes. Expliquons brièvement les enjeux des discussions dans le domaine des systèmes experts juridiques.

⁶BRUITLOG (en Prolog) puis BRUITEXPERT (dégradé en Turbo Pascal). Des articles, une thèse et un livre (D. Bourcier, *La décision artificielle Le droit la machine et l'humain*, PUF, Les voies du droit, 1995) ont été publiés sur cette expérience. Les références seront fournies et explicitées.

2.2 Les modèles concurrents

Traiter le droit dans un ordinateur de façon à ce que *certain*s raisonnements propres à la fonction juridique puissent être simulés: tel est le projet général des systèmes experts juridiques.

On limite en général l'expertise (juridique) à la transposition des énoncés de la norme en règles de production. Deux arguments militeraient en faveur d'une transposition aisée du droit en terme d'expertise, telle qu'elle est représentée dans les bases de connaissances. D'abord, le modèle du droit le plus courant est celui d'un ensemble de règles. Ensuite il existe une hypothèse communément admise par la psychologie cognitive selon laquelle l'expertise est fondée sur des règles.

Pourtant l'opération de réécriture d'un ensemble de textes normatifs dans la forme de règles de production en <SI, ALORS> semble soulever des difficultés et avoir fait l'objet de peu de généralisation méthodologique: l'examen de la littérature sur ce point montre qu'il existe autant de mode de "passages" que de systèmes experts, et que même, il est tout à fait possible, à partir d'un même ensemble de règles, d'écrire des règles de production différentes⁷. En outre au moment de l'utilisation (et de la validation par un quelconque test de Turing) il ne semble pas que la paraphrase du texte juridique suffise à caractériser l'expertise minimale - ces fameux standards - pour un raisonnement propre à la fonction juridique. Pour éviter que ces systèmes experts ne soient réduits au rôle de checklists⁸ ou de relecteurs naïfs du droit positif, il faut donc approfondir cette notion d'*expertise juridique*.

En réalité, l'opération de traduction d'un ensemble de connaissances textuelles en règles de production met en jeu les conditions de *passage* d'un domaine d'expertise vers un certain formalisme: les résultats sont fonction des contraintes du codage informatique mais aussi des limites de la "computabilité" (le mot manque en français) d'un domaine de connaissance que j'étendrai à la notion de *fonction de connaissance*. La recherche de ces contraintes éclaire d'un jour nouveau la double question de la normativité dans la langue -ce que Foucault appelait la "police discursive"- et de la discrétionarité dans la décision. Le paradigme objectiviste, appelé aussi positiviste (comme en théorie du droit) correspond à la première génération de modèles en Intelligence artificielle. Il est fondé sur la correspondance entre connaissance et réalité. Le terme "réalité", pour le droit, correspond au

⁷T. BENCH-CAPON (ed.), Knowledge based-systems and legal applications, Academic Press, 1991

⁸Ph. LEITH, The computerized lawyer, Springer-verlag, London, 1991, p. 198

texte de droit positif: l' article de loi , etc. On l'oppose en général au paradigme constructiviste "faisant participer le sujet de connaissance à la construction d'un monde qu'il (le sujet) amène à exister"⁹ .

Pour les positivistes donc, le droit s'exprime sous forme de textes normatifs. L'approche la plus *objective* consistera donc à "découper " la surface du texte suivant une certaine lecture. Mais cette lecture, très diverse comme on le remarquera dans l'examen des différents modes de formalisation, prend souvent appui sur le modèle computationnel pour acquérir les connaissances nécessaires et non sur le processus de construction de l'interprète. Les systèmes informatiques qui en résultent rendent compte de l'insuffisance de la plupart des approches syntaxiques "textes-règles".

2.2 La lecture des règles

Prenons la décision du maire en matière d'atteinte à la tranquillité publique. Donnons un exemple de situation à laquelle il devait faire face: une discothèque bruyante dans une station balnéaire.

Il faut d'abord lire les textes de lois (400 pages). Que disaient ces textes de lois sur la décision à prendre? Peu de choses: les lois ne sont pas des modèles de décision.

Regardons maintenant les recherches concernant cette étape de lecture pour en dégager soit une méthodologie soit éventuellement une solution.

D'abord une confusion est souvent faite entre l'étape de "reconnaissance des règles " et celle de leur formalisation dans l'ordinateur.

Pour certains cette phase est typiquement "humaine", signifiant en cela que toute tentative d'automatisation paraît vaine et toute méthode difficile à appliquer. C'est la position de SUSSKIND¹⁰: " L'étape de représentation des connaissances juridiques est un processus pour lequel nous ne connaissons aucun algorithme (puisque nous ne connaissons même pas d'algorithme pour celui du langage naturel)".

Au contraire, pour d'autres juristes, cette phase ne paraît pas poser de problèmes spécifiques: le texte juridique est considéré comme pouvant bénéficier d'une *transposition directe* dans la forme de règles.

⁹J. STEWART, Sciences cognitives : enjeux scientifiques et politiques, in N° spécial IA/Sciences cognitives: quels enjeux? in revue Terminal, n° 55, oct. nov. 1991, p.33-37

¹⁰R. SUSSKIND, op. cit. p. 46

Pour GREENLEAF (1992), l'approche qu'il développe dans PRIVACY WORKSTATION (une station de travail sur le thème de la vie privée) est "d'adhérer à la formulation de la législation aussi loin que possible", la règle n'étant alors que la "paraphrase" du texte d'origine¹¹.

Une autre façon de régler cette question de la transposition de la langue naturelle sous forme de règles est de choisir des domaines où "la loi est exprimée en termes d'objets concrets et d'acteurs qui ont des relations mathématiquement définies" (EDWARDS, 1992)¹². C'est le cas du domaine du droit fiscal ou du droit de successions par exemple.

Mais pour la plupart des concepteurs, la logique est considérée comme l'outil idéal pour analyser la structure d'un texte et dévoiler les ambiguïtés: les travaux de L.E. ALLEN (1963)¹³ sur l'explicitation de UNLESS (à moins que) dans les textes juridiques sont significatives à cet égard. Précisons que cette démarche a précédé l'écriture de base de règles pour les systèmes experts.

D'autres chercheurs liés à l'intelligence artificielle ont considéré par la suite qu'on pouvait faire plus que d'utiliser une liste de connecteurs (ET, OU, SAUF, SI, ALORS, SI ET SEULEMENT SI) pour traduire la législation en règles claires. STAMPER¹⁴ a été à l'origine d'un langage LEGOL (Legally oriented language) qui distinguait notamment dans son modèle deux types de règles d'application: les règles V qui peuvent être invoquées et des règles H qui peuvent être écartées (une partie de la règle spécifie pour chacune d'elles par qui et quand).

Ce langage tentait d'aller plus loin que la simple lecture logique de la loi en apportant un méta-niveau d'analyse. Mais il fut critiqué comme gérant des règles plus "administratives" que juridiques. Cette distinction a l'intérêt de faire intervenir le contexte décisionnel dans la syntaxe sous une forme codée.

Mais on peut aussi utiliser directement un langage de programmation logique (LISP ou PROLOG) pour à la fois interpréter et écrire des règles de droit. L'interpréteur PROLOG est utilisé alors pour traiter une série de faits et de règles, et prévoir les conséquences logiques d'une règle ou d'un fait. Prenons l'exemple bien connu : si l'on indique

¹¹G. GREENLEAF, The Privacy workstation, 2nd international Conference on substantive technology in legal education, Chicago, July 30-August 2, 1992

¹²Lilian EDWARDS, Building an Intestate Succession Adviser: compartmentalisation and creativity in decision support systems, in Journal of Law and Information science, vol.3 n° 1, Un. Tasmania, 1992

¹³L. E. ALLEN & M. CALDWELL, 1963, Modern logic and judicial decision making : a sketch of one view in H.W. BAADE ed., Jurimetrics, Basic books, New-york,

¹⁴R. STAMPER, LEGOL: modelling legal rules by computer in NIBLETT Computer science and law: an advanced course , Cambridge University press, 1980

que BOULE est un chat et qu'un chat a quatre pattes, on pourra savoir que Boule a quatre pattes. Si vous demandez combien BOULE a de pattes et que l'ordinateur répond 4, vous pouvez lui demander (à l'ordinateur) d'expliquer pourquoi. PROLOG est en effet à la fois un langage et un démonstrateur de théorème. Il présente l'avantage de pouvoir représenter des règles dans une forme logique et de les rendre directement exécutables par la machine : il peut donc être utilisé pour traiter les langues naturelles et même en "vérifier" la lecture.

L'équipe de logiciens d'Imperial college a utilisé PROLOG pour analyser le British Nationality Act. La formalisation du texte est fondé sur le mode essais-erreurs. L'énoncé de chaque article est décrit (sinon traduit) suivant la logique des prédicats du premier ordre sous forme de règle en <SI ALORS> et est remanié au fur et à mesure de la lecture pour rester cohérent avec le dernier article. Mais surtout cette équipe de logiciens a voulu prouver qu'un expert (juriste) n'était pas nécessaire pour formaliser un texte de loi. Ce que permet cette écriture c'est un traitement par ordinateur et l'ordinateur en "posant des questions à l'utilisateur" peut à son tour lui en poser (comme pour le chat Boule) et interagir avec l'utilisateur pour une "lecture de l'acte". Mais cette expérience a suscité beaucoup de critiques.

2.3 Une hypothèse opératoire: la représentation de la fonction décisionnelle dans le texte

Une nouvelle approche du sens des règles juridiques devenait nécessaire face aux nouvelles constatations suivantes:

- 1- Aucun des textes juridiques n'est formulé sous forme de règles.
Il faut repenser l'article, la disposition, l'énoncé sous forme de règles.
- 2- Il n'y a pas de rapport entre la particularité des cas et la généralité de la règle.
Il faut une connaissance du monde supérieure à la loi pour faire un système expert.
- 3- Les règles dans les textes de loi sont formulées à des moments différents pour des objectifs différents.
Il faut écrire les liens internes entre les textes de façon à ce qu'ils deviennent convergents et cohérents par rapport à des situations réelles.

4- Le droit gère des mécanismes d'interactions institutionnelles à l'aide de règles implicites, internes à la langue.

Il faut déconstruire les règles pour introduire la notion de pouvoir et d'autonomie dans la machine.

5- Les règles gèrent leur propre incertitude

Cette incertitude concerne l'évènement (tout n'est pas prévisible), le contexte (plusieurs acteurs peuvent avoir à agir), l'évolution du monde des faits et des règles (voir le modèle dont nous parlions plus haut).

Pour tenir compte de ce sens perdu entre loi et décision, il a fallu insérer les notions de *densité normative* et de *force d'un système de règles*. Ces notions permettent d'expliquer le degré d'autonomie ou de contrainte qu'il fallait représenter dans la machine.

Prenons l'exemple d'un ensemble de règles ayant pour objectif d'aider le maire à décider s'il s'agit d'une infraction pour "bruits de voisinage":

Si émergence > 3 dB sauf si bruit ambiant < 30 dB

ET SI période=nocturne

ET SI durée (t) = 30 s > t > 1 mn

ET SI instrument-mesure = conforme

ET SI élément-matériel = négligence-de-prendre-des-précautions appropriées

ET SI élément-intentionnel = oui

ALORS Infraction= bruit de voisinage

Nous avons ici un ensemble d'énoncés qui ne sont *marqués* ni par rapport à leurs référents (qu'est ce qu'un *élément intentionnel*?) ni par rapport aux acteurs (*qui peut* définir ce qu'est un *élément intentionnel*?).

Le système expert doit définir les possibilités d'interprétation entre l'homme et la machine, soit pendant la session soit par rapport au résultat de la machine, si l'on envisage l'interaction des systèmes humains et artificiels dans une même tâche globale.

Tout modèle représentant des règles doit donc inclure le degré d'autonomie d'un décideur devant appliquer une règle, ou un ensemble de règles, par rapport à ces quatre éléments:

l'émetteur (le législateur, l'administration...)

le destinataire (le maire, le préfet, le fonctionnaire...)

mais surtout :

le degré d'ouverture des conditions et des actions (concepts précis, détaillé, etc.)

la force de l'implication (obligation, possibilité, tolérance, ...)

Ces deux dernières composantes posent des questions fondamentales en langue (extension, intension, appartenance, etc...) mais aussi en pragmatique institutionnelle (justification, motivation, etc.).

On peut éclairer les deux dernières composantes par les énoncés suivants, indiquant les différents degrés de contrainte décisionnelle impliqués

Vous devez interdire (décision sans condition, pouvoir ou non pouvoir?)

S'il y a urgence, vous devez interdire (le décideur peut définir l'urgence)

Si c'est la nuit, alors vous pouvez interdire (le décideur peut définir ce qu'est la nuit et il a le choix d'interdire ou non)

S'il est 22 h, vous pouvez interdire (le décideur ne peut définir ce qu'est la nuit mais il a celui de choisir d'interdire ou non)

Si le bruit émergent est de 3 dB, vous ne pouvez pas interdire (le décideur n'a aucun pouvoir : ni celui de définir ce qu'est un bruit gênant, ni celui du choix de son action)

etc...

L'alternative a donc consisté à introduire la fonction de décideur dans la machine: pour cela, c'est la **rationalité** même du raisonnement qui est suivi et non la substance de la règle. Le système pourra proposer plusieurs stratégies:

- requalifier les faits
- utiliser des règles de justification (des exemples de jurisprudence orientés dans le sens de sa décision, des dossiers déjà traité dans ce sens...)
- vérifier les mesures techniques
- rendre cohérente les qualifications
- choisir en fonction des buts souhaités ou justifiables
- sélectionner une autre stratégie
- etc..

Les règles ne sont pas donc pas seulement des règles impératives du type SI <infraction> ALORS <sanction>; elles sont des règles d'assistance, de contrôle ou d'argumentation. La décision finale de qualification et le choix d'action ne pourront être élaborées qu'en interaction, en "donnant la main" au décideur. Dans ce modèle, la conception de la collaboration homme -machine prend tout son sens. En effet la condition d'étape de jugement- ce que KANT spécifiait comme conditions de la dérivation d'un ensemble de règles- peut être insérée dans la base de connaissances. Dans le contexte institutionnel municipal, le décideur final qui a un large pouvoir discrétionnaire peut à la fois rejeter ou modifier la décision, mais aussi garder la capacité de jugement.

Pour représenter les règles dans une "décision artificielle", il faut donc comprendre ces jeux de délégation d'interprétation, et les règles qui ferment, ou ouvrent, les choix discrétionnaires. La discrétion est un pouvoir sans règles *a priori* et sans justification *a posteriori*. Ainsi le droit définit une hiérarchie entre les machines : des machines consultantes, des machines critiques, des machines automates (ou autonomes), des machines assistantes, des machines discrétionnaires etc. Le droit crée des espace de décision qui ne peuvent être évalués que par rapport aux acteurs et au référent qui sont en jeu. Qu'il s'agisse de machines bureaucratiques ou de machines technologiques.

En proposant de raisonner sur le modèle décisionnel à partir de la fonction juridique, on adopte une attitude plus constructiviste qu'objectiviste par rapport au droit. Les objets de connaissance que sont les règles de droit n'existent pas en dehors du sujet qui interprète . Dans ces conditions, il faut savoir quel est le pouvoir du sujet par rapport à la loi. Le pouvoir se vérifie au degré d'ouverture c'est-à-dire aux modes d'interaction de l'homme et de la machine, mais aussi à la place du système dans l'institution. Suivant ce modèle , les systèmes experts pourront être considérés comme "assistants" avec un degré de détermination et de précision faible. En revanche , le système de simulation de la décision s'adressant à des agents sans pouvoir de décision (systèmes experts dits classiques) seront formalisés en suivant les différentes strates de structuration du droit , avec des règles heuristiques et des règles de sens commun contraignantes.

Il paraît donc nécessaire de relier règles de droit et modèle décisionnel et d'insérer le paramètre de la normativité ou de l'autonomie de l'utilisateur par rapport au système. L'ajout d'une règle peut rendre le système invalide non parce que la règle est fausse mais parce qu'elle ne s'applique pas au type de décision ou de raisonnement impliqué.

Pour conclure sur les machines à dire le droit: le modèle Frankenstein?

Représenter des systèmes de règles dans des machines a obligé à approfondir la notion de système juridique *ouvert* ou *fermé* (et le continuum).

Faut-il inclure le droit, qui d'une certaine façon est la science des systèmes de règles et des jugements normatifs, dans les sciences cognitives? Peu de connaissances échappent à la norme parce que tout modèle est aussi un ensemble de normes. Alors en ce sens peut être que Faust, comme le droit, sera sauvé par son idéal.

Il s'est élevé de nombreuses voix pour considérer que le droit ne pouvait être informatisé (WEIZENBAUM): une question d'éthique. Le droit a lui-même interdit l'informatisation de la décision quand celles ci "traitent" des données liées à la vie privée. Mais il s'agit d'un autre débat sur ce que nous faisons de nos modèles quand il deviennent des machines dans nos organisations.

Mais "techniquement", le droit peut-il être simulé dans une machine? Cette question a-t-elle un sens? Rappelons cette pensée du métaphysicien et juriste Aurel David, auteur de l'ouvrage "*Le cybernétique et l'humain*":

"La décision du capitaine sera confiée au pilote c'est à dire tôt ou tard aux machines. Il en résultera un intense démoralisation de la zone prétendue humaine et une extension quasi illimitée de la pensée asservie. Car les véritables buts sont très peu nombreux et que tout le reste n'est que moyen, soustrait aux désirs et à l'attention du capitaine".

Qu'est-ce qui sépare nos modèles de nos machines, Faust de Frankenstein?

Le Diagramme, le Modèle et le Signe.

Bernard MORAND

GREYC UPRESA CNRS 6072, Université et ISMRA de Caen
Email : morand@iut3.unicaen.fr

Résumé

Cet article développe l'idée selon laquelle le concept de modèle peut être explicité au sein d'une théorie du signe, comme une information créée dans un but d'information. Nous nous appuyons sur des propositions de C.S. Peirce qui permettent de comprendre la polysémie du terme lui-même : image créée par imitation, figure à caractère manipulable ou expression d'une structure formalisée. Nous vérifions cette approche sémiotique en étudiant le cas des diagrammes Entité-Association utilisés par les analystes de systèmes d'information ainsi que leur traduction en schéma relationnel de base de données. Nous mettons en évidence que l'algorithme de traduction interrompt la chaîne des significations entre ces deux modèles, ce qui pose alors un problème pour l'interprétation externe de l'application informatique.

1. Les traits sémantiques du terme « modèle ».

Le mot *modèle* de la langue naturelle est très largement polysémique : dans un travail précédent⁽¹⁾ nous notions trois traits sémantiques principaux relevés dans les définitions données par les dictionnaires. Le premier trait prend *modèle* au sens d'un objet que l'on peut reproduire par copie ou imitation, par exemple le moule de fonderie, un modèle d'écriture, le modèle du peintre ou le patron de la couturière. Dans cette acception du terme, la modélisation consiste en une opération de référenciation dans laquelle l'objet à produire doit respecter les caractères de son modèle, de la même manière que l'on moule une occurrence sur son type. On trouve donc aussi l'idée que l'opération peut être réitérée : le modèle permet de reproduire plusieurs objets à son image. On peut encore dégager deux sens dérivés de ce premier trait. Soit le modèle constitue le premier exemplaire d'une série à partir duquel on va engendrer des occurrences et on parle alors de *prototype*. Soit il s'agit à l'inverse d'une occurrence en principe unique dont la fonction est d'imiter les caractères du modèle, on parlera de *maquette* ou *modèle réduit*.

Le second trait sémantique est celui qui fait de *modèle* un objet idéal représentatif d'une catégorie, d'un ordre ou d'une qualité : un maître à penser sert de modèle pour son élève du fait de ses qualités intellectuelles, un enfant modèle est un exemple de sagesse et le poisson est un modèle des animaux aquatiques. Le caractère opératoire, manipulable ou encore expérimentable du *modèle* au sens du premier trait est réduit au profit du critère de qualité en elle-même. On peut cependant trouver des recouvrements entre les deux sens : le modèle du peintre ou du sculpteur est à la fois copiable et, en principe, objet idéal en tant qu'il présente des qualités particulières (la beauté, la force, ...etc.). Le troisième trait sémantique est beaucoup plus récent, contemporain de la logique moderne qui fait notamment suite aux travaux de Tarski, et utilisé dans l'expression de *théorie des modèles*. Un *modèle* est un couple (Domaine, Interprétation) qui rend vraie une formule logique. Il consiste en la donnée d'un ensemble d'objets d'un domaine, la donnée d'une interprétation de relations entre ces objets pour lesquelles il s'agit d'établir que ces deux données sont le cas de la formule logique. Cette acception de *modèle* sert en particulier à définir la validité d'une déduction logique :

« la proposition X **suit logiquement** les propositions de la classe K si et seulement si tout modèle de la classe K est aussi modèle de la proposition X » (Tarski, cité par Gochet⁽²⁾, p. 330).

Un modèle est donc une structure formellement définie par un domaine et une interprétation que l'on utilise comme moyen pour valider ou invalider une inférence. Dans un sens élargi, un *modèle* est une structure formalisée d'objets qui pourra être utilisée comme support d'activités de raisonnement : spécification de relations entre objets logiques, tests d'hypothèses, comparaison de structures, ...etc.

Nous notions dans le même travail l'étrange similarité entre ces trois acceptions du mot avec les stades du développement de l'intelligence chez l'enfant tels que Piaget les propose. Il distingue en effet le stade sensori-moteur dans lequel l'enfant produit des imitations réflexes des phénomènes de son environnement, puis le stade des images mentales dans lequel se développe une activité de représentation et enfin le stade des structurations opératoires à caractère de raisonnement logique. Nous notions une autre similarité, tout aussi étrange, avec les différentes manières dont on peut envisager une méthode scientifique. En effet, on oppose souvent la méthode expérimentale à une méthode par représentations abstraites qui mathématise ses objets ou encore à une méthode herméneutique qui vise davantage une explication au moyen de structures articulées de lois. Parlant du point de vue de la conception des systèmes d'information, nous tirions la conclusion que :

« tout l'art de la conception pourrait bien résider dans la **conjonction** de l'**imitation**, de la **figuration** à caractère expérimental et du **raisonnement** à caractère opératoire ».

Nous reprenons ici ce pronostic pour expliciter la nature de la conjonction présumée. On peut observer que les différences spécifiques représentées par les traits sémantiques du mot *modèle* se font sur la base d'un genre commun : un modèle est, dans tous les cas, quelque chose qui tient lieu d'autre chose, en bref une représentation. Il se trouve par ailleurs que « quelque chose qui est mis pour une autre chose », *aliquid stat pro aliquo* selon les médiévaux, constitue une définition élémentaire du **signe**. Peirce étend cette définition de la façon suivante :

« un signe, ou representamen, est quelque chose qui tient lieu pour quelqu'un de quelque chose à quelque égard ou en quelque qualité » (C.P. 2.228¹).

La section 2 développe quelques points caractéristiques de la théorie sémiotique peircéenne en vue de justifier notre argument selon lequel les modèles sont nécessairement des signes. La section 3 expose le cas des diagrammes utilisés en conception de systèmes d'information pour montrer qu'ils présentent des propriétés caractéristiques du signe. La section 4 identifie quelques problèmes posés par le processus de modélisation des systèmes informatiques en proposant des réponses possibles, du point de vue d'une théorie sémiotique. Nous montrons comment les termes de *diagramme*, *modèle* et *information*, peuvent s'articuler au sein d'un même champ sémantique, celui de *signe*.

2. Quelques caractéristiques du signe selon C.S. Peirce.

Mathématicien, philosophe, sémioticien et logicien, cet auteur américain décédé en 1914 est resté largement ignoré des milieux académiques si l'on excepte les coups de chapeau de convenance au « précurseur » pour ses graphes existentiels et ses quantificateurs logiques, le raisonnement abductif ou sa définition du signe. Le **pragmatisme** dont il fut le fondateur explique probablement ce désintérêt pour un auteur dont les travaux n'ont souvent été considérés qu'au travers du miroir très déformant transmis par Morris et Carnap : le pragmatisme y est réduit à l'effet pratique du signe en contexte. Or, les travaux de Peirce constituent une théorie en filiation directe d'une logique du signe et une telle réduction n'a pu s'opérer qu'en gommant l'ensemble de son système philosophique. Si l'on en juge par les publications^(4, 5),

¹Les références aux *Collected Papers* de Peirce, sans autre indication de source, renvoient à la traduction donnée par G. Deledalle⁽³⁾.

6, 7), il semble heureusement en passe d'être réhabilité en France. Nous reprenons sa définition du signe dans la figure 1 :

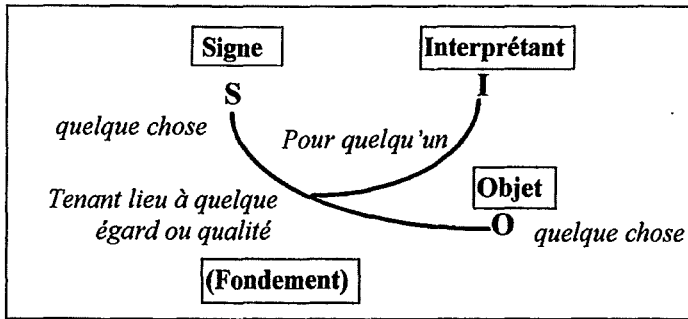


Figure 1 : Le concept de Signe selon Peirce.

Chacun des trois pôles de cette relation triadique est lui-même un signe, c'est-à-dire constitutif de la même relation. L'objet O (que l'on nomme quelquefois dans d'autres traditions *réfèrent* ou *signifié* ou encore *contenu*) d'un signe S (que l'on nomme quelquefois *signifiant*) est lui-même signe. Ceci veut dire que toute pensée a propos d'un monde extérieur ne peut s'opérer que par signes, que donc les « objets » sur lesquels une pensée pourrait s'exercer ne sont pas donnés et en tout cas, pas par l'intuition. Ils n'ont pas de « réalité » autre que celle de signes (perçus). Il en résulte que le seul postulat métaphysique dont nous avons besoin est celui du caractère observable de ces phénomènes ; il est donc possible d'en faire des études de manière expérimentale. Ici réside selon nous le principe véritablement pragmatiste de Peirce souvent cité et trop souvent mal interprété :

« Considérer quels sont les effets pratiques que nous pensons pouvoir être produits par l'objet de notre conception. La conception complète de tous ces effets est la conception complète de l'objet »⁽⁴⁾, p. 297.

Cette maxime fondatrice nous dispense du postulat positiviste selon lequel il existerait un ordre « naturel » du réel, indépendant de sujets pensants dont la tâche serait de découvrir des lois auxquelles cet ordre obéirait. L'interprétant I est lui aussi signe, et non une « personne » en dépit de l'expression « pour quelqu'un », méprise explicitement vulgarisée par Morris et ses successeurs. Il s'agit d'un sujet - de - pensée dont le caractère de signe nous dispense d'un autre postulat, de caractère psychologique celui-ci, relatif à des modèles mentaux génétiquement innés ou structurellement acquis. Ici

réside la seconde face du pragmatisme de Peirce qui n'interdit, **par ailleurs**, ni une science de la psychologie ni une science de la neurobiologie par exemple : ce sont simplement d'autres sujets. En définitive, selon Peirce toute pensée est en signes et les signes sont instruments de pensée. Une conséquence de ceci est le caractère quelque peu dérangent de la sémosis : un phénomène sans limites discernables puisqu'un signe **S** appelle son interprétant **I** qui est lui-même objet **O** d'un autre signe **S** appelant quelque autre interprétant et ceci à l'infini.

Une autre caractéristique essentielle est la propriété d'**insécabilité** de la relation triadique **O, I, S**. En d'autres termes, il est impossible de décomposer le signe en, d'une part, une relation de référence et, d'autre part, une relation d'interprétation. Cette propriété renvoie à deux aspects reliés de l'oeuvre de Peirce : sa théorie des catégories au plan de la philosophie et sa théorie des relations au plan de la logique. Pour ce qui concerne la première, l'auteur distingue la catégorie de la « priméité », celle de la chose en soi, des qualités et des possibles qui précèdent toute existence posée par un sujet (exemple : un sentiment), la catégorie de la « secondéité » qui est celle des relations dyadiques où s'exercent des forces entre un sujet et un objet (exemple le rapport de cause à effet) et, pour finir la catégorie de la « tiercéité » qui est de l'ordre de la généralité (exemple : la convention, la loi ou l'habitude). Peirce pose une relation d'ordre sur ces catégories : un troisième comporte un second, lequel comporte nécessairement un premier mais l'inverse est absurde. En tant qu'il est un troisième, le signe contient de manière immédiate son premier **O** par la présentation qu'il en effectue. Il contient de deux manières son second, **I** : de manière immédiate en tant qu'un signe n'est signe que pour un autre signe **I**, et de manière dynamique en tant qu'il exerce sur **I** un effet. Pour ce qui concerne la logique des relations, la découverte de Peirce est précisément la suivante : il y a dans toute relation à trois éléments quelque chose de plus que dans la décomposition analytique de chacun de ces éléments pris deux à deux. Il donne l'exemple suivant : si **A** donne **B** à **C** (relation triadique), cela ne veut pas dire que **A** jette **B** et qu'ensuite **C** ramasse **B** ; on ne peut décomposer le triplet sans perte. Par contre, toute relation de degré supérieur serait décomposable, sans perte, en relations triadiques selon R. Burch⁽⁸⁾.

Deux conséquences de cette théorie à la fois logique et sémiotique sont à opposer à certaines vues courantes sur le signe. Tout d'abord l'objet **O** ne transmet rien au signe **S** : le signifiant ne permet en lui-même ni une connaissance ni même une identification du signifié, il en tient seulement lieu pour quelque interprétant et il ne peut pas non plus « véhiculer » quelque connaissance que ce soit. Ceci s'oppose donc aux conceptions usuelles

de la référence largement utilisées par exemple en Représentation des Connaissances. Ensuite, le procès d'interprétation est fondamentalement de nature dialogique et diachronique en ce que I est contraint de viser le même objet O que S lui-même, procès dans lequel I ne fait que compléter sa propre connaissance antérieure de O. La cognition prend donc la forme d'un processus mû par l'échange et non celle d'une structure munie de propriétés particulières et en elles-mêmes objectives.

Nous terminons cette rapide présentation par le tableau des divisions du signe que Peirce considérait comme une validation expérimentale de sa théorie. Il commence (C.P. 2.243) par établir une division trichotomique du signe selon qu'il est considéré en lui-même (priméité), considéré dans son rapport à son objet (secondéité), ou dans la façon dont son interprétant le représente (tiercéité). Chacune de ces trois divisions est ensuite reprise une à une (C.P. 2.244 à C.P. 2.252), à nouveau sous la détermination des trois catégories philosophiques pour établir ce que l'on pourrait appeler une grammaire du signe (figure 2) :

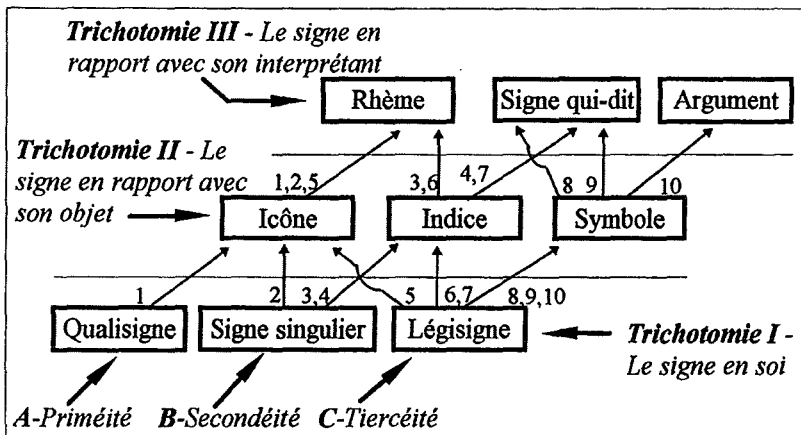


Figure 2 : Tableau des divisions du signe

Chaque élément simple du tableau peut être défini de la façon suivante :

- Qualisigne** : Apparence dénuée d'identité, instantanée et arbitraire.
- Signe singulier** : Événement individuel existant une seule fois.
- Légisigne** : Loi générale signifiant au moyen de répliques singulières.
- Icône** : Substituée à l'objet en vertu de caractères propres analogues.
- Indice** : Substitué à l'objet en vertu du fait qu'il en est réellement affecté.

Symbole : Substitué à l'objet en vertu d'une loi, habitude ou disposition naturelle.

Rhème : Interprété comme une simple possibilité qualitative.

Signe qui-dit : Interprété comme le signe d'un existant actuel.

Argument : Interprété en tant que loi faisant signe (abduction, induction, déduction).

Le tableau de regroupement est ainsi construit :

Premièrement². Une ligne de la figure désigne une partie constitutive du signe et trois lignes sont nécessaires pour qu'il soit complet : un caractère propre, un objet, un interprétant. Une ligne montre la division en éléments simples, une relation de différence donc, sur un trait catégorique donné. Par exemple un signe peut être du fait de ses caractères propres, qualisigne ou signe singulier ou légisigne et rien d'autre. Par rapport à l'objet qu'il dénote un signe peut être icône, indice ou symbole. Un signe peut être interprété (connoté) comme un rhème ou un signe qui-dit ou un argument.

Deuxièmement. Une diagonale est utilisée pour figurer l'effet de la relation d'ordre (notée *A, B, C* dans la figure) qui s'impose aux capacités de connexion d'un élément simple avec d'autres. Un élément simple ne peut se connecter qu'avec un autre élément simple situé sur la même diagonale ou à sa gauche. Par exemple un qualisigne ne peut être qu'icône pour son objet et ne peut être interprété que comme rhème. Un signe singulier peut être soit icône soit indice de son objet mais pas symbole. Un indice peut être interprété comme rhème ou signe qui-dit mais pas comme argument (sauf, précisément, par un héros de roman comme Sherlock Holmes).

Troisièmement (C.P. 2.254 à 2.264). Il est possible de bâtir au moyen des deux règles précédentes et compte tenu de la relation d'ordre sur les constituants (notée *I, II, III* dans la figure 2) une liste complète et suffisante de 10 classes de signes (notées par les chiffres dans la même figure). Les caractères mis entre parenthèses sont déductibles par construction³ :

1. Le qualisigne (iconique, rhématique) : ex. *la crainte de l'accident, avant de monter dans sa voiture.*
2. Le signe singulier iconique (rhématique) : ex. *la figure 2 elle-même, comme diagramme individuel.*
3. Le signe singulier indiciaire (rhématique) : ex. *l'empreinte du pied de Vendredi sur la plage.*
4. Le signe singulier (indiciaire) qui-dit : ex. *la girouette.*

² Cette présentation est faite délibérément pour indiquer que la structure des divisions du signe obéirait elle-même aux règles qu'elle prescrit (ce qui reste cependant à démontrer).

³ La structure de la relation entre ces dix classes a été formalisée par Marty ⁽⁹⁾ en treillis.

5. Le légisigne iconique (rhématique) : ex. *le rectangle et l'arc, comme forme générale de la figure 2.*
6. Le légisigne indiciaire rhématique : ex. *un pronom démonstratif.*
7. Le légisigne indiciaire qui-dit : ex. *le mot Allô prononcé au téléphone*
8. Le (légisigne) symbolique rhématique : ex. *un nom commun de la langue.*
9. Le (légisigne) symbolique qui-dit : ex. *une proposition.*
10. L'argument (légisigne symbolique) : ex. *la règle de passage des prémisses à la conclusion du syllogisme.*

A titre d'exemple, la girouette est un signe singulier en elle-même. Par rapport à son objet, elle indique la direction du vent et elle en est directement affectée par un lien de causalité (secondéité) ; c'est donc un indice. Son interprétant doit admettre que si elle tourne, alors c'est que le vent a tourné (un second à nouveau : l'interprétant est lui aussi affecté réellement par la rotation de la girouette) et c'est un signe qui-dit. Au total, la girouette est un signe singulier indiciaire qui dit. Ce tableau des divisions du signe auquel Peirce est parvenu vers la fin de sa vie a fait l'objet de nombreuses hésitations de sa part. Il est douteux qu'il l'ait considéré comme définitif et, conformément à sa méthode, il demande à être vérifié par l'expérience. Notre propre expérience en conception de systèmes d'information nous conduit à penser que les modèles qui y sont construits ne sont autres que des signes. Si cette hypothèse est valide, alors la théorie sémiotique peircienne doit fournir un schéma explicatif de ces modèles. Nous examinons maintenant ce point.

3. Les « diagrammes » de conception des systèmes d'information.

Soit le cas archétypique suivant, que l'on peut considérer comme un état des choses du système d'information ou, déjà, un énoncé de l'Univers du Discours :

(A) : *Les commandes de produits faites par les clients.*

Dans une perspective de mémorisation de cet état de choses par un système d'information automatisé, l'ingénieur analyste utilisera le très populaire modèle Entité-Association pour tracer le schéma « semi - formel », figure 3.

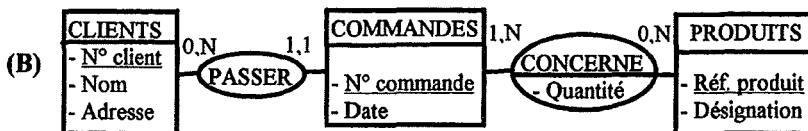


Figure 3 : Un modèle Entité-Association

Les rectangles sont des entités, les ovales des associations qui relient les entités entre elles. Entités et associations sont nommées. Les entités sont décrites par des propriétés dont au moins une est un identifiant (attribut souligné dont la valeur est réputée différente pour chacune des occurrences de l'entité). Les associations n'ont pas d'identifiants propres et peuvent avoir ou non des propriétés. Les caractères portés sur les liens sont des cardinalités qui statuent la façon dont une occurrence d'entité peut être reliée à une occurrence d'association : par exemple 0,N porté sur le lien Client - Passer signifie qu'un exemplaire de client peut passer 0 ou plusieurs commandes. En fait les explications que nous venons de donner peuvent être présentées dans un méta-modèle (simplifié) construit selon les mêmes conventions (figure 4).

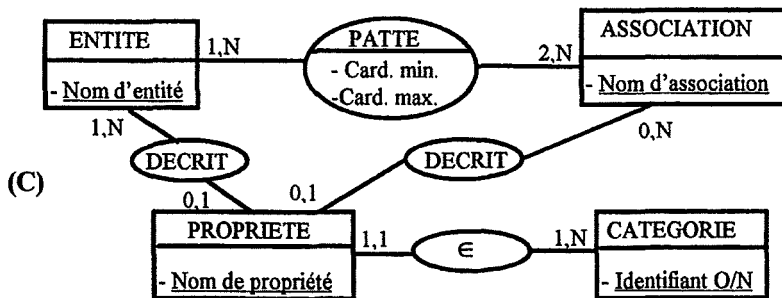


Figure 4 : Méta-modèle Entité-Association

En quoi de tels schémas non seulement rentrent-ils dans la catégorie générale du signe, mais de plus participent-ils de telle ou telle classe de signes selon la proposition de Peirce ? L'expression A est elle-même déjà un signe S dont l'objet O peut être un énoncé linguistique écrit sur un cahier des charges par exemple, une suite de sons émis par la secrétaire comptable, une « information » déjà instituée sur un formulaire de bon de commande, etc. En tant que signe, il peut produire un effet sur une très grande variété d'interprétants I et en particulier sur le signe matérialisé par le schéma B. En tant qu'il est signe S, le schéma B lui-même a pour objet l'énoncé A et un interprétant potentiel : le schéma C. On pourrait redéployer la même analyse à propos du schéma C lui-même mais nous arrêtons là : il nous suffit de comprendre que vient de se mettre en place une chaîne de significations actionnée par des signes. Envisageons maintenant la question suivante : de quelle(s) sorte(s) de signes s'agit-il ? En prenant B à témoin, on peut dire qu'il s'agit en lui-même d'un signe singulier, occurrence ou réplique de C

dans laquelle ont été instanciées des entités, des associations et des propriétés dans un agencement prescrit par C (une syntaxe). Par rapport à son objet A, B est un indice dans la mesure où il manifeste l'existence de A aux yeux de quelque interprétant⁴. Dans son rapport à son interprétant, ici C, B est un signe qui dit, un signe qui informe au sens propre C en procédant au découpage des entités et des associations, à l'affectation des propriétés relativement à l'énoncé A. En résumé, B est un signe singulier indiciaire qui dit. Au regard de la théorie sémiotique il est donc de même nature que la girouette vue précédemment, ce qui peut sembler quelque peu étrange.

De fait, l'analyse que nous venons de donner pourrait être considérée comme une lecture possible, un commentaire parmi d'autres qui viendrait de la sémiotique et qui ne serait d'aucun effet, ni sur les pratiques de modélisation ni sur la formalisation théorique du domaine de la conception des systèmes d'information. Nous argumentons en faveur une position exactement inverse : ce dont il s'agit ici, comme dans le cas de la girouette, c'est de la compréhension d'un processus de production et d'interprétation de signes ou sémosis. C'est dans la manière d'agir de B au sein d'un système d'information que celui-ci est signe. Voyons maintenant cela de plus près. En vertu du principe d'emboîtement des catégories, un signe singulier indiciaire qui dit contient aussi : *i*) en tant que signe singulier : des qualisignes, *ii*) en tant qu'indice : des icônes, *iii*) en tant que signe qui dit : des rhèmes. Tout d'abord, B possède en commun avec A des qualités (icônes) qui sont les mots de la langue naturelle *Clients, Commandes, Produits* (nous avons souvent observé que la performance communicative d'un modèle entité - association tient au moins autant dans la pertinence du choix des signes linguistiques qu'il contient que dans sa bonne formation syntaxique). Ensuite, en tant qu'indice, B dirige l'attention de tout interprétant potentiel sur A en mettant ces contenus en relation de la même manière que A : l'exemple B montre à l'évidence, le redoublement dans le modèle sous la forme entité - association - entité(s) de la forme sujet - verbe - complément(s) de l'énoncé A. Enfin, B laisse à tout interprétant potentiel un espace de possibilités dans le découpage et l'agencement des parties représentées (aspect rhématique que Peirce nomme le caractère « vague » du signe). A titre d'exemple ici, on peut imaginer de faire de *Adresse* une entité de plein droit, liée par une as-

⁴ C'est précisément cette caractéristique qui nous semble imprimer sa marque à l'informatique de gestion : Par l'enregistrement, notamment dans des bases de données, de ces « manifestations d'existence », elle a pour fonction d'attester que les événements relatés ce sont bien produits. C'est bien pourquoi, dans le cas présent un problème de preuve (au sens de Peirce : un problème d'interprétant) se pose : de quel droit la commande informatisée peut-elle tenir lieu de l'événement commande ?

sociation vers *Client* (ce serait même recommandé si les clients peuvent posséder plusieurs adresses différentes). On peut imaginer à l'inverse de faire du *Client* une pure propriété de la *Commande* (et c'est effectivement le cas dans le commerce de détail), etc.

Bref, **B** est un *modèle manipulable par un interprétant*. En employant le verbe « imaginer », nous venons de considérer que l'interprétant peut être un auteur ou un lecteur du modèle. Nous venons donc d'opérer un déplacement en direction de la production du modèle, la sémosis, quittant la position d'une analytique du signe. Bien sûr, toutes les manipulations ne sont pas autorisées parce que certaines mises en relation peuvent être incompatibles avec **A** comme dans le cas de la figure 5 :

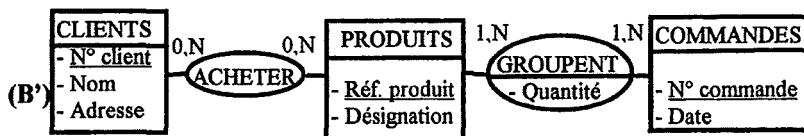


Figure 5 : Un diagramme incorrect

On perd ici l'information qui permettrait de savoir quelles sont les commandes d'un client (on sait seulement quels produits les clients ont acheté et, indépendamment, de quels produits se composent les commandes). Ces représentations iconiques et indiciaires, manipulables sous le contrôle d'une norme, peuvent être nommées des **DIAGRAMMES**. Le schéma **B** apparaît donc comme un membre d'une grande famille qui va de la formule du physicien comme $U=RI$ aux expressions de l'algèbre comme $2x + 3y = 0$ en passant par les représentations des molécules de la chimie comme $H-O-H$ ⁽¹⁰⁾. Leur caractéristique commune est d'être des objets d'expériences sur lesquels des raisonnements peuvent se développer et grâce auxquels, comme le dit joliment Peirce, il est possible de voir la pensée en actes. En particulier, la disposition topologique des parties du diagramme y joue un rôle important : par exemple *Quantité* dans le schéma **B**, est placé dans une position qui indique sa double dépendance, à la fois du *Produit* et de la *Commande*. Observons enfin que ces propriétés ne sont pas réunies dans le cas de **C**, légisigne symbolique dont le caractère iconique se réduit à réutiliser les qualités que sont le rectangle et l'ovale aux fins de désignation des concepts d'*entité* et d'*association* sans que les parties qu'il met en relation soient manipulables. Pour plus de clarté, nous réserverons à ce seul dernier cas le terme de **MODÈLE**, décision purement terminologique et arbitraire qui permet cependant de souligner la propriété de reproductibilité des légisignes. Si

l'on observe de l'extérieur la chaîne de significations A, B, C on constate qu'il s'y opère une mise en forme progressive des contenus, doublée d'une réduction de l'incertitude des déterminations pour un interprétant. Nous voyons aujourd'hui dans ces caractéristiques des traits spécifiques de la notion d'INFORMATION. Nous examinons maintenant comment l'informatique articule cette chaîne de significations.

4. Du « diagramme » au « modèle » dans la conception des systèmes informatiques.

A propos de l'énoncé A, on peut établir un autre schéma qui s'écrira de la façon suivante :

(D)

R1 (NUMCLI, NOMCLI, ADCLI)
 R2 (NUMCOM, DATCOM, #NUMCLI)
 R3 (NUMPRO, DESPRO)
 R4 (#NUMCOM, #NUMPRO, QUANTITE)

Figure 6 : Un modèle relationnel

Il s'agit de l'expression d'une structure des données telle qu'elle sera implémentée dans un ordinateur⁵. Elle respecte les canons établis par la théorie du modèle relationnel. Une relation est un sous-ensemble du produit cartésien de ses domaines. Un domaine est défini comme un ensemble de valeurs. Dans le schéma de la figure 6, R1 note une relation qui comporte les domaines NUMCLI, NOMCLI et ADCLI. Le modèle définit également les notions de clé primaire et de clé étrangère (respectivement notées par un souligné et un # dans la figure 6). Ce qui nous intéresse ici est le caractère de signe de D. Son objet O est un certain nombre de marques sur un disque magnétique. Un interprétant potentiel I est le modèle relationnel lui-même que nous avons brièvement présenté de façon littérale ci-dessus. Mais D est un signe singulier qui dit son objet O, une base de données particulière, laquelle est un fait existant.

En conséquence et contrairement à une illusion référentielle très répandue, D n'est pas indice de A. En effet, dans R1 l'aspect iconique de

⁵ A vrai dire et pour des raisons d'optimisation, l'implémentation effective pourra être quelque peu différente. On distingue alors le modèle logique D du modèle physique ; on dit aussi que le premier est un modèle *externe* masquant l'implémentation effective aux yeux de l'utilisateur alors que le second est un modèle *interne*.

NUMCLI, NOMCLI, ADLI n'évoque respectivement le N° *client*, son *Nom* et son *Adresse* que par le **simulacre** d'une référence alphabétique à ces objets externes. NUMCLI, NOMCLI, ADCLI pourraient tout aussi bien s'écrire x, y et z sans que la signification de **D** en soit pour autant altérée. Ce sera à la charge du programme informatique, opérant sur les données de la base, de garantir que les caractères *Pierre Dupont* qui s'inscrivent sur un écran constituent bien (aux yeux d'un lecteur) le nom d'un client. Ce point somme toute banal a cependant fait couler beaucoup d'encre dans les années 80 parce que l'on a réussi à « transcoder » **B** en **D**. Nous sommes donc, d'un point de vue sémiotique, dans la situation montrée par la figure 7. On y voit clairement que la chaîne des significations y est interrompue par une relation dyadique, le programme de transcodage.

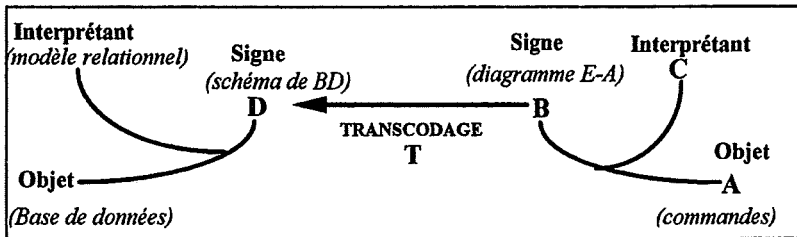


Figure 7 Transcodage du diagramme E-A en schéma de BD

On peut se demander si le programme de transcodage n'est pas un signe authentique (au sens de la triade peircéenne) de **B** pour **D**. Il est clair que l'objet de **T**, ce dont **T** tient lieu, ne peut être le diagramme Entité - Association. En fait, **T** est seulement une fonction de réécriture syntaxique de **B** en **D** gouvernée par des règles applicables à toute transformation d'un schéma E-A en schéma relationnel. Nous n'exposerons pas ici ces règles simples (au nombre de trois) mais nous pensons avoir montré que la chaîne des significations se trouve ainsi interrompue. Pourtant, si l'on observe ce complexe de l'extérieur⁶, le procédé fonctionne puisque nous pourrions utiliser la base de données informatique comme signe de **A** et, en conséquence, lui attribuer du sens. C'est qu'il y a ici selon nous, et d'un point de vue sémiotique, **simulacre** à nouveau. Les gens du métier n'en sont d'ailleurs pas dupes, ce qui est le propre d'un simulacre, puisqu'ils ont convenu de qualifier **T** de « moulinette » : ils savent que le procédé peut dépanner provisoirement.

⁶ C'est-à-dire du point de vue d'une représentation de l'informatique par opposition à une représentation en informatique, distinction que nous empruntons à E. Bataud⁽¹¹⁾.

sus, indépendant du précédent, sans qu'il soit possible de capitaliser les connaissances acquises. C'est effectivement ce qui se passe dans la suite de cette histoire, que nous ne pouvons raconter ici, avec l'arrivée de la conception orientée objets.

Si l'on souhaite rétablir une véritable chaîne de significations entre les commandes et la base de données, il faut un signe *S* dont l'objet soit *B* et dont *D* puisse devenir un interprétant (figure 9) :

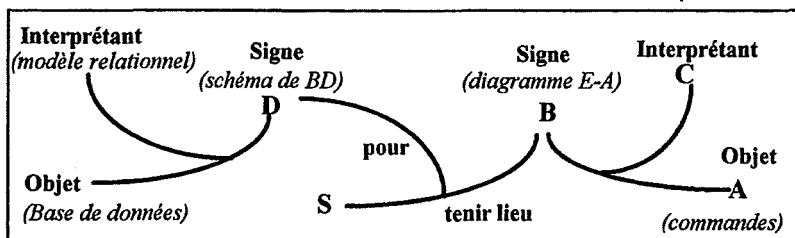


Figure 9 : Une problématique de recherche

Il nous semble que ce signe *S* ne peut être que l'expression sémiotique d'un *utilisateur* s'informant des commandes au moyen du schéma de la BD. En fait c'est ce qui manque à *T*, le programme de transcodage. Une autre direction de travail complémentaire consisterait à élaborer un légisigne *L* dont *C* (le méta-modèle E-A) serait l'objet tout en étant capable de s'adresser au modèle relationnel (en tant qu'interprétant d'une BD relationnelle en général). En termes sémiotiques, *L* devrait **manifester** (ou **garantir** ou seulement **illustrer**) que *B* est à *A* ce que *D* est à la base de données, soit très exactement une métaphore. Bien que ne se réclamant pas d'une approche sémiotique, c'est la voie qui est suivie dans les problématiques de recherche développées dans le domaine de l'intégration de schéma : le problème y est, par exemple, de trouver une représentation fédérant des bases de données hétérogènes.

5. Conclusion.

Une approche sémiotique des phénomènes informationnels nous conduit donc à mettre en cause un paradigme essentiel, bien que le plus souvent implicite, de la pensée informatique. Un système informatique ne peut être un système autonome de traitement de l'information. Le « sens » de son activité ne peut être attribué que par un interprétant extérieur, qu'il s'agisse d'un humain ou non. Il nous faut donc substituer à l'idée de traitement au-

tomatique celle d'interaction. Un corollaire de cette position est qu'il ne s'agit pas pour les concepteurs de systèmes informatiques d'introduire du sens dans les machines, ce qui supposerait à la fois une certaine matérialité et une certaine objectivité du sens alors qu'il s'agit d'un effet sur un interprétant. Il s'agit plutôt de créer des systèmes hommes - machines, socialement insérés, dont l'effectivité engendre des significations, c'est-à-dire de l'information. Dans cette perspective, le rôle du scientifique est de formuler des propositions théoriques sur la nature de l'information : une analytique du signe, sur le processus selon lequel elle se produit : une génétique du signe, et sur les moyens par lesquels le complexe homme - ordinateur peut effectuer un tel processus : le fonctionnement du signe ⁽¹²⁾. Un domaine de prédilection en est à l'évidence les principes de formation des modèles, c'est-à-dire la conception de l'information. Nous avons montré que les traits descriptifs du terme de modèle : image copiée, figure manipulable et structure formalisée, loin de s'opposer, constituent une totalité dont une théorie du signe permet l'entendement. Celle de Peirce propose une organisation en trois catégories imbriquées qui font du modèle formel un troisième qui contient analytiquement un second et un premier :

« A tout symbole doivent être attachés de façon organique ses index de réaction et ses icônes de qualité » (C.P. 5.119).

Cette théorie fournit également un principe de compréhension en diachronie : comment émergent progressivement les formes à partir des observables. D'autres théories du signe lui sont certainement opposables mais le rapport qu'entretient le créateur avec son modèle ne peut être à notre avis que de l'ordre du sémiotique, c'est-à-dire de la production d'informations. Ce ne peut donc être un accident de l'histoire que l'expression A (*les commandes de produits par les clients*), soit devenue la forme moderne, complexifiée et enrichie, de cet antique signe observable sur nos foires aux bestiaux d'antan : le serrement de mains d'un acheteur et d'un vendeur accompagné de la formule rituelle « *Topez là* ».

6. Références.

- ¹Morand, B. (1995) Statut épistémologique des modèles dans la conception des systèmes d'information. *Ingénierie des Systèmes d'Information*, Hermès, 3, 5, 665-700.
- ²Gochet P. & Gribomont P. (1990), *Logique, méthodes pour l'informatique fondamentale*, Hermès.
- ³Deledalle G. (1978). *Charles S. Peirce. Ecrits sur le signe*. Rassemblés, traduits et commentés par G. Deledalle, Seuil.

- ⁴Chenu J. (1984). *Peirce. Textes anticartésiens*. Présentation et traduction de J. Chenu, Aubier.
- ⁵Peirce C.S.(1898). *Le raisonnement et la logique des choses (Les conférences de Cambridge)*, Introduction K.L. Ketner & H. Putnam, Traduction Ch. Chauviré, P. Thibaud & Cl. Tiercelin, Editions du Cerf .
- ⁶Claudine Tiercelin Cl. (1993), *La pensée - signe. Etudes sur C.S. Peirce*, Editions J. Chambon.
- ⁷Chauviré Ch. (1995). *Peirce et la signification, introduction à la logique du vague*, PUF.
- ⁸Burch R. (1991). *A Peircean Reduction Thesis and the Foundations of Topological Logic*, Texas Technical University Press.
- ⁹Marty R. (1990). *L'algèbre des signes*, John Benjamins, Amsterdam - Philadelphia.
- ¹⁰Marty R. La sémiotique selon Robert Marty, <http://www.univ-perp.fr/web/RCH/LSH/SEMIOTICS/MARTY/>
- ¹¹Batard E. Représentations en informatique et représentation de l'informatique, *Actes du colloque interdisciplinaire 1997 Les modèles de représentation : quelles alternatives?*, Association F. Gonseth, Institut de la méthode, Bienne (Suisse).
- ¹²Morand B. Les sens de la signification. Pour une théorie à priori du signe. *Les Cahiers du GREYC*, Université de Caen, 1998.

Conscience avec sciences

Hugues Bersini
IRIDIA - ULB
50, av. Franklin Roosevelt
1050 Bruxelles

1. Introduction : Conscience : La déferlante philosophique

Ce petit écrit n'a pour seule ambition que de participer au débat dans lequel s'engluent actuellement un nombre impressionnant d'acteurs philosophiques et scientifiques. Ce débat a pour objet l'explication scientifique de la conscience en tant qu'expérience sensible et subjective et le statut de cette explication dans le champ des sciences cognitives. Un premier étonnement est légitime, il servira d'introduction. Pourquoi les étales de librairies aujourd'hui s'affaissent sous le poids de livres au titre aussi ambitieux (pour ne citer que les plus lus) que « *Consciousness explained* », « *Towards a science of consciousness* », « *The Astonishing Hypothesis : The scientific search of the soul* », « *... The Seat of the Soul* », « *A cognitive theory of consciousness* », « *The self and its brain* », « *... A Biological Theory of Consciousness* », « *The Emperor's New Mind* », « *The Conscious Mind : In Search of a Fundamental Theory* » ainsi que sous le poids du prestige de leurs auteurs : prix nobel : Crick, Edelman, Eccles, mathématicien d'élite : Penrose, philosophes de référence : Dennet, Searle, Churchland, etc...

Parce qu'elle est à la base de tout, il n'y pas lieu de s'étonner qu'en tant que sujet d'étude pour elle-même, la conscience soit l'objet de multiples convoitises scientifiques. Elle semble même se présenter pour certain comme le Graal scientifique, le dernier rempart au bulldozer scientifique, à la théorie du tout. Ce déficit apparaît encore plus vital pour la communauté scientifique et philosophique américaine, à l'arrivisme et aux envies de triomphalisme clairement affichée, et qui présente, comme trait distinctif, de ne pas s'encombrer plus qu'il ne faut des traditions ou même des errements intellectuels du passé.

Historiquement, Descartes semble avoir été le premier qui, par son adhésion à un dualisme ontologique et son expression du doute fondamental, capable de se défaire de tout sauf d'elle-même, à poser clairement le problème soulevé par toute tentative d'explication de la conscience. Dès le départ, il caractérise le phénomène conscient comme intrinsèquement subjectif, n'existant que pour un « moi » et s'évanouissant comme par magie en dehors de ce « moi ». Depuis Descartes, fondamentalement, on ne peut affirmer que le problème ait trouvé dans la philosophie contemporaine une expression différente. Sans doute, peut-on se limiter à constater que l'informatique a permis de reposer le problème d'une façon nouvelle « plus branchée », un vieux problème dans des habits neufs.

En effet, l'autre raison de cette déferlante philosophique est à rechercher dans le projet de l'intelligence artificielle au sens fort. Celui-ci a sorti de sa léthargie ce problème ancestral en aiguisant les sens des philosophes américains, tels Searle, Dreyfus, Dennet, Nagel, et beaucoup d'autres, se sentant quelque peu déstabilisé par l'ambition démesurée de ce projet ? Ceux-ci, comme nombre de leurs disciples, considèrent maintenant comme un exercice obligé de disserter sur ce thème et de trouver là, pour pas mal d'entre eux, l'occasion, lassante pour leur lecteur, de faire étalage, sinon de leurs incohérences du moins de leur inconstance. Par exemple, et pour se limiter aux plus célèbres, quand Searle cherche fondamentalement, par la « chambre chinoise », à séparer la sémantique de la syntaxe, il adhère à une forme de dualisme. Pourtant, dans ses mêmes écrits, il ramène toute manifestation psychologique au seul matériel cérébral (tout en alternant sans arrêt entre la position matérialiste et fonctionnaliste) pour finir dans son livre le plus récent par affirmer que « *The world contains subjectivity as a rock bottom element.... In the sense in which I am here using the term, subjective refers to an ontological category, ...* ».

Il n'en va pas beaucoup mieux pour Dennet qui, en prétendant « expliquer la conscience » sur la couverture, laisse le lecteur Gros-Jean comme devant à la fermeture du livre. On imagine difficilement qu'un être sensé comme Dennet puisse adhérer à la thèse du couple Churchland, défendant avec acharnement l'élimination de la conscience comme une manifestation psychologique, et néanmoins il s'obstine à traiter le problème par une opiniâtreté lassante à démontrer qu'il n'est qu'illusoire.

Nagel est à l'origine d'un de ces coups de pub de génie, un slogan du tonnerre, qui marque une communauté philosophique a tout jamais : le

« *what's like to be a bat* » par lequel il explique que l'explication du phénomène conscient ne peut éviter de se heurter à l'ultime barrière du sensible, au point singulier du cogito, surface imperméable à toute explicitation. Dès lors, quel étonnement par la suite quand il tient à se présenter comme un matérialiste de la première guerre (bien qu'il éprouve tout de même quelques difficultés à caser la conscience dans les neurones). Popper est assez détonnant également quand, tout à la fois, il prétend, avec autorité, être capable de séparer le bon grain scientifique de l'ivraie et co-signe un livre où la conscience s'explique tout bonnement par un phénomène quantique. Je garde Mc Ginn pour la fin, imputant à nos limitations cognitives l'impossibilité de comprendre comment l'eau de nos neurones se transforme dans le vin de la conscience (tout comme les fourmis ne peuvent comprendre la mécanique quantique, il dit "*an understanding of consciousness is beyond the reach of the human mind, cognitively we are not equipped to understand it in the way we understand other phenomena we experience in the physical world*") et qui, plus loin, dans le même livre, attribue le mystère de la conscience à son caractère intrinsèquement non spatial. J'aurais pu aussi vous parler des non philosophes et de leur acharnement suicidaire, malgré le talent qu'on leur reconnaît dans leur chasse gardée, à s'exposer poitrine nue, et ceci quand ils prétendent également élucider le mystère de la conscience, aux balles de leurs adversaires et envieux.

Fermons là cette parenthèse pour nous repencher sur le projet de l'IA forte. De fait, et sans conteste, les machines pensantes existent aujourd'hui. L'IA les met au point. Elles sont capables de reproduire nombre de nos mécanismes cognitifs: inférences sérielles et parallèles, analogie, apprentissage, mémorisation, automatisation, raisonnement par défaut, taxonomie cognitive, ... Mais ces mêmes machines sont-elles conscientes? Allons jusqu'à imaginer la plus incroyable d'entre elles, capable de réussir le test de Turing et même de nous confondre sur son apparence physique en passant quelques jours chez le sculpteur américain De Andrea (dans la littérature philosophique on traite généralement ces créatures de « zombie »), cette copie conforme sera-elle consciente? De Hal à Terminator, que de films et d'ouvrages de sciences fictions ont trouvé là matière première de scénario et clef du ressort dramatique (Terminator est-il ému? Hal a-t-il des pulsions de mort?).

Je pense que ces deux phénomènes, l'existence aujourd'hui d'une science par trop gloutonne ainsi que les succès de l'IA faible (mais qui ne

demande qu'à se fortifier), sont à la base de ce déferlement philosophique dont l'origine principalement américaine est certainement à rattacher à l'esprit pionnier ainsi qu'au manque de recul philosophique ou simplement littéraire.

2. La raison du mystère : pourquoi la conscience apparaît-elle comme énigmatique ? ?

L'accès à la conscience semble être possible par deux portes: l'une, qui bien que privée, fondamentalement subjective et introspective, soit à la portée de tout un chacun, l'autre qui bien qu'objective et publique, soit réservée à une communauté scientifique plus triée, composée de neurophysiologues, neuropsychologues et autre neuroscientifiques. Lapalissade d'envergure, la conscience est la première et la plus simple des choses dont nous pouvons être conscients. Mais comble de malchance, seuls nous pouvons en être conscients. Seuls nous, dans ce que nous avons de plus privés, savons ce qu'il en est d'être nous. D'autre part, tant les neurologues que les psychologues aspirent, souvent avec succès, à une caractérisation cognitive ou neurologique des phénomènes qui chez nous s'accompagnent de conscience. Je sais que l'on skie à Rochebrune, l'ordinateur pourrait le savoir également car le symbole « Rochebrune » peut-être mécaniquement connecté au symbole « ski ». Mais cet ordinateur le sait-il de la même façon que je le sais, ressent-il ce que je ressens lorsque je pense à cela ? Je sais que ma vision de la neige est blanche, l'ordinateur pourrait également caractériser la neige comme blanche, mais s'agit-il du même blanc, etc.

C'est de l'existence de ces deux portes d'accès à la conscience, une porte franchissable à la première personne, l'autre à la troisième, mais sans connexion possible, que naît l'énigme de la conscience. L'une donne accès à une expérience privée et sensible, c'est l'expression du « MIND ». L'autre offre la possibilité d'une caractérisation scientifique, en dévoilant même les structures et mécanismes causaux qui chez nous n'ont d'existence que « conscients », c'est l'expression du « BODY ». Par ailleurs, ce « BODY » peut n'avoir d'existence que matérielle ou fonctionnelle, selon que l'on pense que la fonction seule suffit aux phénomènes cognitifs ou que la matière cérébrale possède quelque chose d'irréductible à une simple expression fonctionnelle. Un des apports majeurs de l'IA aura sans doute été la possibilité offerte aux psychologues, par l'entremise de l'ordinateur,

de faire «des sciences cognitives», en n'identifiant comme objet d'investigation pour ces sciences que les objets pouvant se doter d'une existence informatique, pouvant se caser soit dans la mémoire soit dans un programme.

Ce gouffre infranchissable, de nature épistémique ou ontologique, entre ces deux portes d'accès est, à lire et à relire les ouvrages philosophiques, la clef de l'énigme, ce qui distincte, pour Chalmers (le dernier bateleur en date et dont le livre, de très bonne facture par ailleurs, fait couler beaucoup d'encre) le « easy du hard problem de la conscience». Lisons le : « *Even when we have explained the performance of all the cognitive and behavioural fonctions in the vicinity of experience - perceptual discrimination, categorization, internal acces, verbal report - there may still remain a further unanswered question: Why is the performance of these fonctions accompanied by experience ?.... This further question is the key question in the problem of consciousness. Why doesn't all this information-processing go on in the dark, free on any feel ?* ».

L'ordinateur raisonne presque comme moi (je vous en prie), parfois moins bien, parfois mieux, mais la question de son expérience privée semble dénuée de sens. Que les activations de neurones puissent entrer en synchronie (comme expliquer par Crick), c'est bien possible. De fait, l'ensemble de la population mondiale, les yeux rivés sur un électroencéphalogramme géant, pourrait s'accorder là-dessus. Mais que cette même synchronie, soit (toujours selon Crick) le support matériel de l'activité consciente, voilà une couleuvre qui semble beaucoup plus difficile à avaler car, jamais au grand jamais, l'ensemble de la population mondiale ne pourra ressentir le phénomène conscient dans ce même électroencéphalogramme et où que ce soit d'ailleurs.

3 Les premiers résultats de l'enquête

De nombreux philosophes se sont évidemment penchés sur cette énigme abyssale, cet infranchissable gouffre, et ont proposé d'adopter quelques premières positions timides ou fortes, qui se catégorisent comme suit :

- les éliminativistes durs : la conscience n'existe pas comme entité ontologique, c'est l'équivalent du « phlogiston », un accident historique, un mauvais souvenir pour les sciences de demain. Ne souriez pas, ils existent, j'en ai même rencontré des heureux. On en croise beaucoup au pays des églises.

- les éliminativistes mous : la conscience existe et elle est parfaitement identique aux phénomènes cérébraux qui la génèrent. Ne me demandez pas comment, c'est comme cela. 50Hz d'oscillations synchrone d'un assemblage particulier de neurones, et la perception bleue du bleu (appelée également « qualia » par les philosophes), il ne s'agit là que d'un même et unique phénomène. Ce sont les chantres de la position dite de la « supervenience » de l'esprit sur le corps.

- les « onfaitcommesiste » : dans nos modélisations cognitives des phénomènes qui, chez l'humain, ne se manifestent que consciemment, tout est réuni pour que le modèle soit également « véhicule de conscience ». Rien n'est absent, tout est là. Si la conscience ne suit pas, c'est qu'elle n'apporte rien de plus. Cela reste de toute façon le mieux que l'on puisse faire, le plus que l'on puisse modéliser. Il ne faut pas amalgamer les « onfaitcommesiste » aux éliminativistes car, ils ne nient pas le phénomène, ni ne l'assimilent entièrement au mental, mais ils prétendent qu'en dehors de leurs modèles cognitifs rien n'existe comme phénomène cognitif. C'est évidemment une position très voisine de l'épiphénoménalisme.

- les épiphénoménalistes : la conscience est au cerveau ce que la Suisse est aux guerres mondiales : stratégiquement neutre. La conscience n'a aucun rôle à jouer. Elle accompagne passivement les phénomènes cognitifs comme le sifflement du train (exemple originel de Thomas Huxley) accompagne passivement le déplacement du train. Elle est causalement inactive. La sensation du bleu n'apporte rien à l'activité neuronale responsable de la perception du bleu. La conscience chapeaute, mais tout se passe en dessous du chapeau. Le monde pourrait être peuplé de Zombies (dans « Blade Runner », il l'était), êtres dénués de toute sensation, mais pour le reste en tout point nos semblables (et allez donc les débusquer!).

- les capitulateurs : ceux pour qui le mystère reste entier. Ceux qui ne veulent pas ou plus chercher à comprendre où le bleu du bleu peut bien se nicher dans l'activité cérébrale, ceux pour qui le problème est sans solution, et ce peut-être car nous sommes, en tant qu'humain, limité par ce même état d'humain à ne comprendre que certaines choses.
- les nouveaux dualistes : ceux-ci proposent deux types d'ontologie, l'une matérielle, l'autre mentale. La conscience devenant une nouvelle donnée fondamentale de la science comme les forces qui régissent la nature. La chute de la pomme ou la perception du bleu sont deux phénomènes qui demandent pour leur compréhension un nouvel ingrédient. La conscience se manifeste aux deux niveaux ontologiques et se soumet à des modes d'analyse différents suivant le niveau d'étude. La plupart des écrits reliant le mystère de la conscience à la mécanique quantique se situent dans cette catégorie. C'est également la position adoptée par Chalmers, le dernier auteur star de la déferlante, qui réclame pour la conscience un nouveau type d'explication, non réductionniste, dont une première propriété pourrait être la cohérence structurelle.

Souvent, des philosophes font une jolie bouillabaisse de tout cela, ou mieux encore, ils passent d'une position à l'autre suivant leurs humeurs ou l'adversaire qu'ils affrontent, car de nombreux écrits sur la conscience se font lors de grandes joutes internationales ou l'important est moins ce que l'on dit que d'avoir le dernier mot.

4. Ma position : la conscience en tant qu'expérience sensible ne peut être un objet de science.

Je vais ici défendre une forme d'éliminativisme, différent des deux présentés plus haut. En effet, loin de moi l'idée de nier le phénomène conscient ou de l'assimiler entièrement à son substrat neuronal. Mon objectif ici est d'argumenter et de plaider le refus d'octroyer à la conscience sensible le statut d'objet de science. Il est essentiel pour cela de définir ce que j'entends par objet de science.

Pour qu'un objet du monde devienne objet de science, il faut avant tout une communauté d'êtres humains disposée à l'étudier. Jusque là pas de problème, la conscience satisfait à ce pré-requis. Mais la deuxième condition est plus sévère et nettement plus sélective. Je vais l'énoncer de manière

extrêmement schématique mais plus de finesse n'apporterait rien à l'argumentation. Il faut que l'objet se prête avec succès à un processus d'objectivation. Processus qui, en toute première instance, force cet objet non seulement à apparaître comme une donnée fondamentale de nos perceptions mais bien comme une donnée et sa description sur lesquelles s'accorde la communauté dans sa majorité. Chacun dans cette communauté doit pouvoir le décrire d'une manière commune et univoque. L'accord s'établit bien sûr par l'entremise d'un dialogue et la mise à plat d'un même type de symbolisation. L'objet doit donc se prêter à une description symbolique en tant qu'objet (description qui peut s'accompagner de mesures), dont l'énoncé ou les résultats font également l'unanimité. S'il est vrai que chacun perçoit pour soi, il n'en reste pas moins vrai que dans de nombreux cas, nous avons l'intime certitude de percevoir la même chose. Même s'il s'agit là d'une illusion, l'illusion est totale. Tout le monde peut voir une pomme tombée, mesurer le temps de sa chute, lire la présence d'une force gravitationnelle dans une équation, suivre la résolution de cette équation et calculer le temps de la chute. Le tour est joué.

Tout le monde peut également se mettre autour d'une table, voir un objet bleu et se mettre d'accord sur le fait que cet objet soit bleu. Le problème est, et il est bien connu, que le processus d'objectivation s'arrête là, et que c'est un rien trop court lorsqu'on s'intéresse à la conscience. En fait, personne ne peut expliquer ce que cela fait ou est de voir du bleu, au point même bien sûr, classique expérience philosophique, que chacun pourrait percevoir ce bleu comme une couleur différente et ne pas le savoir.

La conscience peut-elle dès lors être appréhendée comme objet de sciences. Je ne pense pas, car elle est par essence le plus mauvais candidat à ce processus d'objectivation. Elle n'a de manifestation que privée et s'annihile en dehors du moi. Elle est en amont de tout et ne relève que de la sensation. Elle résiste donc à toute extériorisation et certainement à se présenter comme donnée du sensible collectivement perçue de manière univoque.

L'impossibilité d'une caractérisation objectivable du phénomène sensible se retrouve dans l'incapacité totale et définitive pour chacun d'entre nous, de percevoir dans tout autre que nous, humain ou animal, un être conscient. D'ailleurs, vous n'êtes tous que des zombies et je vous plains sincèrement.

Je pense également que ma position est légèrement plus nuancée qu'une séparation nette entre la perspective objective qui rendrait le problème simple et la perspective subjective qui le rend compliquée. Je ne me retrouve pas dans cette séparation 1ère/3ème personne mais préfère caractériser les objets de science, qui dès le départ sont le résultat de perception individuelle et subjective, par la facilité avec laquelle ils se prêtent à l'exercice d'objectivation. En d'autres termes, je préfère une version graduelle et surtout active (en préférant objectivation à objectif) de l'argument. C'est cette caractérisation qui fait, par principe, de la part sensible de la conscience le plus mauvais candidat à cet examen. Les sciences cognitives, plus que toute autre, recourent à l'introspection et à la perspective subjective pour identifier les objets de leur science comme: le raisonnement, la mémoire, la catégorisation, ensemble de phénomènes qui ne semblent n'exister que pour moi. Mais c'est évidemment la possibilité d'abstraire, de mettre sur la place publique, une certaine composante, la composante informatique, de ces objets mentaux, qui confère à cette discipline le statut de science.

Le fait de dénier à la conscience sensible le statut d'objet de science est-elle une position en tout point assimilable à l'une de celles présentées dans le chapitre précédent ?

Il est des positions qui manifestement sont à l'opposé de celle-ci. Les nouveaux dualistes ne peuvent être qu'en profond désaccord puisqu'ils partent du postulat que l'on peut faire une science de la conscience par l'entremise d'une nouvelle ontologie. Il s'agit certainement là de la position la plus diamétralement opposée à la mienne.

Qu'en est-il des éliminativistes durs ou mous. Là aussi, je les perçois comme assez distants, car soit ils nient purement et simplement le phénomène conscient (et pourtant j'en suis la preuve vivante) soit ils l'assimilent tout aussi brutalement au phénomène physique, ce que nulle part, je ne fais. Je suis persuadé, tout comme vous, du phénomène conscient, j'exprime simplement un doute fondamental quant à la possibilité de l'expliquer ou le modéliser scientifiquement.

Certaines positions se rapprochent de la mienne telles les « onfaitcommesiste » ou les « épiphénomalistes » qui acceptent de gommer le phénomène conscient de leurs études. Mais là encore, quelques bémols s'imposent. Je me démarque des épiphénomalistes quand ceux-ci donnent à

l'affirmation que la conscience est causalement neutre un statut scientifique. Dès lors qu'on ne peut parler du phénomène sensible en des termes scientifiques, qu'une telle affirmation de neutralité ait des velléités scientifiques, et elle perd toute sa consistance. Je me retrouve plus à l'aise avec les "onfaitcommesiste" qui poussent la démarche scientifique au plus loin dans le cerveau, à l'horizon de la cognition et à l'orée de la conscience, et dépensent une énergie conséquente à faire des objets cognitifs qu'ils rencontrent des objets de science. Mais là encore, plus ils se rapprochent des épiphénomalistes en affirmant que tout se passe comme si le phénomène sensible accompagnait par défaut le processus cognitif, plus je m'éloigne d'eux en soulignant les écarts dangereux qu'ils exécutent sans précaution aucune entre l'objectivable et le non-objectivable.

Restent les capitulateurs, catégorie avec laquelle je me sens bien sûr le plus d'affinité. Refuser aux phénomènes sensibles le statut d'objet scientifique est sans conteste une forme de capitulation. C'est une forme de capitulation qui ne se limite d'ailleurs nullement à la conscience sensible mais qui en général tend à se dresser contre l'omnivorisisme scientifique ainsi qu'à restreindre le champ des sciences aux seuls objets qui se prêtent au mieux à l'exercice d'objectivation. Je tiens cependant à marquer quelque distance avec ceux qui font du mystère de la conscience un mystère semblable aux ovnis ou la mort de lady Di, avec le lointain espoir d'une révélation ou d'un nouvel éclairage. Non, la capitulation se doit d'être franche et définitive, et la conscience sensible doit se résigner à jeter ses armes au pied des sciences.

5. Conclusions: Conscience sans sciences , oui et alors ??

Que la conscience résiste à l'investigation scientifique ne constitue nullement un obstacle à faire de celle-ci une notion opératoire, sérieuse et bien utile. La science n'a pas droit de cité partout. Aussi évident que cela puisse paraître, la notion de conscience sensible est essentielle à nos grilles d'appréhension du fonctionnement d'autrui. Indépendamment de mon impossibilité à vérifier la conscience de l'autre, le présupposé de son existence et, plus encore, d'une manifestation semblable à la mienne, est capitale au bon déroulement de nos relations. Cette attitude "instrumentaliste", "projective" et "mimétique" est également celle proposée par Dennet dans certains de ces livres comme étant la seule capable de conférer de la conscience aux machines. Là encore, je m'y oppose car je

répète que “charger” la conscience dans les machine implique à priori une identification objectivale de la conscience, pratique, par principe, impossible.

S'il est défendable que Popper ne daigne pas voir dans la psychanalyse une pratique scientifique, il n'en reste pas moins vrai que cette pratique, qui fait de la notion opératoire de “conscience” et “inconscience” ses fondements, peut être appréciée avec le plus grand sérieux, si ce n'est que par l'analyse statistique de ces résultats thérapeutiques. Je crois également très possible de faire une étude du phénomène subjectif sincère et puissante, de révéler à chacun, par une démarche introspective soutenue, des clefs d'accès à la conscience jusque là insoupçonnées. Le travail pourrait être à ce point systématique et sérieux que beaucoup lui attribuerait sans hésitation le label de “scientifique”. Les expériences “blindsight”, la pratique “psychiatrique” nous éclairent chaque jour un peu plus sur le rapport entre le phénomène sensible, le contrôle de nos actes et les traitements chimiques. Les articles qui sont écrits sont pris par tous avec le plus grand sérieux et font partie de la science médicale. Les aventures philosophiques au pays du clonage, des translateurs, et des greffes neuronales qui nous laissent inchangés, sont parfois des perles métaphysiques aux répercussions importantes. Les psychologues et autres psy ne nous ont pas attendu afin d'expliquer le phénomène sensible de manière telle que chacun se comprenne mieux, s'analyse mieux, se contrôle mieux ou se comporte mieux.

J'arrête tellement il en devient léger même suspect d'écrire de pareilles évidences. En substance, ceci devrait suffire à vous convaincre que refuser à la conscience l'entrée dans le club très privé des objets de science, d'abord ne fait pas de l'affiliation à ce club la panacée universelle pour la compréhension et l'amélioration de la condition humaine, mais surtout ne fait pas de tout autre travail d'investigation ayant pour objet le phénomène sensible un travail dénué d'intérêt. Il s'agit simplement de centrer ce travail dans un cadre interprétatif et opératoire approprié et tout malentendu peut être facilement évité.

Modèle, théorie, langage : le point de vue de la logique mathématique

Dominique Luzeaux*

luzeaux@etca.fr

* Laboratoire Perception pour la robotique

CTME/GIP, 16bis, av. Prieur de la Côte d'Or, 94114 Arcueil Cedex, France

Résumé

Nous allons principalement nous intéresser dans cet article aux relations entre modèle et langage, en tâchant de cerner précisément ce qu'est un modèle, comment il est exprimé (d'où le rôle prépondérant que va jouer le langage sous-jacent), et en quoi le langage d'expression va limiter sa portée. L'étude sommaire de cette problématique permettra ensuite de répondre aux diverses questions que l'on peut se poser, quant aux relations que l'on peut entretenir avec les modèles. Nous développerons le point de vue de la logique mathématique, en essayant de ne pas rentrer sur un plan technique, puis nous nous tournerons vers le monde de la physique théorique, ou plus exactement vers le monde de la théorie de la physique.

1 Définitions de base

1.1 Langages

Avant toutes choses, donnons-nous un langage par lequel nous pourrions exprimer des assertions. Communément, les langages utilisés en logique mathématique sont composés d'un vocabulaire de connecteurs logiques (et, ou, non, quel qu'il soit...) et de divers ensembles de symboles que l'on aime à considérer séparément pour simplifier les développements ultérieurs : symboles de variables, symboles de constantes, symboles de fonctions et symboles de prédicats. Ces divers ensembles sont supposés infinis, afin de pouvoir y « piocher » sans restriction.

Un langage sera dit *d'ordre zéro* si le lexique de symboles ne contient que des constantes ou des variables et les connecteurs logiques sont restreints à « et », « non » (et « ou » et « implique », par combinaison des précédents¹).

1. Tout au moins dans le cadre des logiques classiques car, par exemple pour les logiques intuitionnistes.

Si les divers ensembles de symboles précités sont utilisés et que les connecteurs logiques comprennent « quel qu'il soit » et « il existe », on parlera de *langages du premier ordre* lorsque ces deux derniers connecteurs peuvent s'appliquer uniquement sur les symboles de variables, et de langages du *second ordre* lorsqu'ils s'appliquent aussi sur les symboles de fonctions et de prédicats.

La notion de langage, comme elle a été présentée ici, ne doit pas être vue comme restreinte à des jeux mathématiques arides et décorrélés de toute réalité. À titre d'exemple, la simple introduction d'autres quanteurs, comme « il est possible », « quelquefois » ou « une infinité de fois », permet d'exprimer, par une extension triviale de ce qui précède, des concepts très variés. En fait, cette richesse va découler des paragraphes suivants, qui vont aborder sous l'angle de la logique mathématique les aspects syntaxiques et sémantiques.

1.2 Théorie

Une fois le langage posé, il est possible de se fixer des règles définissant les *expressions bien formées*. Par exemple, « p ou q » est une telle expression en logique des propositions alors que « p ou » ne l'est pas. Parmi ces expressions, un ensemble va être choisi et constituer les *axiomes*. Il ne reste plus qu'à se donner des procédures mécaniques de transformation d'expressions, que l'on appelle souvent *règles d'inférence* (comme le *modus ponens* qui permet en logique classique, à partir des expressions p et $p \supset q$, d'inférer l'expression q), et toutes les briques de base sont en place : il suffit d'appliquer de manière itérative les règles d'inférence sur les axiomes pour engendrer de nouvelles expressions, que l'on appellera *théorèmes*, puis d'utiliser ces derniers conjointement avec les axiomes et de leur appliquer les transformations autorisées, engendrant ainsi un ensemble potentiellement infiniment dénombrable de théorèmes. Cet ensemble constitue une *théorie*.

Les théories intéressantes – on les appellera *cohérentes* – sont manifestement celles où on ne peut montrer simultanément qu'une expression et sa négation sont des théorèmes, car dans ce cas il est facile de voir que toute expression bien formée est un théorème, ce qui enlève quelque peu de son intérêt à la procédure de preuve !

concernées par la constructibilité des raisonnements, il n'y a pas les lois habituelles d'interdéfinissabilité des quanteurs ; en particulier seule est valide : $(\neg p \vee \neg q) \rightarrow \neg(p \wedge q)$.

Ce qu'il convient de remarquer est d'abord le caractère purement syntaxique de la notion de théorie en logique, ainsi que l'hypothèse sous-jacente de finitisme (hilbertiste, pour être précis) : les axiomes sont en nombre fini et les règles de transformation, prises par définition comme des procédures mécaniques, sont appliquées un nombre fini de fois sur les théorèmes disponibles pour en engendrer d'autres.

Il est clair que l'on peut relâcher cette hypothèse finitiste, et considérer des *schémas d'axiomes* en place des axiomes (cf. Bourbaki pour la définition de la théorie des ensembles comme base des mathématiques), c'est-à-dire une procédure engendrant un nombre infini d'axiomes, ainsi que des applications infinies des règles d'inférence dans le processus de génération de la théorie (ceci est à la base des inductions transfinites par exemple). Cela modifie énormément les propriétés métamathématiques des théories associées, mais ne change rien quant à notre discussion qui insiste sur la *définition purement syntaxique d'une théorie*, comme ensemble des expressions engendrables à partir d'axiomes par des procédés de réécriture.

1.3 Modèle

C'est par la notion de modèle que va être introduite la sémantique. On suppose que sont donnés un langage et une axiomatique, c'est-à-dire toutes les règles, axiomes et procédures mentionnés ci-dessus. Pour rester à un niveau élémentaire, nous nous plaçons dans le cadre de la logique classique du premier ordre.

Donnons-nous une *structure*, constituée d'un ensemble appelé *domaine*, de fonctions dont les variables ainsi que l'image prennent leurs valeurs dans l'ensemble précédent, et de relations définies avec les mêmes contraintes. À titre d'exemple, l'ensemble des entiers naturels avec la fonction successeur et la relation d'addition usuelle ($R(x, y, z)$ si et seulement si $x + y = z$) est une structure possible. On définit alors une *interprétation* qui associe un élément du domaine à chaque symbole de variable (libre) du langage, une relation à chaque symbole de prédicat et une fonction à chaque symbole de fonction. En d'autres termes, l'interprétation associe la structure à l'axiomatique, et une expression bien formée va donc pouvoir être interprétée dans cette structure, ce qui va permettre de lui attribuer une valeur de vérité (la présentation donnée ici suit la définition de

la vérité issue des travaux pionniers de Tarski).

Il est clair que selon la structure une même expression formelle peut être vraie ou non : dans l'ensemble des entiers naturels, l'expression $((\forall x)(x = 0))$ est fausse, alors qu'elle est vraie dans la structure restreinte au singleton $\{0\}$, avec les restrictions évidentes des relations et fonctions sur les entiers. C'est tout l'intérêt de la notion de modèle : soit une structure et une expression bien formée ; on dira que la structure est un *modèle* de cette expression si cette dernière y est vraie. On dira aussi que l'expression est satisfaite dans ladite structure.

Au vu des paragraphes précédents, nous sommes maintenant en possession d'un côté de notions formalisant la *preuve* (comment dériver mécaniquement des théorèmes à partir d'axiomes de départ) purement syntaxique, et de l'autre côté de notions formalisant la *vérité* (associer à une expression d'un langage une valeur de vérité).

Avant de voir les liens entre ces deux notions, signalons qu'il est possible d'étendre la notion classique de vérité, et de considérer des logiques multivaluées (vrai, faux, indéterminé), voire avec une infinité de valuations (logiques de Lukasiewicz, logique floue dans le cas d'un continuum de valeurs de vérité). Par ailleurs, il est aussi possible d'envisager des modèles non monotones, c'est-à-dire que la valeur de vérité d'une expression peut varier en fonction par exemple d'un paramètre particulier (cf. logiques modales, logiques temporelles) ; en fait, cela revient à considérer des classes de structures en place d'une simple structure et de définir des relations d'accessibilité entre les diverses structures.

2 Langages et modèles : limitations et approximations

On peut se poser la question du lien entre vérité et preuve : tout d'abord il serait souhaitable que tout théorème fût vrai dans tout modèle. Cette propriété s'appelle l'*adéquation* de la logique, et elle est vérifiée entre autres par les logiques classiques des propositions et des prédicats.

La réciproque est bien plus intéressante, dans la mesure où elle tente de répondre à une question fondamentale de l'épistémologie :

la vérité est-elle mécaniquement déductible? En effet, si pour une logique adéquate, toute expression valide est nécessairement un théorème, cela signifie que les propositions vraies sont en fait exactement les formules syntaxiques que l'on peut dériver mécaniquement à partir des axiomes et des règles d'inférence. Cette question est à l'origine de la popularité de la logique et des excès de ce que certains appellent la gödelite. En effet, ladite propriété est la complétude sémantique de la logique, et, par exemple, Kurt Gödel a montré que la logique classique des prédicats était complète. Par contre, il a montré que l'arithmétique du premier ordre (c'est-à-dire la formulation des axiomes de Peano dans un langage du premier ordre, définissant les entiers naturels munis de l'addition et de la multiplication) était incomplète (en supposant qu'elle est cohérente), et plus généralement que toute théorie du premier ordre cohérente permettant au moins de manipuler les entiers naturels avec l'addition et la multiplication l'était aussi. Il a également montré qu'il n'était pas possible de montrer la cohérence de l'arithmétique du premier ordre (dans cette même théorie).

Revenons sur ces derniers résultats, qui montrent les relations entre les notions de modèle et de théorie, et entre langage et théorie. En effet, le fait que l'arithmétique du premier ordre soit incomplète n'entraîne pas, comme certains ont voulu le croire, que toute notion de vérité est au-delà des capacités des machines. Elle signifie en fait qu'il existe des propositions de l'arithmétique vraies mais non démontrables à partir d'un nombre fini d'axiomes exprimés dans un langage du premier ordre et en utilisant des procédures finitistes de déduction. C'est donc la puissance d'expression du langage formel et la notion formelle de preuve qui sont en cause dans cette incomplétude.

Il en est de même pour l'inaccessibilité de la preuve de la cohérence de l'arithmétique, qui montre simplement qu'il est nécessaire de sortir du cadre finitaire. D'ailleurs, Gentzen a donné une preuve de la cohérence de l'arithmétique, mais en utilisant l'induction transfinitie, ce qui est intéressant épistémologiquement car implique la nécessité de mettre en jeu des principes de démonstration plus puissants que ceux qu'un réalisme mathématique semblerait vouloir mettre en œuvre.

Notons que la logique nous donne ici une réponse à certaines questions : notamment, elle distingue clairement l'aspect démontrabilité

formelle, au sens d'une déduction suivant des principes syntaxiques donnés a priori, d'une adéquation avec une structure, qui pourrait être par exemple une représentation du monde physique observé. Elle met en évidence les limitations intrinsèques de cette démontrabilité formelle si l'on suit les principes de déduction classiques depuis la logique aristotélicienne, basés sur des fondements combinatoires finis. Par ailleurs, elle pose la recherche d'une théorie comme un rapport de sujet à objet, et la recherche d'un modèle comme le rapport de l'objet au sujet. Nous reviendrons sur ce point dans le paragraphe consacré à la modélisation en physique.

Nous avons vu de manière cachée le lien étroit entre le langage et les deux notions de théorie et de modèle ; mais il apparaît d'une façon flagrante avec la notion de modèles non standard. Revenons par exemple à la logique du premier ordre, qui est la plus utilisée en mathématique..., et qui pourtant est incapable de caractériser, à isomorphisme près, une structure infinie donnée (théorème de Löwenheim-Skolem) : en d'autres termes, toute théorie ayant un modèle infini en possède un second non isomorphe au premier, c'est-à-dire qu'il est impossible de les mettre en correspondance. Ceci est valable entre autres pour des structures aussi familières que les entiers naturels ou les réels. Il n'existe pas d'axiomatisation du premier ordre qui caractérise de manière unique ce que nous entendons par les entiers naturels : les axiomes classiques (il existe un nombre zéro, une fonction successeur, un schéma d'induction) sont satisfaits dans la structure habituelle, mais aussi dans d'autres structures appelées non standard (il faut imaginer quelque chose qui ressemble du point de vue de l'ordre aux rationnels positifs, mais avec des copies de l'ensemble standard des entiers relatifs à la place de chaque rationnel non nul, et une copie de l'ensemble standard des entiers naturels à la place de zéro, donc comme segment initial).

Cette présence de modèles non standard souligne qu'il existe des formules qui ne sont pas vraies dans tous les modèles, et par conséquent non démontrables, au sens syntaxique du terme. De plus, ces divers modèles non standard sont certes non isomorphes, mais indiscernables du point de vue de la théorie ; par exemple, tous les modèles standard ou non standard des entiers naturels satisfont les axiomes de Peano. Là où réside la différence entre ces modèles, et donc leur non-isomorphie, est qu'il y a d'autres formules (non expressibles dans

la théorie engendrée par les axiomes de Peano) qui seront satisfaites dans certains modèles et pas dans d'autres. En fait, l'incomplétude elle-même ne peut se voir que si on sort de la théorie et du langage choisi.

Par contre, si on passe à une axiomatique du second ordre, la situation n'est plus la même : d'une part il y a catégoricité, c'est-à-dire un seul modèle à isomorphisme près, et d'autre part l'axiomatisation est finie. C'est une conséquence du pouvoir d'expression bien plus fort du langage (peut-être trop : en logique du second ordre, on peut démontrer l'hypothèse du continu or cette dernière est indépendante des autres axiomes définissant la théorie des ensembles « usuelle », tout au moins en premier ordre, ce qui revient à dire que la logique du second ordre supprime des modèles de la théorie des ensembles, ce que rien dans la pratique des mathématiques ne justifie, si on adopte un certain réalisme ontologique des mathématiques). L'inconvénient est la perte de certaines propriétés : la logique du second ordre n'est pas complète par exemple, donc plus d'équivalence entre preuve et validité, et toute notion de conséquence ne peut être donnée que par la sémantique, ce qui peut paraître gênant, car la sémantique est fondée sur la donnée d'ensembles, donc sur l'existence d'une formalisation a priori d'une théorie des ensembles, d'où problème éventuel de circularité². D'autres propriétés utiles sont aussi perdues, comme la compacité, qui dit que si un ensemble de formules a des modèles finis arbitrairement grands, il a un modèle infini (propriété très pratique, par exemple, pour montrer que la notion de groupe fini n'est pas finiment axiomatisable par un langage du premier ordre).

Le choix du langage d'expression apparaît alors comme crucial pour une adéquation entre théorie et modèle, et il conditionne cette adéquation. Notons que les modèles non standard trouvent des applications pour certaines démonstrations, dans la mesure où ils les simplifient ; les physiciens du XVIII^e siècle n'hésitaient pas à utiliser des réels non standard infinitésimaux (sans le savoir, du moins comme exposé ici !) en physique pour calculer des dérivées par exemple ou

2. La logique du second ordre nécessite une théorie des ensembles, que l'on est donc obligé de construire par une logique du premier ordre, où l'on a bien équivalence entre preuve et validité, ce qui permet de définir sans recours à d'autres objets, des modèles par exemple d'une théorie des ensembles. Un de ces modèles est choisi pour construire la logique du second ordre, qui va permettre de définir une théorie des ensembles du second ordre, avec disparition de certains modèles exhibés dans le cadre du premier ordre du fait du théorème de Löwenheim-Skolem. Ce « boot-strapping » qui crée des chimères pour se fonder avant de les éliminer a de quoi intriguer le moins réaliste des mathématiciens...

des limites de suites. Mais cette mathématique non standard, même si elle simplifie certaines preuves, n'apporte rien de nouveau dans le principe (c'est une extension conservative), ce qui est intellectuellement compréhensible, puisqu'avec un autre langage d'expression ces modèles n'existent plus. Allant cependant un peu plus loin dans cette direction, nous sommes en droit de nous poser des questions fondamentales sur l'objectivité des êtres mathématiques...

3 Le point de vue des topoi

Comment ne pas évoquer brièvement la théorie des catégories, et en particulier les topoi, quand on recherche les liens entre langage, théories et modèles. En effet, le concept de topos est défini a priori pour subsumer ce que sont les univers possibles pour une mathématique et non pour fonder un univers particulier³. La présentation catégorielle est caractérisée par une indistinction entre syntaxe et sémantique, et contrairement à une formalisation particulière comme, par exemple, l'axiomatisation de Zermelo-Fraenkel de la théorie des ensembles, qui revient à une décision ontologique, la théorie des topoi est une description logique des ontologies possibles. De manière plus imagée, les topoi offrent une description diagrammatique (des flèches et des points – une caricature, mais à peine...), dont les propriétés (c'est-à-dire l'organisation des flèches) entraînent des choix ontologiques : à titre anecdotique, si un topos est « bien pointé », la logique de tout univers ainsi décrit est classique (entre autres, on a nécessairement le tiers exclu, alors que sinon on ne l'a pas); si un topos valide l'axiome du choix (par l'existence systématique de certaines flèches), il est booléen et sa logique est classique.

Pour résumer, nous avons soutenu précédemment que la recherche d'une théorie dont un modèle est donné (ce qui est couramment appelé dans la langue courante une *modélisation*) est fondamentalement conditionnée par le choix d'un langage formel, et les topoi font apparaître cette démarche comme la quête d'une certaine configuration d'un diagramme. La question de la création d'un modèle et des relations vis-à-vis de lui se ramèneraient dans ce cadre à un choix de description, en amont de toute décision ontologique.

3. C'est le point de vue optimiste des topoi vus comme langage fondationnel des mathématiques.

4 Liens avec la modélisation en physique

En sciences physiques, et notamment en relativité et en physique quantique, on retrouve les notions exposées ci-dessus : partant d'axiomes et de principes (définis pour développer un ordre logique des connaissances), des conséquences logiques sont déduites. Le concept de modèle apparaît lors de la phase de « validation expérimentale » (l'emploi de « validation », plutôt que « vérification » n'est pas innocent). C'est à ce niveau qu'il convient d'être prudent. En effet, du point de vue de la logique, il est nécessaire de disposer d'une structure pour définir l'interprétation réalisant la liaison syntaxe-sémantique ; cette structure elle-même est un objet mathématique descriptible dans un certain langage (et en fait dans une autre théorie « préexistante », du moins en regard de ladite structure), et concevable a priori dans sa globalité (car justement définissable via une certaine théorie). A contrario, la validation expérimentale se réfère à notre monde, que d'une part nous n'embrassons pas dans sa globalité, et que d'autre part nous ne pouvons sonder en toute liberté, pour des raisons technologiques. L'accès aux éléments du domaine (les données expérimentales) n'est donc pas immédiat, ni même possible. Là réside une des difficultés d'établir la liaison entre théorie et modèle, et la relation que nous pouvons avoir vis-à-vis de toute modélisation en sciences physiques est donc fondamentalement limitée.

Ceci peut entraîner des impossibilités d'établir une relation entre données observationnelles et théorèmes. Cela tient au fait que les règles de correspondance entre la réalité et la théorie (en d'autres termes l'interprétation) ne peuvent être définies dans l'absolu, comme c'est le cas dans le cadre logico-mathématique, mais sont nécessairement limitées par l'accès lui aussi nécessairement limité aux éléments du domaine.

Certains expriment cette limitation comme l'indécidabilité phénoménologique et l'expliquent par une relation autoréférentielle au niveau de la procédure expérimentale, dans la mesure où un appareil analyse l'équivalent de ses propres constituants.

Un autre point mérite d'être souligné dans la relation modèle-théorie en sciences physiques : un seul modèle est disponible, et le physicien se doit de choisir parmi des théories celle qui peut être

mise en relation avec *le* modèle. Ce renversement de perspective par rapport à la logique, où l'on adopte le point de vue d'une théorie et des divers modèles pouvant être mis en relation avec elle, se traduit par la problématique du choix des axiomes et règles d'inférence établissant une théorie. De nombreuses querelles philosophiques ont porté sur le choix des axiomes : sont-ils préexistants ?, quel est le rôle de l'intuition dans leur définition ? ; on s'est même posé la question de leur vérité, ce que le lecteur aura compris comme étant une mécompréhension, ou au moins un abus, de la relation théorie-modèle : le problème réside dans un choix d'axiomes tels que la théorie qui en découle puisse être mise en correspondance avec le seul modèle disponible.

Ceci dit, il convient de remarquer que la tâche du physicien théorique dans sa pratique de la modélisation est intrinsèquement limitée, car toute théorie qu'il choisira, même si elle admet le monde réel comme modèle, a de fortes chances d'accepter d'autres modèles (c'est l'incomplétude sémantique, que nous avons décrite ci-dessus, entraînant l'existence de modèles dits non standard, au sens où on ne peut les confondre avec le modèle standard privilégié), dès que la théorie a un minimum de pouvoir d'expression.

De plus, comme nous l'avons déjà souligné dans le paragraphe consacré aux relations entre langages et modèles, il n'y a aucune chance de discerner le modèle standard du non standard, si ce n'est en passant à une métathéorie. Ce qui signifie que d'une part le physicien peut définir une théorie acceptant d'autres modèles que celui qu'il considère comme standard, et d'autre part il risque de ne pas disposer de moyens de s'en rendre compte. D'où une inaccessibilité de principe, qu'il convient de ne pas confondre avec inconnaissabilité.

5 Conclusion

Finalement, la présentation logico-mathématique permet de distinguer clairement la théorie du modèle, et étudie leurs liens respectifs en regard du langage formel utilisé. Elle permet d'en analyser les limitations, tant intensives qu'extensives.

Les implications des divers résultats abstraits, obtenus dans le cadre logique, dans l'élaboration de théories en sciences physiques répondent à certaines interrogations que l'on attribue quelquefois à

tort à des impossibilités techniques, ou métaphysiques, en les désignant comme nécessaires conclusions du choix de langages d'expression (mais n'est-ce pas un nouveau retour à la métaphysique?). Pour y remédier, ou tout au moins s'en donner l'impression, il est possible de construire des hiérarchies de théories (et de langages) qui créent une irrésistible ascension vers une métathéorie unifiée sous forme de point d'accrétion de réseaux de théories (par exemple, les travaux de G. Ludwig en théorisation de la physique quantique vont dans cette direction).

Pour conclure, soulignons que nous sommes conscient d'avoir déployé une certaine arrogance propre au faitichiste de la métamathématique, dont les fondements souffrent pourtant aussi de certaines faiblesses (pataphysiques?) qu'il reproche aux autres...

Références

- [1] S. Abramsky, D.M. Gabbay, and T.S.E. Maibaum. *Handbook of Logic in Computer Science, vol. 1: Mathematical Structures*. Oxford University Press, 1992.
- [2] A. Badiou. Platon et/ou Aristote-Leibniz. Théorie des ensembles et théorie des topos sous l'œil du philosophe. In M. Panza and J.-M. Salanskis, editors, *L'objectivité mathématique : platonismes et structures formelles*. Masson, 1995.
- [3] H.D. Ebbinghaus, D. Flum, and W. Thomas. *Einführung in die mathematische Logik*. Springer Verlag, 4 Auflage, 1996.
- [4] M. Felden. *Le modèle géométrique de la physique : l'espace et le problème de l'interprétation en relativité et en physique quantique*. Masson, Nouveaux horizons scientifiques, 1996.
- [5] Y. Gauthier. *Logique et fondements des mathématiques*. Diderot Éditeur, Arts et Sciences, 1997.
- [6] J. Largeault. *La logique*. Presses Universitaires de France, Que sais-je?, vol. 225, 1993.
- [7] G. Ludwig. *Die Grundlagen einer physikalischen Theorie*. Springer Verlag, 2 Auflage, 1990.
- [8] D. van Dalen. *Logic and structure*. Springer Verlag, 1994.

Modèles abstraits en perception artificielle

Jacques Blanc-Talon*

blanc@etca.fr

* Laboratoire Perception pour la robotique

CTME/GIP, 16bis, av. Prieur de la Côte d'Or, 94114 Arcueil Cedex, France

Résumé

Nous discutons de deux approches pour la vision artificielle, opposées dans la nature de la modélisation qu'elles apportent. La première considère l'image numérique comme la simple mesure physique d'un phénomène plus complexe et, comme telle, vise à la généraliser dans une démarche constructive; la limite de la construction peut d'une certaine manière s'appeler l'image perçue et être identifiée à un modèle abstrait de l'image. La seconde approche prend l'image mesurée, qui remplace alors l'image perçue, comme point de départ d'un processus de structuration géométrique; le modèle de la perception est alors un modèle basé sur une approximation du réel, qui ne généralise pas l'image numérique mais fédère sa mesure et les actions structurantes. Nous prenons comme exemples respectivement l'approche multifractale et la théorie multi-échelle, appliquées au traitement d'images et à la vision artificielle. De manière générale, nous opposons l'interpolation du mesuré et l'approximation du perçu.

1 Vision, perception et cognition

La perception est l'ensemble des processus qui président à une élaboration abstraite du réel perçu, c'est-à-dire pour paraphraser Platon, l'exercice raisonné des sens qui autorise à des objets physiques différents dans le temps ou l'espace de posséder ou non la même identité. La *perception artificielle*, dont le champ d'application est généralement restreint au seul processus de vision (bien que cette restriction ne soit pas forcément limitative dans sa richesse)¹, est l'ensemble des mécanismes mis en place pour construire une représentation symbolique du réel mesuré.

1. Il faut ici être juste et mentionner qu'il existe un certain nombre d'autres travaux, par exemple sur la localisation acoustique 3D ou la robotique avec actionneurs permettant d'émettre un retour d'information tactile. De même, nous ne parlerons pas de reconnaissance de la parole, etc.

Faisons preuve de pragmatisme en jugeant la perception naturelle (pour faire opposition à la perception artificielle) comme la référence en la matière et donc comme le but ultime à atteindre (avant peut-être un jour de le dépasser). Dans ce cas, c'est cette même raison orchestrant les sens qui va également interpréter et apprécier, comme le Roi jugeant son artisan horloger, les résultats des rouages de l'artificielle perception. Cette remarque est importante car elle indique les limites globales de tout système de vision artificielle qui ne sont ni plus ni moins que celles d'une sorte de modèle sous-jacent et de ses hypothèses d'application. Ce sont ces limites que la recherche a pour but de dépasser ; ce faisant, on repousse parfois une question fondamentale sur la manière dont elles peuvent être atteintes. Ce point est discuté dans la partie 2 et nous conduit à adopter la dichotomie approximation/interpolation.

La transposition entre sens et raison d'un côté, et capteurs² et algorithme de l'autre, s'effectue en formalisant les acteurs mentaux que nous pensons entrevoir dans l'acte de perception d'une part, en délimitant le réel avec les aides conjuguées de la physique et de la géométrie d'autre part, puis en modélisant d'une manière ou d'une autre l'action des premiers sur les sous-espaces réels retenus en second. En d'autres termes, on se donne des « instances » d'une part dont l'évaluation à partir des éléments physiques possibles d'autre part permet de réaliser la perception automatique, puis on cherche quelle démarche logique d'ensemble conduit effectivement à percevoir quelque chose. Cette démarche revient nécessairement à dégager des modèles plus ou moins formels du monde physique, de la perception et du raisonnement. « Le modèle est un outil de pensée utile et inévitable parce qu'il nous permet de penser en termes familiers à des choses qui ne le sont pas » (Bridgman).

Si les modèles de la perception sont très variés dans leur nature et leur mode d'exercice, ils ont en commun une certaine décomposition mécaniste de la perception qui est justifiée, de manière plus ou moins affirmée, soit par référence à l'approche d'inspiration biologique, soit par le vœu pieux que l'assemblage de modèles très généraux produise un modèle réputé universel, ce sans souci ni d'une quelconque « suffisance » physique, ni de leur cohérence mathématique (à moins que ce ne soit dans l'impuissance de les vérifier).

2. Caméra visible ou infrarouge, antenne radar, télémètre laser, etc.

Ce double souci d'*universalité* et de *cohérence* est central aujourd'hui dans nombre de disciplines. Dire d'un modèle qu'il est universel revient à vérifier effectivement qu'il peut produire n'importe quelle manifestation du phénomène qu'il modélise ; un excellent exemple est la machine de Turing qui est un modèle universel en théorie de la calculabilité (Thayse). Tout est alors prévu par le modèle qui est identifié au phénomène et il ne reste plus qu'à trouver les bons paramètres, ce qui ne serait pas sans rappeler une froide déclaration de Lord Kelvin si le spécialiste ne se rendait pas compte que dans le cas général, justement, le calcul des paramètres n'est souvent pas réalisable. La condition d'universalité peut pourtant être comprise différemment suivant qu'on effectue de la modélisation, qui vise à un but d'*identification*, ou de l'approximation, dans un but de *synthèse*.

L'affirmation de cohérence logique d'un ensemble de modèles interconnectés, quant à elle, signifie que deux informations utilisées par des modèles différents ne peuvent jamais être incompatibles et que les résultats de l'ensemble tout entier ne peuvent pas aller à l'encontre des hypothèses des modèles partiels. C'est un point en général très difficile à vérifier, ne serait-ce qu'en raison des domaines un peu « flous » de certaines assertions.³ Une approche pratique consiste alors à fusionner les informations délivrées par les différents modèles, ce qui peut être fait de différentes manières (nous renvoyons le lecteur à l'ouvrage de Mongi Abidi pour un exposé des principales méthodes) et permet de repousser le problème de la cohérence en dehors d'un cadre applicatif immédiat, voire même de gommer des possibles incohérences par un traitement approprié. De cette manière, même lorsque différentes modélisations successives ne sont pas pleinement compatibles, leurs incohérences ponctuelles peuvent être artificiellement écartées.⁴

Ainsi dans les systèmes automatiques la perception est-elle considérée comme une certaine forme d'intelligence qui peut être soit une intelligence dynamique (et nous rentrons de plein pied dans le domaine de l'intelligence artificielle), soit un certain ensemble de connaissances directement enfouies dans les modèles. Ce second point, un peu trop embelli par la désignation générique d'*a priori*,

3. Au sens propre comme au figuré.

4. L'être humain utilise des informations externes pour résoudre certaines illusions d'optique.

mérite quelque attention. Par exemple, l'utilisation d'un algorithme de flot optique⁵ pour recadrer un objet en mouvement dans le champ d'une caméra implique entre autres, sans que cela soit souvent bien considéré par la suite, que l'on sait qu'on observe un objet non déformable en déplacement non chaotique. Ce genre de connaissances influe, certainement de manière plus fine que simplement binaire, sur les résultats de tout système perceptif qui les emploie.

Ces dernières années ont vu naître des modèles pour lesquels on ne peut guère avoir intuition des implications, ainsi que des nouvelles applications de la perception artificielle pour lesquelles nous ne disposons pas d'un acquis préalable. Concernant les premiers, on est souvent en peine d'en exprimer l'a priori. Concernant les secondes, l'acquis injecté dans les modèles perceptifs n'est souvent qu'une transposition de l'expérimentation dans des applications plus simples : Cordier a déjà discuté de la difficulté des professionnels d'exprimer leur savoir-faire en mots, puis de traduire ces mots dans un cadre formel. Par exemple dans le traitement des images radar, le néophyte ne comprend pas ce qu'il voit. Il utilise son expérience du domaine visible, clairement inappropriée.

L'utilisation des modèles que nous avons baptisés de *modèles abstraits* (par manque d'un meilleur nom), au départ essentiellement confinés aux mathématiques ou à la physique, se répand aujourd'hui dans d'autres sciences, et notamment en vision par ordinateur. Nous allons tenter d'en aborder deux types : le modèle fractal et multifractal, et la théorie multi-échelle. Comment ces modèles se placent-ils dans la reconstruction d'une perception, à quel degré d'universalité peuvent-ils prétendre et quel biais leur structuration a priori ajoute-t-elle de manière implicite ? Notre propos est de montrer qu'on peut les opposer sous les critères assez particuliers et rarement abordés qui sont ceux d'approximation et d'interpolation.

2 Interpoler le mesuré ou approcher le perçu ?

Il existe des différences fondamentales entre l'action de *synthétiser* et celle d'*identifier* un même phénomène, c'est-à-dire dans notre

5. Technique permettant, sous certaines hypothèses physiques, de retrouver le déplacement d'un motif d'une image à une autre. Parmi ces hypothèses figure celle d'une certaine continuité dans le déplacement.

cas un ensemble de données qui code toute l'information observable. Rappelons au passage que ce caractère d'observabilité des données ne présuppose en rien que l'on puisse uniformément les modéliser à une précision quelconque : par exemple, les phénomènes chaotiques peuvent être correctement approchés sous de bonnes hypothèses, mais seulement sur de petits voisinages autour de certains points de fonctionnement (pas d'a priori subversif ici). Les deux approches sont à l'évidence duales, en ce sens qu'elles utilisent parfois les mêmes modèles et que la synthèse autorise l'identification.

On retrouve cette dualité en analyse numérique classique suivant qu'on effectue de l'*interpolation* ou de l'*approximation* de fonctions ; en reconnaissance syntaxique des formes dans la dualité *analyse/inférence* suivant que l'on reconnaît ou que l'on engendre le langage des formes ; ou pareillement encore en géométrie fractale (nous avons déjà discuté de ce point dans d'autres publications).

En identification, on dispose d'un ensemble de points ordonnés que le modèle doit pouvoir toujours rencontrer : toutefois en dehors de ces valeurs, l'*interpolant* peut avoir n'importe quel comportement, conditionné par son nombre de degrés de liberté fixé par avance au nombre de données (auquel peut venir s'ajouter un nombre de paramètres internes). L'*interpolant* calculé permet ensuite de trouver automatiquement des valeurs intermédiaires entre chaque couple de données. On ne dispose en général d'aucune précision supplémentaire sur la qualité de ces prédictions.

Dans le cas de la synthèse, on part du même ensemble ordonné qui baigne cette fois-ci dans un espace métrique (un espace muni d'une distance mathématique⁶) et également d'un critère d'erreur basé sur la distance. Le nombre de degrés de liberté du modèle est bien moindre, c'est-à-dire qu'il **n'interpole pas** mais approche l'ensemble de données. Par construction, il ne s'écarte jamais de ces données au-delà d'une erreur préfixée ; en général, le modèle affiche d'autres propriétés (stabilité, retard de la sortie par rapport à une variation d'entrée, etc.) auxquelles nous ne nous intéressons pas ici.

Dans son essence et avec l'acception platonicienne du mot, la perception humaine peut être envisagée comme une certaine approximation contrôlée du réel perçu (et il ne nous importe pas de discuter

6. C'est-à-dire qui vérifie les axiomes élémentaires de positivité, symétrie, non-dégénérescence et inégalité triangulaire.

davantage ni de ce que pourrait être cette approximation, ni des sens et implications des mots « contrôlé » et « réel »), tandis que la perception artificielle devient une certaine interpolation du réel mesuré (ou perçu). Par exemple, un homme né maladroit pourra malgré tout apprendre à conduire une voiture, puis un avion et un bateau ; sa conduite sera probablement la même à chaque fois bien qu'adaptée au type du véhicule, mais elle sera efficace car il sera capable d'approcher à une précision suffisante sa trajectoire, sans certainement jamais passer par les repères de conduite. Par contre, construire selon une démarche classique par blocs (i.e. *top-down*) un système de vision permettant à un robot d'accomplir la même tâche demandera de redéfinir dans chacun des cas le modèle afin qu'il puisse effectivement interpoler les données qui lui sont fournies.

Nous mettons en évidence au cours de la partie suivante cette nuance entre le mesuré et le perçu : d'une certaine manière, cette démarcation transpose dans le cadre de la vision artificielle les réflexions sur l'aspect objectif et l'aspect subjectif de l'espace. Notre propos est de montrer au travers de deux exemples comment cette interpolation peut elle-même être construite autour d'un ensemble de processus interpolants, ou bien au contraire être approchée par un modèle généralisateur. Cette présentation approchée d'une interpolation peut d'abord sembler paradoxale, mais elle nous conduira à une conclusion inattendue et personnelle sur les modèles utilisés en perception artificielle.

3 Modèle multifractal et théorie multi-échelle

L'émergence relativement récente de la physique fractale⁷ a bouleversé notre compréhension du monde qui nous entoure. Tout d'abord simples jouets mathématiques, les *fractals* sont devenus un moyen de description puissant de phénomènes liés à la turbulence, à la non-linéarité, et au chaos ; rapidement, ils sont apparus dans des domaines d'où l'on aurait pu les penser absents, comme en algèbre ou en physique quantique (la théorie des nombres ayant fourni quant à elle les premiers « monstres » bien avant la lettre). Mais au-delà de

⁷ Ce nom a été proposé par B. Mandelbrot, et est utilisé indifféremment comme adjectif ou nom commun, en français et en anglais ; le pluriel de « fractal » demeure « fractals ».

la simple description, le succès des recherches, tant en identification qu'en synthèse des phénomènes sous-jacents, a permis une utilisation insoupçonnée et fructueuse des modèles fractals. On retrouve ainsi une même description élégante de phénomènes aussi variés que la croissance de plantes, le tissu pulmonaire, la géométrie des nuages, la cartographie des fleuves ou des régions côtières, le système cardiovasculaire, les feux de forêt, et même certains amas d'étoiles.

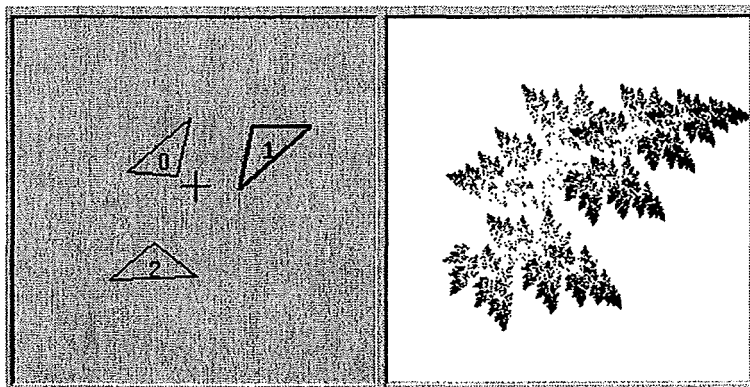


FIG. 1 - Un fractal et ses 3 générateurs. À droite, chaque zone semblable à un triangle de gauche contient les 3 autres zones triangulaires. Les variations d'intensité indiquent le nombre de fois que chaque point a été visité.

En fait, ce sont les multifractals qui devraient trouver une utilisation générique en vision artificielle ; toutefois, afin de rendre la discussion plus claire, expliquons d'abord ce qu'est un fractal auto-similaire dans le plan, c'est-à-dire le type de fractal le plus simple. Par construction, c'est un ensemble contenu dans un rectangle fini qui possède la propriété particulière suivante : celle de contenir en son intérieur une infinité de copies de lui-même (figure 1).

Pour trouver le processus qui produit ce fractal, on sélectionne d'abord l'ensemble des plus grosses de ces copies qui sont toutes différentes les unes des autres. Par hypothèse, on suppose qu'elles proviennent d'une réduction du fractal entier d'un certain facteur, comme si une photocopieuse avait été utilisée. Si on imagine qu'elles ont été tirées sur papier calque, alors on doit pouvoir ensuite ajuster et coller ensemble dans n'importe quelle position toutes ces copies

de manière à retrouver le fractal de départ. Le fractal est dit auto-similaire parce qu'il est constitué du collage d'une infinité de copies similaires à lui-même (théorème du collage de Barnsley).

De quelle manière les fractals sont-ils utilisés en vision artificielle? Dans un premier temps, la méthode la plus brutale et pourtant la plus efficace consiste à « fractaliser » sans autre forme de procès ce qui présente dans une image, de manière raisonnablement évidente, des traits caractéristiques de la fractalité. Les nuages atmosphériques, par exemple, offrent une telle particularité et on peut aisément imaginer inclure un modèle fractal de nuage dans un système de vision. Dans le champ du capteur, la nuée sera interpolée par le modèle et les paramètres discriminants transmis à des algorithmes subséquents pouvant les prendre en compte : on aura identifié le nuage.

Ce qui peut être plus difficile à justifier, dans un second temps, est que l'on peut chercher à étendre, voire à généraliser, cet emploi du modèle fractal, et ce dans au moins deux directions. La première est assez technique puisqu'elle consiste en quelque sorte à traduire la terminologie des modèles classiques dans la terminologie fractale afin de trouver des interpolants analogues. De cette manière, on peut choisir ou calculer un modèle fractal qui interpole toutes les données et présente, dans le même temps, les mêmes grandeurs caractéristiques principales. Du point de vue de la complexité, cela semble normal puisque la géométrie fractale, sans être universelle, est plus générale que la géométrie euclidienne (Barnsley).

La seconde direction de généralisation est plus difficile à justifier. Elle se base sur une propriété importante des fractals qui est de ne jamais pouvoir être véritablement observés puisqu'il faudrait, au moins dans le cas géométrique, pouvoir les scruter avec une précision infinie. Dès lors, puisque le fractal véritable est inaccessible, rien n'empêche de calculer une suite d'interpolants fractals dont la limite n'interpole plus aucune donnée physique, mais converge vers l'approximation d'un fractal hypothétique. Ce modèle fractal, rejeté à l'infini, approche uniformément les données.

Les *multifractals* sont une généralisation physique et mathématique des premiers. En effet, la géométrie fractale basée sur les seuls ensembles auto-similaires est par trop simpliste pour aborder des phénomènes physiques un peu génériques. De manière simplifiée, un phénomène multifractal est tel qu'en chacun de ses éléments il

existe un comportement fractal différent de celui de ses voisins. Par exemple, une place publique sur laquelle se seraient regroupés des mics-macs, des abénaquis, des algonquins et des inuits pour un évènement annuel. . . Chaque groupe ethnique aurait une répartition fractale différente, variant suivant sa position par rapport au centre d'attraction, et la population qui les rassemble en afficherait une autre, variable dans le temps. Les grandeurs caractéristiques d'un multifractal dépendent des coordonnées (spatiales et temporelles) suivant des lois particulières ; de manière plus précise, ces lois décrivent des transferts de quantité d'information, d'énergie, etc., entre différentes échelles des coordonnées.

En vision artificielle, l'image numérique peut être identifiée à un phénomène multifractal via une justification mathématique assez complexe qui laisse, dans l'application, un grand nombre de degrés de liberté. Sans être particulièrement critique mais plutôt simplement réaliste, force est d'admettre qu'on laisse donc choir un choix ad hoc des paramètres sur le tableau noir d'une théorie. Pourtant, calculer exactement les paramètres du modèle permettrait de disposer d'une description extrêmement riche et compacte du phénomène, ce qui est une motivation suffisante pour son application.

Une fois de plus, le multifractal véritable ne peut pas être observé, puisqu'il faudrait cette fois-ci disposer d'une précision et d'une puissance d'analyse infinies. Les multifractals permettent pourtant de justifier mathématiquement de quelque chose d'assez particulier et nouveau : la fractalité est un *effet de bord* de quelque chose de plus fondamental⁸, en d'autres termes, on dispose d'un *modèle abstrait* algébrique dont une propriété permet d'exprimer le comportement fractal, et par voie de conséquence les formes observées par notre système de perception artificielle. Ce modèle permet de transporter les structures géométriques perceptibles du global vers le local en assurant la cohérence algébrique.

Opposons maintenant à ceci une approche quelque peu antinomique.

La théorie multi-échelle est une théorie naissante d'une quinzaine d'années, du moins dans sa formalisation, ainsi que le souligne T. Lindeberg. Sa légitime aspiration est de proposer une formalisation idoine pour manipuler des structures géométriques à différentes

8. La propriété de nilpotence dans les groupes de Gromov et de Heisenberg.

échelles, aspiration motivée par la constatation que la perception des détails est gouvernée par le degré d'attention que l'on porte à notre sujet. Il ne s'agit pas à proprement parler d'une théorie générale des structures géométriques puisque l'information structurelle n'est contenue que dans l'image numérique, mais c'est une théorie qui fournit un cadre de traitement des structures adapté à la perception artificielle.

En effet, on peut à partir d'un signal donné produire une famille de signaux dérivés (dont les représentations sont des images de moins en moins précises) en supprimant successivement des structures de moindre importance alors que l'on franchit l'espace des échelles de la plus fine vers la plus grossière en faisant varier un paramètre d'analyse ; on se trouve en quelque sorte dans la position d'un myope qui s'éloigne du centre de son attention. Quelle que soit l'échelle, les caractéristiques des images peuvent être obtenues soit en combinant des objets géométriques, soit en effectuant des opérations mathématiques sur l'image. Ceci n'est possible qu'en raison des formalisations précises, ainsi que le clament les détracteurs de la théorie, de la causalité, de l'invariance d'échelle et du comportement des structures.

À l'inverse, on peut imaginer de raffiner les signaux les plus grossiers pour obtenir de nouveaux signaux de plus en plus fins. On pourrait alors mathématiquement franchir l'image mesurée pour, à la limite mathématique, obtenir comme dans le cas multifractal une image de précision infinie qui approcherait un modèle abstrait de l'image perçue.

Ceci en fait n'est pas prévu et n'a pas de sens dans la théorie multi-échelle. Le transport de la forme s'effectue cette fois-ci d'un continuum local contenant toute l'information de détail vers le global, considéré d'une certaine manière comme plus riche en information sémantique, ce qui est notre entendement naturel du processus d'abstraction des formes géométriques.

4 Conclusion ... liminaire

Il est intéressant de constater qu'il existe en perception artificielle une sorte de syncrétisme tacite entre les modèles relevant d'une interpolation et ceux relevant d'une approximation, utilisés de manière indifférenciée.

Pourtant, nous avons montré qu'il existait des différences fondamentales dans les principes mis en cause entre les deux approches et que cela conduisait à des divergences de fond sur l'universalité de la perception réalisée. Tandis que les premiers proposent une perception artificielle qui se rapproche de la perception naturelle ou plus précisément de l'approche d'inspiration biologique, les seconds en réfèrent à des modèles que nous avons qualifiés d'abstraits, en général exprimés par un passage à une limite mathématique, qui tentent de généraliser l'image mesurée avec, admettons-le, un souci seulement secondaire de corrélation à l'intuitif.

Ainsi, on voit que l'approche multifractale va nous conduire (nous employons le futur car l'approche théorique est aujourd'hui loin d'être mature) à l'approximation d'une abstraction complexe de l'image mesurée. On peut dire que cette approximation permet de rejeter à une certaine limite mathématique ce qu'on peut considérer comme le phénomène réellement perçu. Concernant la théorie multi-échelle, le passage à la limite annihile au contraire toute l'information, permettant de considérer l'image numérique directement comme l'image perçue et non plus comme l'image mesurée.

Le passage à une certaine limite est un aspect obligé d'une modélisation qui permet de généraliser le concept sous-jacent. Dans un contexte d'interpolation, « le passage à un nombre illimité de variables et d'exemples conduit à effectuer de l'approximation » (Davis). Bien sûr, la transposition brutale des principes de l'analyse mathématique aux sciences cognitives et de la perception peut conduire à des interprétations abusives basées sur trop peu de constatations. Pourtant, c'est un fait que toute théorie contient en elle-même les germes de sa généralisation.

Plutôt que de conclure par une opinion personnelle favorable ou laconique, nous donnons comme argument supplémentaire (et récréatif) la théorie des graphes transfinis (d'après le livre de Zemanian) qui est basée sur la théorie des graphes.

Un graphe est un objet mathématique qui correspond, dans sa formalisation première, très exactement à la notion intuitive qu'il évoque : des sommets reliés par des arcs, orientés ou non, porteurs de valeurs numériques ou pas. Rien n'empêche de faire tendre le nombre de sommets (l'ordre) du graphe vers l'infini : cette généralisation première conduit à la notion de graphe infini, dont le passage

à la limite peut lui même être envisagé au moyen de la théorie des nombres ordinaux⁹.

Dans cette théorie donc, si le concept d'infini est toujours essentiel, une grandeur rejetée à l'infini n'est plus inaccessible et peut être observée. Le besoin d'une telle approche semble être apparu en théorie des graphes infinis appliquée aux circuits électriques, bien que nous ayons eu des besoins similaires en approximation de formes fractales.

Ainsi, les mathématiques transfinites nous permettent de contourner l'obstacle et de boucler cet exposé ... par l'infini !

Références

- [1] Abidi (Mongi A.) et Gonzales (Rafael C.). – *Data Fusion in Robotics and Machine Intelligence*. – Academic Press, 1992.
- [2] Barnsley (M.). – *Fractals everywhere*. – Academic Press, 1988.
- [3] Blanc-Talon (Jacques). – Une méthode mixte numérique/symbolique pour l'inférence grammaticale de fractals. In : *Dixièmes Journées Francophones d'Apprentissage*. – Grenoble, France, avril 1995.
- [4] Bridgman (Percy W.). – *The logic of modern physics*. – Mac Millan, 1927.
- [5] Cordier (F.), Denhière (G.), George (C.), Crépault (J.), Hoc (J.-M.) et Richard (J.-F.). – *Traité de Psychologie Cognitive*, chap. Connaissances et représentations. – Dunod, 1990 volume 2.
- [6] Davis (Philip J.). – *Interpolation and approximation*. – Dover, 1975.
- [7] Lindeberg (Tony). – *Scale-Space Theory in Computer Vision*. – Kluwer, 1997.
- [8] Thaysé (André) et et al. – *Approche logique de l'intelligence artificielle*. – Dunod, 1988.
- [9] Zemanian (Armen H.). – *Transfiniteness for graphs, electrical networks and random walks*. – Birkhäuser, 1996.

⁹ La théorie des ordinaux traite des nombres comme par exemple les cardinaux d'ensembles infinis tels \mathbb{N} et \mathbb{R} .

Modèles neurophysiologiques sur la détérioration de la mémoire conceptuelle chez les patients d'Alzheimer : confrontation de deux modèles.

Herminia PERAITA *
José Luis DOBATO.**

**Departamento de Psicología Básica I.U.N.E.D.
Ciudad universitaria s/n. 28040. Madrid*

***Hospital Nuestra Señora de Sónsoles.INSALUD. Avila.*

1. Résumé.

Cette communication se propose d'analyser un modèle (Damasio, 1989, 1990) fondé sur une architecture neurale autre que l'architecture classique, laquelle s'inspirait des architectures localisationnistes. L'objectif de Damasio est d'expliquer les processus de mémoire et de reconnaissance d'objets, d'événements, de situations, etc., ainsi que la détérioration des catégories sémantiques relevée dans certaines démences sémantiques (Alzheimer, corps de Levy) et autres types de lésions du SNC.

Ce modèle est important dans la mesure où il remet en question certaines théories ou explications neurologiques traditionnelles, qui, jusqu'il y a peu de temps, cherchaient à interpréter des "dissociations sémantiques" de plusieurs types et modalités. L'objectif de ce travail est d'évaluer les apports du modèle de Damasio (1989, 1990), en partant de certains résultats expérimentaux préliminaires d'une batterie de tests (E.M.S.D.T.A : Évaluation de la Mémoire Sémantique dans la Démence Type Alzheimer - Perais, Galeote, González-Labra et Sánchez Bernados, 1997) qui, dans une certaine mesure, constituent un ensemble de valeur prédictive permettant d'évaluer la détérioration de la représentation en mémoire sémantique de catégories d'objets naturels et d'objets artificiels, ainsi que de catégories supraordonnées et subordonnées/de base.

Au cours de cet exposé, nous nous efforcerons de décrire non seulement l'architecture neurale, les structures critiques ou fondamentales de cette architecture, ainsi que la notion de zones de convergence et le "binding problem", mais aussi les restrictions qui organisent la connaissance dans le S.N.C. : restrictions neurobiologiques et restrictions relevant du milieu.

Si ces dernières peuvent être décrites à plusieurs niveaux -physique, chimique, biologique- celles qui nous intéressent ici se situent au niveau psychologique, soit les catégories qui déterminent le sens des choses qui nous entourent : objets aussi bien naturels que fabriqués et événements. L'interaction entre ces deux pôles

(neurobiologique et milieu physique) conduit à un processus de groupement, ou catégorisation, en fonction de la similitude de la structure physique des objets, ainsi que de leur localisation dans l'espace, de la séquence temporelle des faits et de leur coïncidence dans le temps.

Partant de ce modèle, on analysera les restrictions neurobiologiques et celles relevant du milieu, la formation de domaines (sémantiques) et la complexité conceptuelle, pour ensuite étudier la possibilité d'élargir ce modèle à d'autres domaines cognitifs. Cette analyse fait apparaître comme nécessaire la reconceptualisation de l'idée, plus ou moins traditionnelle, de représentation du sens (meaning), en faisant intervenir, pour l'aborder, l'idée centrale de temporalité et de représentations distribuées.

On se fondera sur cette présentation de données empiriques qui, à notre avis, viennent à l'appui du modèle de Damasio.

2. Introduction

La mémoire est un système cognitif très complexe, qui comprend de multiples aspects et composantes. L'une des dichotomies les plus classiques permettant d'analyser la structure de ce système, dichotomie mentionnée pour la première fois par Tulving, consiste à faire la distinction entre mémoire épisodique et mémoire sémantique et vaut aussi bien pour l'analyse du traitement normal que pour celle du traitement pathologique.

La mémoire épisodique correspond au souvenir de faits, d'épisodes ou d'événements vécus personnellement, et qui par conséquent renvoient à des lieux, à des personnes et à un temps spécifiques. Au contraire, la mémoire sémantique, appelée par certains auteurs "mémoire conceptuelle", et composante de la mémoire à long terme, se rapporte à des connaissances culturellement partagées, suracquises et qui, par conséquent, ne renvoient à aucun lieu et à aucun temps spécifique. Plus précisément, la mémoire sémantique renvoie à des objets et à des concepts, ainsi qu'aux mots et à leur sens.

Le processus de détérioration de la mémoire sémantique dans la maladie d'Alzheimer entraîne, outre d'énormes problèmes de désignation, l'altération ou la perte sélective de certaines catégories sémantiques par rapport à d'autres. Par exemple : la détérioration des catégories liées aux objets animés par rapport à celles relatives aux objets inanimés. Ce fait est des plus contestés dans la littérature psychologique et neuropsychologique actuelle, les constatations faites ne pouvant en aucun cas s'imposer de manière définitive fautive, entre autres choses, d'une explication satisfaisante. De plus, la dissociation inverse a été constatée chez d'autres patients, à savoir une détérioration des catégories liées aux objets inanimés.

Par ailleurs, il semble que ces patients traitent les dessins plus aisément et plus rapidement que les mots, ce qui peut s'expliquer par un type de prétraitement sémantique des clés picturales relatives aux structures des objets.

Le présent travail a pour but de dégager une évidence confirmant, ou infirmant, la dissociation décrite en premier lieu, suivant laquelle ces patients ont davantage de difficultés à traiter les catégories animées.

Les explications, et/ou hypothèses, permettant d'expliquer cette dissociation, catégories animées/catégories inanimées, sont fondées sur les arguments suivants :

1°. Existence d'une représentation séparée, aux niveaux anatomique et fonctionnel, de certaines catégories conceptuelles dans la mémoire.

2°. Existence d'une représentation plus complexe des catégories animées, en raison des traits et des attributs structurels (sensoriels) qui les composent, par rapport aux autres traits structurels (fonctionnels) qui caractérisent les catégories inanimées.

3° Effet de certaines variables, telles que plus grande complexité visuelle, moindre familiarité et moindre fréquence d'utilisation dans les catégories animées.

4° Importance spécifique de certaines clés visuelles, ce qui se traduit par un degré élevé de similitude structurelle dans certaines catégories inanimées et par un degré faible dans les catégories animées.

3. Modèle de Damasio.

L'idée centrale de ce modèle (Damasio, 1989,1990) est qu'il n'existe pas de système neural servant de support à certains domaines conceptuels (contrairement à ce que certaines données neuropsychologiques récentes, issues de sujets présentant des lésions focales du SNC, prétendent nous faire croire). En revanche, il existe bien des systèmes neuraux servant de support à certains ensembles distinctifs de traits, ou de caractéristiques, de nature et de statut différents, mais de portée intercatégorielle.

3.1. La théorie des plans et des niveaux hiérarchiques (Damasio, 1989, 1990)

Damasio a élaboré un modèle, fondé à la fois sur des concepts neuro-anatomiques et comportant des composantes psychologiques de type démences, les aphasies et les psychoses. Ces problèmes restent inexplicables dans le cadre des théories neurologiques et psychologiques classiques.associationniste, pour essayer d'expliquer les problèmes liés à la mémoire sémantique ou épisodique qui apparaissent dans les lésions cérébrales focales, les

3.1.1. Structures critiques.

Fondamentalement, le modèle de Damasio postule l'existence de deux types de structures critiques, du point de vue neuroanatomique et cognitif :

A. - Aires associatives primaires, monomodales :

La représentation mnésique des éléments des attributs sensori-moteurs (mais pas les attributs globaux) serait localisée au niveau du cortex associatif sensori-moteur primaire des quatre lobes. Cette représentation est dénommée "TRACES", et constituerait la représentation sensorielle monomodale (visuelle, tactile, olfactive, auditive ...) d'un fragment d'attribut. Les aires associatives, d'un point de vue neuroanatomique, assureraient le débit (ou, mieux, le "courant descendant") du flux d'information lié aux zones de convergence.

B. - Zones de convergence :

Au niveau du cortex associatif multimodal (ou "amodal"), il s'agit plutôt de "courant ascendant" du réseau cognitif. Les codes de liaison, sous forme de patterns synaptiques d'activation, y sont emmagasinés ; ils sont en principe induits par la réalité perçue - bien qu'ils puissent devenir ultérieurement indépendants de cette réalité. Sur leur base, un mécanisme de rétroactivation permet d'activer les traces des aires associatives primaires, déterminant ainsi la représentation cognitive d'un attribut donné, celle des différents attributs de l'exemplaire considéré, et celle de divers autres attributs liés à l'événement ou à la catégorie. Ces patterns synaptiques ne seraient pas en eux-mêmes (pas plus que les zones où ils figurent) dépositaires de la connaissance, mais plutôt des "pivots" de rétroactivation qui, pour provoquer l'événement cognitif, doivent réactiver les traces des zones associatives primaires ; par conséquent, la connaissance n'est pas emmagasinée sous une forme stable, mais générée à chaque acte cognitif, à partir de l'activation des codes de liaison et des traces, permettant ainsi le souvenir de la réalité, ou bien alors une recombinaison des traces sous une nouvelle forme, déterminant de nouvelles entités cognitives, non nécessairement préalablement perçues, qui constitueraient la base de la pensée créative.

3.1.2. Les différents types de zones de convergence

Damasio classe les zones de convergence en fonction de leur pattern d'activation ; il ébauche aussi un autre classement, fondé sur les niveaux hiérarchiques.

Classement en fonction du pattern d'activation

- Type 1. : Pré et rétro-activation des traces des aires associatives primaires *simultanées* ; ces aires sont fondamentales pour l'activité cognitive épisodique, et leur exemplaire paradigmatique est l'hippocampe.

- Type 2. : Pré et rétro-activation des traces *séquentielles*, et fonction d'un modèle temporel ou spatial. Ces aires sont fondamentales pour les tâches

procédurales ; leur exemplaire paradigmatique seraient les ganglions de la base et le cérébelleum.

En termes hiérarchiques, les zones de convergence ont deux niveaux principaux :

- Un premier niveau, avec les aires associatives multimodales, dans les quatre lobes.

- Un niveau supérieur, avec les aires associatives amodales, en zone temporale (Hippocampe, amygdale et cortex entorhinal) et frontale (préfrontale, dorsolatérale et orbito-médiane).

Cette hiérarchie dépend des patterns synaptiques des codes de liaison, qui, d'un point de vue cognitif, déterminent un gradient d'épisodicité, mais qui va à l'inverse des théories neurologiques classiques : les plus circonscrits correspondent à une complexité plus grande du pattern synaptique, à un niveau hiérarchique supérieur, et à une plus grande "unicité" ou "épisodicité" de la représentation cognitive ; plus l'ouverture est grande, moins le niveau hiérarchique et la complexité du pattern synaptique sont élevés, et plus grande est leur "généralité" (à l'inverse des modèles neurologiques classiques, qui tiennent pour acquis que la "généralité" des concepts est hiérarchiquement plus élevée, plus circonscrite et avec plus de complexité synaptique).

C'est ainsi que selon le modèle de Damasio, la représentation cognitive d'un objet présenté comme exemplaire unique est plus complète, hiérarchiquement supérieure, correspond à une probabilité plus grande d'avoir un pattern synaptique complexe dans une aire de convergence plus circonscrite et amodale. Enfin, la représentation cognitive des catégories supra-ordonnées de cet exemplaire - dont les patterns synaptiques des codes de liaison sont plus proches des zones multimodales - est plus ouverte.

3.1.3. *Formation de domaines*

Les codes de liaison de certains thèmes, en fonction de la fréquence d'utilisation de ces thèmes, se regroupent de manière plus ou moins constante dans les mêmes réseaux synaptiques, au niveau des zones de convergence, en se dotant de pondération plus élevée, de manière à être activés de façon plus rapide et plus aisée (par exemple, les visages, les nombres, les couleurs, les parties du corps, etc.). C'est ce que Damasio appelle formation de domaines, qui ne constituent absolument pas des "centres neurologiques régionaux des représentations", dans la mesure où pour les (re)créer, il est nécessaire de faire appel à la réactivation des traces des aires associatives primaires.

3.1.4. *Contraintes portant sur la trame*

Pour cet auteur, la constitution de la trame des codes de liaison est susceptible d'être modifiée par l'expérience, avec deux types de contraintes :

a) Neurobiologiques : dépendance - fonction des idées et du thème - du tissu neurologique de chaque individu, génétiquement déterminé.

b) De la réalité : le système neurobiologique se mettrait au contact de la réalité, avec le développement des patterns de connexions synaptiques (les codes de liaison). La complexité contextuelle se trouverait au niveau taxinomique, de sorte que, à plus grande complexité contextuelle correspondrait une plus grande "unicité" (ou "subordination") et à plus petite complexité contextuelle, une plus grande "catégorisation" (ou "supra-ordination"). Les contraintes imposées à la trame par la réalité présentent des connotations typiquement associationnistes, dans la mesure où les codes de liaison comportent des règles de similitude physique, d'emplacement spatial, de sequentialité temporelle et de fréquence d'occurrence.

4. Comparaison avec les modèles traditionnels

4.1. Neurologiques

Pour les modèles neurologiques traditionnels, les aires associatives antérieures amodales emmagasinaient la connaissance générale et abstraite, et les aires multimodales et monomodales, plus postérieures, la connaissance épisodique, établissant ainsi un gradient d'épisodicité postero-antérieur.

Ces modèles se heurtent à un certain nombre de faits usuels de la clinique neurologique : si on postule que la connaissance générale, abstraite et supraordonnée, réside dans les aires antérieures, une lésion de ces structures détruirait totalement la connaissance du monde des sujets porteur de cette lésion ; à l'inverse, une lésion plus postérieure produirait une altération cognitive plus ou moins sélectivement épisodique. Or, ces effets ainsi postulés ne se produisent absolument pas dans un certain nombre de situations cliniques connues : par exemple, dans les syndromes amnésiques purs par lésions de l'hippocampe, du thalamus, ou du lobe cingulaire (aires amodales de convergence) on observe un déficit typique de l'apprentissage et du souvenir épisodique, avec conservation du souvenir, et dans certain cas, de l'apprentissage - sémantique (ces patients sont incapables de mémoriser des événements nouveaux, et dans certains cas également incapables de les ordonner chronologiquement, mais se rappellent parfaitement leur propre identité, et "ce que sont" les choses ou "la façon de les utiliser"). A l'autre extrême du spectre clinique, on observe que des lésions du cortex pariéto-occipital gauche (aire associative paucimodale) peuvent produire le syndrome d'agnosie des couleurs, où les sujets sont incapables non seulement de nommer ou distinguer les couleurs, mais même de se rappeler la couleur

d'objets concrets représentés mentalement ... Autrement dit, une lésion postérieure fait perdre toute connaissance sémantique des couleurs.

Cependant, le modèle de Damasio, avec son gradient d'épisodicité antéro-postérieur, peut parfaitement expliquer les troubles de la mémoire épisodique avec conservation de la connaissance sémantique des syndromes amnésiques purs par lésion focale ... De plus, considérer les zones de convergence, avec leurs codes de liaison, non plus comme dépositaires, mais comme "pivots" de rétroactivation des aires associatives primaires, rendrait tout à fait compte de la perte de la connaissance sémantique des couleurs dans les agnosies des couleurs : bien que les codes de liaison soient intacts dans la zone antérieure, ils ne pourraient rétroactiver les "traces" relatives aux couleurs, du fait des lésions corticales des aires associatives primaires qui ne permettent pas à l'attribut "couleur" d'être activé, quelle que soit l'activité cognitive.

4.2. Psychologiques

Modèle de stockage sémantique :

C'est dans la mesure où le modèle de Damasio est fondé sur les "réseaux neuronaux", qu'il dépasse l'"ultra-localisationnisme" cher aux théories du stockage, tout en offrant une théorie de formation de la trame, avec ses contraintes biologiques et celles liées au milieu, théorie inexistante dans les modèles de stockage.

Modèle de la voie d'accès :

Le modèle en termes de traces et de niveaux hiérarchiques permet d'expliquer les paradoxes qui dérivent de celui en termes de voie d'accès :

- Pour expliquer les déficits "ultrasélectifs" pour certaines catégories spécifiques, ou les "doubles dissociations", on essaye classiquement d'invoquer des "canaux d'accès spécifiques", ou alors des lésions des modèles synaptiques particulièrement vulnérables que constituent le substrat des codes de liaison des concepts concrets.

- En termes de voies d'accès, les différences observées pour la même tâche (par exemple, les troubles des traitements cognitifs de la catégorie "animal", liés aux seuls stimuli verbaux, pas aux stimuli visuels), sont incompréhensibles. Par contre, avec le modèle de Damasio, elles s'expliquent facilement, en postulant une lésion des connexions entre certaines aires associatives primaires monomodales spécifiques et les zones de convergence, avec conservation de ces dernières et de leurs codes de liaison, qui pourraient être activés à partir d'autres aires associatives primaires liées à d'autres modalités sensorielles.

- Les troubles spécifiques de définition des catégories sont aussi difficiles à expliquer en termes de la théorie des voies d'accès ; ils peuvent s'expliquer, à

partir du modèle de Damasio, en postulant une atteinte des aires de convergence et de leurs codes de liaison spécifiques.

Théorie de la familiarité :

Cette théorie, peu susceptible de fournir une explication de l'ensemble des phénomènes décrits, est plus ou moins acceptée du fait de l'influence de la familiarité sur la conservation ou la perte de certains aspects de la connaissance sémantique, influence que Damasio sous tend, étant donnés les critères associationnistes de sa théorie de modification de la trame en fonction des contraintes du milieu (où figure en premier lieu familiarité et fréquence des stimuli).

5. Interprétation des syndromes neuropsychologiques classiques

Le modèle des traces et des niveaux hiérarchiques permet d'expliquer d'autres syndromes neuropsychologiques classiques, comme la prosopagnosie ou l'agnosie visuelle associative, en termes de "niveaux-hiérarchiques d'épisodicité". C'est ainsi qu'un sujet prosopagnosique qui ne reconnaît ni les visages familiers ni ses objets personnels, mais reconnaît par contre les expressions faciales et les objets génériques, traduit ainsi simplement un déficit des codes de liaison d'un niveau hiérarchique plus élevé, plus "épisodique". Les agnosies visuelles associatives, où les exemplaires individuels sont reconnus, mais où la catégorisation et la supraordination posent problème, traduisent, elles, une atteinte des codes de liaison à un niveau inférieur, avec un niveau moins élevé d'épisodicité.

Les aphasies anomiques et motrices seraient dues à l'atteinte de l'"accès au lexique" du système sémantique, ou à celle de la mémoire procédurale du langage, quant aux aphasies sensorielles, elles seraient dues à l'atteinte du système sémantique lui-même.

Enfin, le modèle justifie pleinement les données des recherches expérimentales sur les troubles de la mémoire sémantique, qui semblent passer par deux étapes : une première étape, où les déficits concernent essentiellement les exemplaires et les attributs, et une étape ultérieure, où les niveaux supraordonnés et catégoriels sont atteints. En postulant ici un gradient d'"unicité" antéro-postérieur, et une atteinte anatomo-pathologiques première de l'hippocampe (zone de convergence amodale de niveau hiérarchique supérieur), qui s'étend ensuite aux aires associatives corticales des quatre lobes (mono et multi-modaux), on peut alors logiquement expliquer la chronologie des troubles sémantiques "de l'exemplaire à la catégorie", que mettent en évidence un certain nombre d'auteurs (Schwartz et al, 1979, ...)

6. Bibliografía.

- Bayles, K.A., Tomoeda, Ch.K. y Trosset, M.W. Naming and Categorical Knowledge in Alzheimer's Disease. The Process of Semantic Memory Deterioration. *Brain and Language*, 39, 498-510, 1990.
- Chertkow, H., Bub, D. y Seidenberg, M. Priming and Semantic Memory Loss in Alzheimer's Disease. *Brain and Language*, 36, 420-446, 1989.
- Damasio, A. R. Time-locked Multiregional Retroactivation: A Systems-level Proposal for the Neural Substrates of Recall and Recognition. *Cognition*, 33, 25-62, 1989.
- Damasio, A. R. Category-related Recognition Defects as a Clue to the Neural Substrates of Knowledge. *Trends in Neuropsychology*, 13(3), 95-99, 1990.
- Gainotti, G., Silveri, M.C., Daniele, A. y Giustolisi, L. Neuroanatomical Correlates of Category-specific Semantic Disorders: A Critical Survey. *Memory*, 3(3/4), 247-264, 1995.
- Gainotti, G. y Silveri, M.C. Cognitive and Anatomical Locus of Lesion in a Patient with a Category-specific Semantic Impairment for Living Beings. *Cognitive Neuropsychology*, 13(3), 357-389, 1996.
- Hodges, J.R., Graham, N. y Patterson, K. Charting the Progression in Semantic Dementia: Implications for the Organisation of Semantic Memory. *Memory*, 3(3/4), 463-495, 1995.
- Nebes, R.D. Semantic-memory Function and Dysfunction in Alzheimer's Disease. En T. M. Hess (Ed.). *Aging and Cognition: Knowledge Organization and Utilization*, 1990.
- Peraita, H., Galeote, M.A., González Labra, M.J. et Sánchez Bernardos, M.L. EMSDTA: Batería para la evaluación de la memoria semántica en pacientes con demencia tipo Alzheimer. (En preparación). 1997.
- Sartori, G., Miozzo, M. y Job, R. Category-specific Naming Impairments? Yes. *The Quarterly Journal of Experimental Psychology*, 46A(3), 489-504, 1993.
- Seidenberg, M.S. Cognitive Neuropsychology and Language: The State of the Art. *Cognitive Neuropsychology*, 5(4), 403-426, 1988.

Shallice, T. Specialisation Within the Semantic System. *Cognitive Neuropsychology*, 5(1), 133-142, 1988.

Sheridan, J. y Humphreys, G.W. A Verbal-semantic Category-specific Recognition Impairment. *Cognitive Neuropsychology*, 10(2), 143-184, 1993.

Smith, S., Faust, M., Beeman, M., Kennedy, L. y Perry, D. A Property Level Analysis of Lexical Semantic Representation in Alzheimer's Disease. *Brain and Language*, 49, 263-279, 1995.

Modèles intermédiaires et concepts scientifiques

Nicolas GREGORI

*Laboratoire de Psychologie de l'Interaction – GRC
Université Nancy 2 – BP 33-97
54015 Nancy cedex
gregori@clsh.u-nancy.fr*

Résumé. Au cours de cet article, nous interrogeons le lien entre théorie et modèle en regardant comment des apprenants en classe de première construisent ensemble un modèle du transfert d'énergie. Nous verrons comment, au cours d'une interaction, chacun des deux tente de modéliser le montage électrique qu'ils ont sous les yeux, cela à partir de théories qu'ils élaborent implicitement et qui fondent leur raisonnement.

Mots-clés : théories implicites, analyse de l'interaction, modèles intermédiaires.

1. Introduction

Modéliser le monde environnant est une activité quotidienne. Pour tenter de comprendre ce que nous ne comprenons pas, pour essayer de percevoir ce qui se cache dans des zones qui nous sont inconnues, nous tentons, parfois vainement, d'en réduire la complexité afin de mieux l'appréhender. Ainsi, nous élaborons des modèles à partir de théories que nous avons du monde. Mais ces théories, nous ne les exprimons pas toujours, ce qui ne va pas sans problème puisque le risque encouru est alors que nous ne nous comprenions pas, faute d'une base commune. Or certaines situations nous imposent de faire ce travail de compréhension mutuelle, qu'il s'agisse de situations de travail (Grusenmeyer et Trognon, 1997) ou encore de situations de conception collective de produit (Grégori *et al.*, 1997).

Pour discuter de la question des modèles intermédiaires en construction et des théories qui les sous-tendent, nous proposons l'analyse d'une séquence conversationnelle d'une interaction d'apprentissage, analyse dont les fondements sont décrits dans Brassac (1992), Trognon et Brassac (1992) et Ghiglione et Trognon (1993). Dans cette situation, deux lycéens ont pour tâche d'accomplir une modélisation en termes de transfert énergétique du montage

électrique qu'ils ont sous les yeux. Nous verrons alors que la difficulté de réaliser cette modélisation provient tout autant du fait qu'ils ne dominent pas le concept de transfert – ce qui les entraînent, bien involontairement, à appliquer un modèle qu'ils connaissent mieux, celui de l'électrocinétique – que du manque de communication qu'ils ont à propos des théories qui soutiennent leurs raisonnements.

2. Des théories en contradiction pour l'élaboration d'un modèle du transfert énergétique

2.1. La situation

La situation que nous allons observer est la suivante¹. Deux lycéens, une fille (F) et un garçon (G), en classe de première sont face à un montage constitué d'une pile et d'une ampoule reliés par deux fils conducteurs. Ils ont pour tâche de dessiner sur ordinateur un modèle de la chaîne énergétique, alors qu'ils n'ont pas encore appris ce concept. La seule source d'information dont ils disposent est un dictionnaire électronique qu'il peuvent librement consulter sur l'ordinateur.

Au moment où débute la séquence conversationnelle (G181, qui représente la cent quatre-vingt unième intervention de G), G et F se sont mis d'accord sur le fait que la pile est un réservoir et que l'ampoule est un transformateur. Ils les ont alors représenté graphiquement à l'écran selon les conventions qui leur ont été indiquées. Dans l'extrait reproduit, on s'intéresse au transfert spécifique, de type « travail électrique » entre la pile et l'ampoule, sachant qu'après l'ampoule, la chaîne énergétique prend la forme d'un double transfert vers l'environnement (« rayonnement lumineux » et « chaleur »).

Extrait conversationnel².

1. Cette situation et le corpus conversationnel qui l'accompagne nous ont été livrés en 1995 par M. Baker, de l'université de Lyon III, et ont déjà fait l'objet d'un travail au sein du GRC ayant réuni D. Brixhe, V. Collet-Gérardin, J. De Almeida, M. Musiol et nous-même. L'analyse présentée ici reprend et complète le travail réalisé alors. Nous tenons donc à remercier nos collègues chercheurs pour la part que nous leur devons.

2. Par convention de transcription, trois points entre crochets "[...]" indiquent une coupure dans le corpus, deux points entre parenthèses "(...)" indiquent un silence d'une à deux secondes et les mentions "a", "b", "c" sur les interventions (par ex. G185a, G185b) renvoient à une seule intervention que nous avons scindée pour la pertinence de l'analyse.

- G181 donc alors maintenant les transferts, alors le transfert
[...]
- F184 en fait, y faut faire une flèche dans « sach », dans chaque sens, transfert 1, transfert 2, on fait un transfert comme ça et un transfert comme ça qu'on va nommer après
- G185a alors attends euh
- G185b alors de la pile on a un transfert
- F186 on a un transfert qui est euh
- G187 alors qui est [inaudible] par deux fils conducteurs, donc le premier fil conducteur
- F188 donc euh j'fais une flèche comme ça que j'mets, 1, transfert 1 comme ça Didier
- G189 mmh
- F190 et euh j'mets fil, premier fil conducteur ? (..) transfert 1, premier fil conducteur ? (..) et après on met transfert 2, second fil conducteur
- G191 donc mode de transfert, donc premier fil, fil conducteur tu mets en dessous
- F192 on va mettre transfert, et là on en fait un dans l'autre sens, c'est le second
- G193 ah non non non non
- F194 ben oui mais le circuit, faut bien le fermer
- G195 eh ouais mais la pile
- F196a ah oui mais y a pas d'énergie qui, y a pas de cas où en fait
- F196b l'ampoule elle produit pas d'énergie,
- F196c donc le fil qui repart à la pile, c'est juste pour fermer le circuit, c'est pas c'est pas un transfert
- G197a ouais alors attends d'accord
- G197b mais attends y a un pôle négatif, donc ça va du pôle négatif au pôle négatif et du pôle + au pôle +
- F198 non du pôle + au pôle -
- G199 c'est bien ce qui me semblait !
- F200 [rires]
- G201 je me disais que là, bon alors donc y a un truc là (..) c'est pas drôle (..) donc là
[...]
- F206 non mais regarde, y a bien un second transfert pour fermer le circuit, mais en fait c'est pas un transfert, c'est juste pour c'est pour fermer le circuit, pour que l'énergie puisse passer
- G207 attends, le courant circule de la borne positive de la pile à la borne négative de l'ampoule, mais du machin là, du socle
- F208 et ça revient du + au - ou du - au +
- G209a +, - et + au - (..)

- G209b eh ben si c'est bon, y a deux transferts
 F210 mais non y a pas deux transferts
 [...]

G213 mais en tout cas si y en a qu'un, ça marche pas je m'excuse
 F214 ben oui mais non mais tu dis que ça
 G215 ah si en fait y en a qu'un mode de transfert c'est vrai
 [...]

F222 à savoir, si j'suis bien d'accord avec toi qu'y a un deuxième fil qui ferme le circuit mais à savoir si c'est un transfert ou pas
 G223 non mais d'accord, non c'est pas un transfert
 F224 parce qu'elle te dit bien que le transfert c'est la chose qui
 G225 c'est une sorte oui c'est un mode
 F226 c'est un système (..) un transfert
 G227 il faut qu'il soit différent, en fait c'est le même, c'est un fil, si tu veux, regarde il y a
 F228 remarque non parce que regarde dans ton dico, elle te met « transfert : déplacement d'un lieu à un autre », donc on peut l'mettre
 G229 je m'excuse là c'est différent, les différents modes de transfert de l'énergie, un mode de transfert
 F230 le travail, la chaleur, le rayonnement, eh oui non mais alors à ce moment là, le rayonnement c'est un transfert, c'est un mode de transfert le rayonnement lumineux, l'ampoule
 G231 donc ensuite il y a encore une chaîne énergétique
 F232 ben je sais pas hein, ben oui parce qu'après on dit
 G233 allez allez faut se dépêcher alors si y en a une autre encore

Un premier regard sur cette séquence nous montre la difficulté qu'ont F et G de construire le modèle énergétique. Y a-t-il un ou deux transferts ? Quel en est le sens ? Et l'on soupçonne que cette difficulté provient de ce que F et G ne possèdent pas encore la théorie liée au concept de transfert énergétique. Or ils ont un problème à résoudre et, pour ce faire, ils appliquent un modèle qu'ils possèdent, celui d'électrocinétique, qui n'est, malheureusement pour eux, pas adéquat ici.

Nous allons maintenant analyser finement le processus conversationnel pour essayer de comprendre comment F et G ont résolu ce problème.

2.2. Première tentative de modélisation et émergence de la première théorie de F

Au cours de la première intervention, G introduit un nouveau thème de discussion : celui de la modélisation du transfert énergétique entre la pile et l'ampoule. On note une certaine indétermination sur le nombre de transfert, puisque « *le transfert* » peut renvoyer autant à un transfert parmi plusieurs qu'à un seul transfert.

En 184, F prend une position d'autorité et expose de façon explicite comment il faut modéliser le dispositif « pile-ampoule », comme le marque l'emploi de « *en fait, y faut faire...* ». Il y a de façon incontestable un transfert dans chaque sens entre la pile et l'ampoule, qu'elle nomme pour l'instant : « transfert 1 » et « transfert 2 ». Soient respectivement p et q les propositions selon lesquelles « IL Y A UN TRANSFERT DE LA PILE VERS L'AMPOULE » et « IL Y A UN TRANSFERT DE L'AMPOULE VERS LA PILE », on peut dire que F affirme « $p \wedge q$ ». Ce qui apparaît au tableau 1. Malgré la force de l'énonciation, le modèle proposé n'est pas validé par G.

Cette position d'autorité, F la modère par la suite (188), on va voir pourquoi. En 185, G suspend la proposition de résolution du problème que F vient d'exprimer. L'effet en est que la satisfaction de cette proposition est différée, mise en attente. Ce n'est pas un désaccord que G manifeste, il ne dit pas que la solution proposée est erronée. Simplement, il entreprend d'examiner cette résolution terme à terme en la reprenant à partir de la proposition p.

<i>séquentiel</i> ^p	<i>conversationnel</i>			
	<i>ILLOCUTOIRE</i>	<i>COGNITIF</i>		
		<i>action de F</i>	<i>action de G</i>	<i>partagé</i>

3. Dans cette colonne, nous faisons figurer une figure hiérarchique et fonctionnelle de la conversation. Les sigles « I » et « E » représentent respectivement des interventions et des échanges. Un échange prototypique est le couple question/réponse constitué donc de deux interventions. Ces interventions et échanges peuvent être subordonnés ou subordonnants selon qu'ils étaient ou sont étayés par l'échange ou l'intervention d'ordre supérieur ou inférieur. Dans la séquence analysée, l'échange (185b/186) soutient l'intervention 187 qui elle-même, en compagnie de l'échange (184/185a), soutient la déduction 188 qui est l'élément dominant de cette sous partie. On peut alors imaginer que si cette déduction était mise en cause par G, alors F aurait la ressource d'évoquer les interventions et échanges sous-jacents pour la défendre.

F184		résout le problème	$p \wedge q$		
G185a	— F —	suspend l'action de F			
G185b	— I —				
F186	— F —				
G187	— I —	teste la résolution		p	
F188	— F —	propose une action	p		p
G189	— I —	phatique			
F190	— F —	reformule la résolution	$px \wedge qy$		
G191	— I —	explicité l'action de F		px	px
F192	— F —	reformule la solution	$pa \wedge qb$		
p : il y a un transfert de la pile vers l'ampoule ; x : premier fil conducteur ; a : transfert 1					
q : il y a un transfert de l'ampoule vers la pile ; y : second fil conducteur ; a : transfert 2					

Tableau 1. *Emergence de la première théorie du transfert de F.*

En 187, G continue l'action entamée en 185, comme l'indique le connecteur « *alors* ». De plus, il fait un lien explicite entre la notion de *transfert* au départ de la pile « *qui* » et les « *deux fils conducteurs* » qui apparaissent sur le montage électrique. Par la suite, il soumet l'hypothèse qu'il vient de formuler, c'est-à-dire celle du lien entre transfert à partir de la pile et fils conducteurs, à un traitement d'évaluation « *donc* ». En somme, G opère méthodiquement. Après avoir repris la formulation de la modélisation proposée par F à partir de la pile, premier élément du montage, il initie un test de cette hypothèse en partant du premier fil conducteur. Mais cette reformulation est également un aveu. G ne reconnaît immédiatement ni la fausseté ni la vérité du modèle proposé par F. Aussi le met-il à l'épreuve de ses connaissances.

En 188, F ratifie la proposition du lien entre transfert et fil électrique faite par G. On est toujours ici dans la satisfaction de 184 qui consistait en une proposition de résolution du problème posé. On observe en effet que F reprend ses termes d'alors : « [faire] *une flèche, comme ça* [de la pile vers l'ampoule], *transfert 1* ». De plus, F interpelle G pour s'assurer de son accord sur la représentation de la flèche. Ce dernier n'émet pour toute réponse qu'un « *mmh* » dont la fonction est certainement plus de laisser F continuer son raisonnement (fonction phatique) que de satisfaire sa proposition.

F poursuit donc son raisonnement en 190. Le fait qu'il y ait un transfert de la pile vers l'ampoule (p) semble maintenant acquis. Rappelons que les élèves doivent non seulement faire figurer le ou les transferts sur le dessin, mais qu'ils doivent également le ou les nommer. C'est ce à quoi s'attelle maintenant F, en proposant de nommer le transfert entre la pile et l'ampoule « *premier fil conducteur* », que l'on note px, tandis que l'autre transfert, entre l'ampoule et la pile, sera nommé « *second fil conducteur* », qy. Première constatation, F reformule 184 en quelque sorte, puisqu'elle propose toujours de modéliser la chaîne énergétique en faisant apparaître deux transferts p et q. Mais on voit ici le rôle de l'interaction dans cette reformulation, puisque l'action de G, qui consistait à examiner terme à terme la résolution proposée, a conduit F à préciser celle-ci de telle sorte qu'elle énonce maintenant « $px \wedge qy$ ».

Dans la première partie de son intervention 191, G précise l'activité en cours. Lui et F sont désormais en train de spécifier le « *mode de transfert* », à savoir, deuxième partie de son intervention, nommer les transferts. Il satisfait alors la première partie de la proposition de modélisation faite par F, le premier fil conducteur correspond à un (mode de) transfert. De nouveau G se montre très méthodique dans la procédure qu'il met en œuvre pour résoudre le problème. D'une part, il explicite le travail en cours, chose que n'a pas faite F, d'autre part, il ne satisfait pas l'action de F dans son ensemble, mais le fait étape par étape, le premier fil d'abord. Ainsi, la proposition « $px \wedge qy$ » n'est pas intégralement satisfaite, mais uniquement « px ».

De nouveau F poursuit son raisonnement. L'acte de satisfaction réalisé par G (191) est ici validé, si bien que l'état du monde représenté dans la première partie de la proposition de F en 190 est maintenant établi comme un fait : « il y a un transfert qui va de la pile vers l'ampoule, en lien avec le premier fil conducteur » (px). Toutefois, en proposant d'étiqueter "transfert" ce fil conducteur, F fait plus que ratifier px. Elle signifie en effet qu'il y a équivalence entre transfert (a, b) et fil conducteur (x, y). C'est dans la logique du raisonnement qu'elle a tenu jusqu'à présent qu'elle poursuit en proposant d'étiqueter un second transfert dans l'autre sens, c'est-à-dire de l'ampoule vers la pile. Autrement dit, à chaque fil conducteur correspond un transfert, ce que l'on peut formaliser sous la forme $(pa \wedge pb)$. On peut alors traduire la théorie sous-jacente à une telle proposition de modélisation : *à chaque fil conducteur correspond un transfert d'énergie.*

2.3. La deuxième théorie de F et celle de G

En 193, G s'oppose à cette proposition de F, donc ne la satisfait pas. Que G interrompe F au moment précis où il le fait laisse supposer qu'il serait d'accord avec le fait d'étiqueter « transfert » le premier fil conducteur, donc qu'il assimilerait lui aussi *transfert et fil conducteur*. En revanche ce serait à propos du second transfert qu'il s'opposerait à F. Ce qui est certain, c'est que G exprime une véritable divergence avec F. Jusqu'à présent, les actions suspensives de G étaient en effet réalisées sur le mode de la reformulation alors qu'ici il s'agit d'une opposition nette "nonq" (tableau 2).

F justifie alors sa proposition (194) et de ce fait exprime la compréhension qu'elle a de la contestation de G. Celle-ci porte, d'après elle, sur le deuxième terme de son intervention 192. En effet, en rétorquant qu'il faut bien fermer le circuit (r), F désigne le second fil conducteur (de l'ampoule vers la pile) et, par suite, le second transfert, comme objet de la contestation. Ainsi, le transfert indiqué depuis la pile vers l'ampoule demeure valide, mais c'est le second transfert et lui seul qui est en discussion. F reste donc fidèle à la modélisation qu'elle propose depuis le début de la séquence : « un transfert par fil », mais est également capable de la justifier en faisant appel à des connaissances (électrocinétiques) antérieurement acquises. Autrement dit, le modèle qu'elle propose trouve sa validité dans un ensemble de connaissances qu'elle possède et qu'elle est capable d'invoquer si le besoin s'en fait sentir.

Ce qui apparaissait en filigrane depuis le début de la séquence, à savoir le fait que F et G appliquent en fait le modèle de l'électrocinétique pour modéliser le transfert d'énergie est maintenant pleinement manifeste. Dans la suite de l'interaction, on va découvrir l'émergence de la deuxième théorie de F et de celle G, toujours à propos du concept de transfert d'énergie.

G réagit de façon oppositive à la tentative de justification opérée par F, ce qu'indique l'articulation du discours à l'aide d'un « *ouais mais...* » sans contenu (195). Plutôt, avec pour seul contenu un thème de discussion : la pile. On peut simplement dire que, par cette articulation, G prépare une contre-argumentation centrée sur l'élément « pile », donc que la justification de F n'est pas satisfaisante pour G.

<i>séquentiel</i>	<i>conversationnel</i>			
	<i>ILLOCUTOIRE</i>	<i>COGNITIF</i>		
		<i>action de F</i>	<i>action de G</i>	<i>partagé</i>

F192		résout le problème	$pa \wedge qb$		
G193] R	rejette en partie la résolution		nonqb	pa
F194] T	justifie sa proposition	r
G195] T	débuté un contre-argument			
F196a] R	accepte de réviser sa position	
F196b] T	formule une nouvelle théorie			s
F196c] R	conclut	$s \rightarrow \text{non}q$
G197a] T	accepte			
G197b] R	conclut	
F198		contredit G			$u [= \text{nont}]$
G199		valide la contradiction		$u [dc q]$ <i>car nonq →</i> <i>or nont</i> <hr/> <i>non(nonq)</i>	u
p : il y a un transfert de la pile vers l'ampoule ; a : transfert 1					
q : il y a un transfert de l'ampoule vers la pile ; b : transfert 2					
r : il faut fermer le circuit					
s : l'ampoule ne produit pas d'énergie ; u : ça va du pôle positif au pôle négatif					
t : ça va du pôle positif au pôle positif et du pôle négatif au pôle négatif					

Tableau 2. Deuxième théorie du transfert d'énergie de F et théorie de G.

L'intervention suivante de F (196) est centrale dans la résolution du problème car F y propose une nouvelle théorie du transfert d'énergie qui se démarque légèrement du modèle électrocinétique. En fait, F anticipe la contre-argumentation qu'elle prête à G. Que dit-elle en effet ? Que l'ampoule ne produit pas d'énergie. Sur quelle mode le dit-elle ? Sur celui de l'opposition « *mais* ». Enfin, sur quel ton le dit-elle ? Sur celui de l'exclamation « *ah oui !* », exprimant par là qu'elle vient de découvrir, de comprendre, d'apercevoir un fait : l'ampoule ne produit pas d'énergie (proposition s). Dans l'inférence qu'elle produit consécutivement à cette constatation, F exprime un nouveau

savoir : l'ampoule ne produisant pas d'énergie, le fil conducteur qui va de l'ampoule vers la pile ne correspond pas à un transfert.

Autrement dit, il est vrai qu'il faut fermer le circuit, comme F l'a affirmé en 194, mais ce fait ne peut justifier qu'il y ait un transfert de l'ampoule vers la pile. Or c'est justement le sens du connecteur « *ouais mais* » utilisé par G dans l'entame de sa contre-argumentation. D'après Roulet (1985), « oui mais » indique que le locuteur ratifie la pertinence factuelle de la proposition contestée (il est vrai qu'il faut fermer le circuit) mais en dénégue la pertinence inférentielle (ce fait ne peut justifier la position tenue, celle de se servir de cet argument pour affirmer qu'il y a un transfert de l'ampoule vers la pile). De plus, F souligne à nouveau que l'objet du litige porte sur un éventuel second transfert, laissant toujours intact le premier depuis la pile vers l'ampoule.

Mais F n'exprime pas cela uniquement. Elle édicte également une règle permettant de décider s'il y a transfert ou non d'un élément vers un autre. Autrement dit, elle théorise le concept de transfert. Il est admis, cela a fait l'objet d'un mouvement d'intercompréhension, qu'il y a un transfert depuis la pile vers l'ampoule. Or F affirme qu'il n'y pas de transfert depuis l'ampoule vers la pile, ce qui résulte « *donc* » de ce que l'ampoule ne produit pas d'énergie. L'hypothèse de F (notée Hf) est donc la suivante :

— un transfert trouve son origine dans un élément producteur d'énergie (rappelons que pour F – comme pour G apparemment –, un transfert se propage dans un fil conducteur) ;

— si bien que l'on peut formuler cette hypothèse : *un fil conducteur propage un transfert s'il a un élément producteur d'énergie pour origine.*

F propose donc ici une nouvelle modélisation du dispositif « pile-ampoule » avec un seul transfert qui prend origine à la pile du fait qu'elle « produit de l'énergie ». Ce faisant, F produit ce que l'on pourrait appeler un *modèle intermédiaire* du concept de transfert d'énergie au sens des abstractions intermédiaires décrites par Vygotsky (1985). En effet, on s'aperçoit que F, sans réellement en prendre conscience, abandonne le modèle auquel elle recourait jusqu'alors, et qui reposait sur le savoir commun d'un élève de première à propos du sens de circulation de l'électricité, pour en construire un nouveau qui, s'il n'est en fait pas avéré, constitue un pas vers un nouveau savoir scientifique. Sans renoncer aux connaissances qu'elle a formulées au préalable, F est capable de modéliser différemment le dispositif qu'elle a sous les yeux en vertu de la nouvelle théorie qu'elle formule implicitement et à laquelle on a accès par l'analyse de la conversation.

La façon qu'a F de définir un transfert pose problème à G. Ici encore c'est une intervention centrale dans la discussion entre F et G. G va à son tour formuler implicitement une hypothèse qui ne correspond pas à celle de F. C'est donc une opposition entre F et G qui est mise à jour par l'analyse mais dont les acteurs n'ont pas conscience comme nous le verrons par la suite. G suspend la satisfaction de 196. En quelque sorte, on pourrait paraphraser le début de son intervention (197a) de la façon suivante : « acceptons ton hypothèse, il faut alors également accepter que... », la suite faisant l'objet de la deuxième partie de son intervention (197b). G marque donc son accord sur le fait que l'ampoule ne produit pas d'énergie et que le fil conducteur n'a d'autre rôle que celui de fermer le circuit. C'est sur l'inscription en filigrane de Hf, qui n'est d'ailleurs jamais explicitement formulée, donc sur le lien de causalité entre les deux, qu'il marque son désaccord.

Ce désaccord, annoncé par le connecteur oppositif « *mais* », repose sur le sens du courant électrique entre les bornes positive et négative des éléments (pile et ampoule). Sens qu'il vérifie auprès de F. C'est bien du pôle positif vers le pôle négatif que circule le courant électrique. Essayons de comprendre le sens de 197b. G attire l'attention de F sur le fait qu'il y a un pôle négatif et que « *donc ça va du pôle négatif au pôle négatif et du pôle + au pôle +* ». Or il énonce cela alors qu'il lui semble que ce n'est pas juste (199 - 201). Pourquoi donc G formule-t-il une proposition dont il affirme par la suite qu'il doutait de la véracité et pourquoi oriente-t-il la discussion sur le sens du courant électrique alors que cela pourrait paraître non pertinent ?

Une explication que l'on peut donner prend sens dans la formulation de ce qui, pour G, fait un transfert, hypothèse que l'on note Hg : *un transfert trouve son origine à la borne positive d'un élément*. En effet, sachant qu'il y a un transfert au départ de la pile, mais pas au départ de l'ampoule et sachant qu'il n'y a pas de transfert en retour de l'ampoule vers la pile, on aurait le schéma électrique suivant.

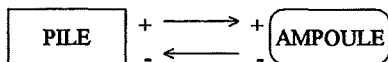


Figure 1. Schéma électrique du dispositif pile-ampoule selon la théorie de G dans le cas d'un transfert unique.

La pile, producteur d'énergie, produirait un transfert qui, selon Hg, prendrait son origine à la borne positive de la pile. Au contraire, l'ampoule, non

producteur d'énergie, ne produirait pas de transfert, si bien que le courant repartirait, selon Hg, de la borne négative de l'ampoule. Mais admettre cela, c'est admettre, comme le montre le schéma, que de la pile vers l'ampoule, l'énergie se déplacerait d'une borne positive à une autre positive et que de l'ampoule vers la pile, elle irait d'une borne négative à une autre négative. C'est justement ce fait que G vérifie auprès de F et sur lequel les deux s'accordent pour dire qu'il est faux, puisque le courant se déplace d'un pôle positif vers un pôle négatif (proposition u).

Ainsi, la question du second transfert est toujours en suspens. G a dans un premier temps (193) refusé l'action de F qui voulait nommer « transfert » le fil conducteur de l'ampoule vers la pile. Cela a amené F à reconsidérer sa théorie du transfert et à conclure qu'il n'y en a pas depuis l'ampoule vers la pile. G se trouve donc dans une situation qu'il n'arrive pas à dépasser (201), dans laquelle (i) il semble penser, désormais avec F, qu'il n'y a pas de deuxième transfert depuis l'ampoule vers la pile mais où (ii) s'il n'y a qu'un seul transfert alors cela se trouve en contradiction avec ce qu'il croit savoir du modèle du transfert d'énergie, qui dans son cas télescope celui de l'électrocinétique.

Cela nous conduit à faire la remarque suivante. G ne perçoit pas que lui et F n'ont pas la même théorie du transfert énergétique. Ce qui nous permet de l'écrire c'est que F et G se trouve dans une situation proche de celle du débat (Trognon et Larrue, 1994) avec deux positions qui s'excluent l'une l'autre. Cependant, G ne métacommunique pas à propos de cette divergence, F non plus d'ailleurs. Or s'il la percevait, il tendrait certainement à démontrer à F que le fondement de son raisonnement n'est pas valide. Autrement dit, G tenterait d'attaquer les conditions de réussite (quelles sont les raisons qui te font dire ce que tu dis ?, voir Searle et Vanderveken, 1985 ; Vanderveken, 1988) de la position de F, et inversement.

2.4. Nouvelle application de la théorie de G

Voyons comment l'interaction se poursuit. En 206, F reformule son hypothèse Hf. Elle interpelle G et focalise son attention sur l'éventuel transfert-retour de l'ampoule vers la pile. Il se passe quelque chose depuis l'ampoule vers la pile, qui permet de fermer le circuit (puisque'il faut que le circuit soit fermé, proposition r) et qui pourrait être de l'ordre du transfert, sans en être un puisque l'origine en est l'ampoule c'est-à-dire, selon Hf, un élément qui ne produit pas d'énergie, donc qui n'initie pas de transfert. En sorte, F est dans un état proche de la dissonance cognitive (Doise *et al.*, 1978) puisqu'elle accorde

à G que le fruit de leur observation commune est qu'il y a transfert entre l'ampoule et la pile (q), mais que son hypothèse Hf implique que ce transfert n'existe pas (nonq). On peut également convoquer le paradoxe de Moore en paraphrasant l'intervention de F par : « il y a un transfert entre l'ampoule et la pile, mais je ne crois pas qu'il y a un transfert ». Pourtant ce n'est pas un état pénible, ni même paradoxal que F exprime, bien au contraire. Simplement le modèle intermédiaire (F n'a bien entendu pas conscience du fait que son modèle est intermédiaire) qu'elle est en train de construire lui permet de *comprendre* qu'il n'y a pas transfert entre l'ampoule et la pile sans savoir comment le dire. C'est la proximité du modèle qu'elle construit, entre modèle électrocinétique et modèle énergétique, qui lui fait sentir qu'il y a un phénomène proche du transfert sans en être un.

Notons en outre le mode d'accomplissement de cet énoncé. Depuis qu'elle a révisé sa théorie du transfert, F soutient fermement qu'il n'y a qu'un seul transfert mais se heurte à G. Celui-ci incline également pour une modélisation avec un seul transfert, mais il est en même temps gêné par la polarité des éléments. G se trouve donc, quant à lui, en dissonance cognitive puisque ses attentes se heurtent à ce qu'il observe. F reprend alors sa démonstration sur un ton presque professorale, elle enjoint son camarade de bien suivre son raisonnement « *non mais regarde* », raisonnement que l'on peut paraphraser ainsi : il y a bien un second transfert nécessité par le besoin de fermer le circuit électrique ($r \rightarrow u$), mais cela n'implique pas qu'il s'agisse réellement d'un transfert. Autrement dit, d'après F, on observe à la fois (i) qu'il faut fermer le circuit ($r \rightarrow u$) et (ii) qu'on n'a pas de transfert depuis l'ampoule vers la pile ($r \wedge \text{non}q$).

Au cours des interventions 207 à 209a, qui forment un échange, on assiste de nouveau à une entente sur le fait que le courant circule de la borne positive de la pile vers celle négative de l'ampoule puis de celle positive de l'ampoule vers celle négative de la pile. En fait il s'agit moins là de s'accorder de nouveau à propos d'une connaissance qui a déjà fait l'objet d'une intercompréhension (proposition u) que de valider l'application de cette proposition au montage électrique auquel ils font face. Le résultat de cette validation est alors que G conclut qu'il y a deux transferts ($u \rightarrow (p \wedge q)$) (209b) puisque si le courant prend à chaque fois origine à la borne positive des éléments, alors on a, selon Hg, un transfert à chaque fois. De la sorte, G sort de son état dissonant et valide l'hypothèse Hg que nous soupçonnions depuis 197. Or comme il ne la verbalise pas et que, d'après nous, F n'entend pas cette hypothèse (pas plus que G

n'entend que F ne l'entend pas), il en résulte que Hg n'est pas discutée alors qu'elle fait problème.

En fait, F n'accepte pas la conclusion de G selon laquelle il y a deux transferts (210). Elle le dit sans argument mais on peut soutenir qu'elle la rejette parce que cela ne cadre pas avec sa théorie du transfert, bien que l'on puisse s'interroger sur sa capacité à formuler cette théorie. F et G sont donc désormais dans une situation diamétralement opposée, avec des théories implicites divergentes qui aboutissent à des modélisations qui ne concordent pas : il y a deux transferts selon G, un seul selon F.

G, en tout cas, semble sûr de son fait (213). S'il n'y a qu'un seul transfert, alors « *ça marche pas* », le courant ne circule pas ($\text{non}(p \wedge q) \rightarrow \text{non}u$). De nouveau nous avons là une trace de Hg, sans que cette hypothèse soit explicitement formulée. Malgré cela, F continue de marquer son désaccord avec G (214), c'est-à-dire avec le fait qu'il y aurait deux transferts, toujours sans argumenter. Ce manque d'argumentation nous mène à penser, avec toujours plus de conviction, que F n'interprète pas la position de G comme résultant de Hg, donc que la source de l'opposition entre elle et G n'est pas entendue. Comme nous le disions plus haut, le fait que F et G ne s'interrogent pas sur le fait qu'ils sont en conflit laisse supposer qu'ils n'entendent pas que l'autre n'a pas la même théorie du transfert, chacun croyant donc que les deux raisonnent à partir des mêmes connaissances et qu'ils cherchent à modéliser le montage d'après la même théorie.

	<i>séquentiel</i>	<i>conversationnel</i>			
		<i>ILLOCUTOIRE</i>	<i>COGNITIF</i>		
			<i>action de F</i>	<i>action de G</i>	<i>partagé</i>
F206		justifie Hf	$(r \rightarrow u) \wedge (r \wedge \text{non}q)$		
G207		interroge u			
F208		valide	$u [\rightarrow (r \wedge \text{non}q)]$		
G209a		valide		u	
G209b		conclut		$u \rightarrow (p \wedge q)$	
F210		rejette la conclusion	$\text{non}(u \rightarrow (p \wedge q))$		
G213		infère		$\text{non}(p \wedge q) \rightarrow \text{non}u$	
F214		initie un contre-			

	argument			
G215	valide le résultat de Hf		non(p \wedge q) [\wedge u]	
F222	interroge le résultat de Hf	[u \rightarrow q]?		
G223	valide le résultat de Hf		nonq	nonq
p : il y a un transfert de la pile vers l'ampoule				
q : il y a un transfert de l'ampoule vers la pile				
r : il faut fermer le circuit ; u : ça va du pôle positif au pôle négatif				

Tableau 3. *Emergence d'une entente sur un transfert unique.*

Nous sommes, avec l'intervention 215, à l'ouverture d'un nouveau temps. G se rallie soudainement à la conception de F selon laquelle il n'y a qu'un seul transfert (nonp \wedge q), il adopte son point de vue, ce qui marque un échange entre 214 et 215. Cependant, ce ralliement se fait sans explication. G affirme juste, mais avec conviction « *ah si en fait [...] c'est vrai* », c'est-à-dire avec des conditions de sincérité fortes, que F a raison.

Passons maintenant à l'intervention 222. Encore une fois, F centre le discours sur le deuxième fil qui relie l'ampoule à la pile. C'est en effet ce fil qui lui pose problème dans son hypothèse de résolution Hf puisqu'elle l'assimile à quelque chose de l'ordre du transfert mais qui n'en est pas un. Cette incapacité de F à différencier ce deuxième fil conducteur d'un éventuel transfert la mène d'ailleurs à douter de ses affirmations précédentes (196 et 206) lorsqu'elle prétendait que le deuxième fil ne correspondait pas à un transfert et ce alors même que G vient de la rejoindre dans ses conclusions. Le modèle intermédiaire du transfert d'énergie qu'elle a construit souffre de quelques faiblesses en sorte. Le ralliement sans argument opéré par G et le fait qu'elle sent peut-être, sans l'exprimer, que elle et G ne se réfèrent pas au même monde en seraient-il à l'origine ? Nous en faisons l'hypothèse mais sans pouvoir le démontrer.

Ce qui est certain, c'est que si F centre le discours sur le deuxième fil électrique c'est aussi parce que c'est cela qui pose problème pour G. Si F n'a pas entendu Hg, elle a en revanche entendu que G manque d'assurance à qualifier ce deuxième fil. G affirme alors de nouveau qu'il n'y a qu'un seul transfert, ou plus précisément que le deuxième fil, reliant l'ampoule à la pile, ne donne

pas lieu à un transfert (nonq). G se trouve donc engagé sur la révision de son jugement fondé sur Hg, mais là encore sans argumentation.

Nous avons tout de même une hypothèse. On peut remarquer que lorsque G s'est opposé à la modélisation que proposait F (193) avec un deuxième transfert en retour depuis l'ampoule vers la pile, c'était après avoir centré le discours sur le thème *mode de transfert*. Par la suite G et F ont discuté de *transfert* jusqu'à l'intervention 215 et l'exclamation de G qui change subitement d'avis, sans raison exprimée, et déclare que « *si en fait c'est vrai y en a qu'un mode de transfert c'est vrai* ». L'on sent ici que l'intonation nous aiderait certainement pour valider ou non cette hypothèse. Quoi qu'il en soit la coïncidence est remarquable, même si nous ne pouvons, dans la mesure des données dont nous disposons, affirmer sa validité. Un autre élément toutefois irait dans ce sens. On peut imaginer que si ce revirement de situation ne provient pas du remplacement thématique que nous venons de décrire, alors c'est qu'il provient d'un changement théorique. Bien sûr, on ne peut attester que celui-ci n'a pas eu lieu, mais on remarque qu'aucun indice verbal ne suppose qu'un tel changement ait été réalisé. Or on peut constater que si le ralliement au résultat « un seul transfert » est issu d'un remplacement thématique, alors il n'est pas nécessaire à G de revoir sa théorie du transfert énergétique. Il lui suffit de différencier *transfert* et *mode de transfert*.

2.5. La confrontation des modèles et leur surpassement

Une action conjointe va permettre aux deux apprenants de surpasser leur opposition. En revoyant leurs positions respectives, ils peuvent entrevoir la solution. Nous allons le voir rapidement.

De 224 à 227, F et G essaient de définir ce qu'est un transfert. En fait, ils ne sont toujours pas parvenus à s'entendre et recourent, sous l'impulsion de F (224), à un élément extérieur (*elle*, renvoyant au professeur) pour s'accorder sur cette notion. On note que G est tout aussi embarrassé que F pour dénommer le deuxième fil. F disait en 206 qu'il y avait bien un deuxième transfert (le deuxième fil) mais que ce n'était pas un transfert. En 227, G dit à propos du transfert « *qu'il faut qu'il soit différent [qu'] en fait c'est le même, c'est un fil* ». La confusion entre courant électrique et transfert d'énergie n'est donc absolument pas levée. Elle est pressentie, ce que montre ces deux interventions, mais n'est pas être exprimée.

En quelque sorte F et G sont là en train de glisser du modèle électrocinétique vers celui énergétique, mais sans disposer des concepts permettant de quitter l'un pour l'autre.

En 228, F fait de nouveau appel à un élément extérieur, le dictionnaire de l'ordinateur, pour définir la notion de transfert. Qu'observe-t-on alors ? Que ce "juge de paix" est lui-même porteur d'ambiguïté puisque F infère de ce qu'elle y trouve que l'on pourrait nommer « transfert » le deuxième fil. G reste quant à lui engagé dans sa croyance qu'il n'y a qu'un seul transfert malgré la définition du dictionnaire donnée par F. Son expressif « *je m'excuse* » est là qui montre que G a quelque raison de ne pas suivre F dans son activité inférentielle. D'ailleurs, ce n'est pas que F se trompe mais que « *là c'est différent* ». G évoque alors le fait qu'il y a plusieurs *modes de transfert* de l'énergie. C'est la première fois que la diversité des modes de transfert est évoquée puis exprimée (230). Cela nous conforte quelque peu dans l'hypothèse que nous avons formulée à propos du changement de position non expliqué de G en 215, hypothèse qui veut que ce changement provienne non pas d'une révision de sa théorie du transfert énergétique, mais d'un remplacement thématique.

A la suite de 230, F cite les différents modes de transfert. Elle semble alors découvrir (« *eh oui* ») que le rayonnement lumineux est un type de transfert. Puis G en conclut (« *donc* ») qu'il y a « *ensuite encore une chaîne énergétique* ». Un nouveau savoir s'installe alors. F et G, en centrant le discours sur la définition de ce qu'est un transfert, viennent de s'accorder sur le fait que les transferts sont de divers types : travail, rayonnement, chaleur. Nous ne sommes plus ici dans une histoire de polarité des éléments de la chaîne énergétique, mais bien dans la notion de transfert. Et la chaîne énergétique qu'ils viennent de découvrir prend son origine à l'ampoule et est de type « rayonnement lumineux ».

A la fin de cette interaction, nous ne savons pas si F et G se sont accordés sur le fait qu'il n'y a qu'un seul transfert entre pile et ampoule, ni s'ils ont été capable de nommer ce transfert, transfert de type « travail électrique ». On ne peut pas dire qu'ils ont construit une représentation partagée au sens de Grusemeyer et Trognon (1997). Nous ne savons pas non plus s'ils comprennent que transfert d'énergie et électrocinétique sont deux notions différentes. Nous ne savons pas si G a révisé Hg ni si F s'est rendu compte de la fausseté de Hf. En revanche, nous savons que F et G sont capables de décrire et de nommer un transfert au départ de l'ampoule de type rayonnement. En fait, il y en a un autre, de type « chaleur », mais c'est une autre histoire.

3. Conclusion

On a vu dans l'analyse de cette interaction comment des apprenants ont résolu un problème de transfert d'énergie qu'ils ne maîtrisent pas encore. Les diverses tentatives de résolution qu'ils ont proposées étaient toutes inscrites dans un modèle qu'ils connaissent, celui de l'électrocinétique. Ils auraient pu être d'accord sur une modélisation, valide ou non, mais ce ne fut pas le cas, si bien qu'ils ont été amenés à argumenter, ce qui nous a permis d'extraire les théories implicites sur lesquels reposaient les arguments.

Le travail d'analyse que nous avons réalisé permet de faire observer que les modèles scientifiques que nous produisons ont pour fonction première d'aider à parler d'un monde encore inconnu et que nous voudrions connaître en vue de l'assimiler (au sens piagétien du terme) à du déjà connu. Ainsi, les modèles sont des instruments cognitifs dont la destination est de nous permettre de comprendre et analyser le monde environnant en le réduisant et qui sont le fruit d'un processus d'objectivation.

Mais ce qui nous a intéressé plus particulièrement ici, c'est l'espace dynamique dans lequel les apprenants ont réalisé leur modélisation. Comme l'écrit Latour (1989), il faut s'attacher à reconstruire l'histoire dans laquelle la science s'établit. Bien sûr, le projet des deux apprenants n'était pas de faire avancer "la" science mais leur science. Le processus est cependant équivalent. F et G avaient l'obligation de s'entendre pour accomplir leur exercice de modélisation. Ils ont alors proposé des solutions et, devant l'échec qu'ils constataient, ont fait appel à des théories implicites pour justifier leurs dires. Mais les théories là encore étaient insuffisantes, et finalement, c'est en adoptant un regard extérieur, celui du professeur ou du dictionnaire, qu'ils ont pu sortir de l'impasse. Autrement dit, c'est en décentrant leur position, c'est en échappant au modèle électrocinétique qui était alors le seul qu'ils connaissaient et qu'ils ont repris sans en avoir conscience, qu'ils ont pu entrevoir une nouvelle réalité et adopter un nouveau modèle.

Certes, on ne peut pas dire qu'à l'issue de l'expérience les deux apprenants maîtrisent le concept de transfert. Cependant, on a vu que F dans un premier temps, puis les deux ont construits des sortes de modèles intermédiaires mi-naturels, mi-scientifiques qui nous font penser que désormais F et G sont en quelque sorte préparés à aborder le concept de transfert dans sa complexité et donc qu'ils sont peut-être plus à même d'en réaliser des modèles valides.

Nous concluons sur l'importance qu'il y a de construire une théorie solide pour modéliser un concept, et de l'explicitier au mieux. F et G ont pêché sur ces deux aspects et ont alors eu le plus grand mal, si jamais ils y sont parvenus, à modéliser le montage électrique en termes de transfert.

Bibliographie

- Baker, M. (1996). Argumentation et co-construction des connaissances. *Interaction et cognitions*, vol. 1(2-3), 157-191.
- Brassac, Ch. (1992). Analyse de conversations et théorie des actes de langage. *Cahiers de linguistique française*, 13, 62-75.
- Doise, W., Deschamps, J.C., Mugny, G. (1978). *Psychologie sociale expérimentale*. A. Colin, Paris.
- Ghiglione, R. et Trognon, A. (1993). *Où va la pragmatique ? De la pragmatique à la psychologie sociale*. PUG, Grenoble.
- Grégori, N., Blanco, E., Brassac, Ch., Garro, O. (1997). Analyse de la distribution en conception par la dynamique des objets intermédiaires. *Actes de la cinquième table ronde sur la conception « Les objets en conception », 01DESIGN'97*. Théoule-sur-Mer, 24-26 septembre 1997, pp. 121-134.
- Grusenmeyer, C. et Trognon A. (1997). Les mécanismes coopératifs en jeu dans les communications de travail : un cadre méthodologique. *Le travail humain*, 60/1, 5-31.
- Latour, B. (1989). *La science en action*. La Découverte, Paris.
- Piaget, J. (1964). *Six études de psychologie*. Denoël, Paris.
- Roulet, E. [Ed.] (1985). *L'articulation du discours en français contemporain*. P. Lang, Berne.
- Searle, J. et Vanderveken, D. (1985). *Foundations of illocutionary logic*. CUP, Cambridge.
- Trognon, A. et Brassac, Ch. (1992). L'enchaînement conversationnel. *Cahiers de linguistique française*, 13, 76-107.
- Trognon, A. et Larrue, J. (1994). *Pragmatique du discours politique*. A. Colin, Paris.
- Vanderveken, D. (1988). *Les actes de discours*. Mardaga, Bruxelles.
- Vygotsky, L. (1985). *Langage et pensée*. Editions sociales, Paris.

Point de vue linguistique sur la mod lisation

Introduction

L'enseignement de la physique met aujourd'hui l'accent sur l'int r t des t ches de mod lisation. La didactique des sciences exp rimentales et en particulier celle de la physique, privil gie une approche de la mod lisation dans laquelle on fournit a priori aux apprenants un mod le, qui sans  tre complet du point de vue de la discipline, est suffisant vis   vis de la t che propos e¹. L'introduction de ce type de mod le, au niveau du secondaire en tout cas, se fait essentiellement gr ce   des textes r dig s en langue naturelle dans un manuel ou dans l'enseignement du professeur.

Il semble donc prometteur, si l'on d sire pr ciser le contenu r ellement apport  par le mod le enseign , de se pencher sur le fonctionnement du texte, et en particulier sur les connaissances v hicul es par la langue. Il semble  galement n cessaire d'analyser de quelle mani re s'entrem lent les d finitions fournies et les connaissances langagi res, d'analyser comment le fonctionnement propre de la langue contraint l'utilisation du mod le.

Dans cette contribution, l'activit  cognitive d'une t che de mod lisation particuli re ayant trait au concept d' nergie, est donc  tudi e d'un point de vue linguistique. L'approche se concentre ici sur l'analyse de l'influence des structures et des contenus associ s aux items lexicaux, tandis que les aspects dialogiques et pragmatiques, relevant d'un autre cadre th orique, sont volontairement laiss s de c t . La dimension de communication n'est pas abord e de front, mais appar it au travers de la n cessit  de proposer une solution commune au probl me pos . Les mots cl s du mod le physique enseign  sont donc l'objet d'une  tude s mantique et syntactico-s mantique, puis des r gles linguistiques sont propos es pour d crire les associations entre les termes du mod le et les entit s du monde r el. Ces r gles traduisent le fait que les hypoth ses  nonc es par les apprenants doivent pr senter une coh sion s mantique minimale. Les t ches  tudi es sont consid r es par leurs concepteurs comme prototypiques, et l'on peut consid rer qu'  certains  gards, elles sont proches de toutes les situations de mod lisation en sciences exp rimentales. Les r sultats de cette  tude ont donc une port e qui d passe le cadre strict de situations d'enseignement.

¹ [Tiberghien et Megalakaki, 1994].

Proposant à la didactique et d'une manière plus générale aux sciences expérimentales un point de vue et une méthodologie d'analyse linguistiques sur le contenu et sur le fonctionnement d'un modèle énoncé en langue naturelle, cette contribution recoupe plusieurs des questions que nous nous poserons. Elle apporte en effet des éléments de réflexion sur la création des modèles, et éclaire certains aspects de leurs limites et de leur fonctionnement.

1. Modélisation en physique, didactique et modélisation

Les conceptions actuelles de la physique affirment une irréductible différence de nature entre les concepts d'un modèle et les objets de la réalité, et l'on admet dans certaines limites, que la théorie représente la réalité physique [Tiberghien, 1993]. Les rapports qu'entretiennent alors le monde de la physique et le monde réel mettent en avant la notion de modèle.

La notion de modèle donnée par Planck [Dans Halbwachs, 1971, p 96] distingue tout d'abord un monde réel, un monde d'objets existants où les transformations physiques arrivent, d'un monde de la physique, monde de concepts pensés où fonctionnent les relations mathématiques construites par la physique. Enfin, il y a le monde sensible, fait d'images perçues, où se situent en particulier les expériences et les mesures de laboratoire. La fonction de la physique est de rendre compte des lois qui règnent dans le monde sensible.

Cette analyse épistémologique conduit à considérer qu'en physique, l'interprétation et la prédiction impliquent un processus de modélisation qui met en jeu trois niveaux : la théorie, le modèle, et le champ expérimental de référence.

1.1 Le rôle assigné à la modélisation par la didactique

Le passage du champ expérimental au monde théorie - modèle constitue un des enjeux majeurs de l'apprentissage de la physique. Partant des difficultés soulevées par l'approche piagétienne, Tiberghien souligne la nécessité d'une approche théorique permettant de dégager des invariants du fonctionnement cognitif de l'apprenant, et affirme que l'analyse fondée sur la modélisation du monde matériel peut jouer ce rôle. Plusieurs hypothèses sont alors avancées quant à l'activité cognitive. La première postule que lorsqu'il interprète une situation matérielle, l'apprenant construit un modèle de la situation ; ce qui implique qu'à l'instar du physicien, il fasse une sélection des objets et des événements qui lui semblent pertinents. Sous-jacente à cette première hypothèse est l'affirmation de la cohérence de l'apprenant. La causalité se voit

reconnaître une fonction cruciale dans la recherche de cette cohérence, elle est considérée comme faisant partie de la théorie de l'élève².

1.2 Les choix didactiques du modèle énergétique :

Le cycle d'enseignement expérimental analysé ici prend pour domaine d'application la modélisation de phénomènes mettant en jeu l'énergie. Les élèves travaillant en dyades doivent établir une relation entre un modèle de l'énergie dont un "germe" leur est donné (figure 1) et les événements qu'ils observent. Ils doivent réaliser une expérience, et construire ensemble une "chaîne énergétique" traduisant dans les termes du modèle les phénomènes observés. La verbalisation est donc le moyen essentiel de prise d'information et d'échange. Le modèle est introduit sous forme textuelle, et comprend la définition des entités théoriques, le principe de conservation de l'énergie, et des règles de cohérence. Le champ de référence contient explicitement pour chaque situation les objets et les phénomènes observables de l'expérience, et implicitement d'autres éléments qui y participent indirectement. Les situations étudiées présentent un aspect très général ; les conclusions obtenues sont donc susceptibles d'outrepasser le cas du modèle de l'énergie, d'outrepasser également le cas des situations d'apprentissage.




<p>L'énergie peut être caractérisée Par ses propriétés:</p> <ul style="list-style-type: none"> -Stockage -Transformation -Transfert <p>Les différents modes de transfert de l'énergie d'un système à un autre sont: le travail, la chaleur, le rayonnement.</p> <p>Par un principe fondamental de conservation :</p> <p>L'énergie se conserve quels que soient les transformations qu'elle subit, ses transferts, et ses formes de stockage.</p>	<p>Pour construire une chaîne énergétique, il faut utiliser les symboles dessinés, et tenir compte des règles suivantes:</p> <p>Réservoir </p> <p>Transfert </p> <p>Transformateur </p> <p>Une chaîne énergétique complète commence et se termine par un réservoir. On indique sous chaque rectangle l'objet correspondant. On indique sous chaque flèche le (ou les) mode de transfert.</p>
---	---

Figure 1 : Un "germe" de modèle : le cas du modèle énergétique

2 [Tiberghien et Megalakaki, 1994, p 6].

2. Point de vue linguistique sur le modèle enseigné

2.1 Les connaissances disponibles et les stratégies suggérées

Le point de vue linguistique adopté ici conduit à mettre l'accent sur les contenus liés aux lexèmes employés dans les définitions, et au fonctionnement de la langue.

Le modèle introduit le mot **énergie**, et le caractérise par trois propriétés et un principe fondamental. Cette définition fait appel aux trois lexèmes **réservoir**, **transformation** et **transfert**. Ces mots ne sont pas redéfinis, mais simplement mis en relation avec le mot **énergie**. On donne également une définition en extension du **mode-de-transfert**, qui pourra dénoter le **travail**, la **chaleur**, le **rayonnement**. Le mot **chaleur**, en revanche, peut fort bien être pris au sens courant dans le cadre de la tâche, puisque le texte ne propose pas de définition théorique. Quant au mot **rayonnement**, il se comporte de manière intermédiaire, et renverra toujours à la lumière visible, aucune définition physique n'étant avancée par le texte. Il y a donc une grande hétérogénéité entre les termes introduits, vis-à-vis du domaine d'emploi comme vis-à-vis de la tâche elle-même. L'intrication entre langue naturelle et langue de spécialité est fort complexe, puisque l'utilisation du sens ordinaire des mots peut être nécessaire ou au contraire exclue, et ce de manière implicite.

2.2 La laxité des termes et les connaissances afférentes

Les principes s'appliquant à une entité théorique peuvent entrer en concurrence avec le comportement de ces entités dans d'autres domaines chaque fois que le mot utilisé pour les nommer appartient aussi à la langue courante, situation fréquente en physique. Il ne s'agit pas là d'admettre cette interaction de significations comme un mal, mais on peut au contraire considérer avec Gaudin que les exigences de la formalisation conduisent à une impasse, et qu'une certaine « laxité » est indispensable³. Cette interaction entre significations provenant de domaines divers ne doit pas être considérée seulement comme une cause d'imprécision, mais aussi comme une source de découverte.

2.3 La définition métaphorique de l'énergie

Le germe de modèle tel qu'il est présenté aux élèves fonde leur compréhension du concept d'énergie sur une métaphore fortement marquée

³ [Gaudin, 1993].

par la propriété de stockage dans un réservoir. Il est vraisemblable que cette image, tout en les aidant à construire un sens du modèle énergétique, apporte par afférence des propriétés plus ou moins indésirables et non contrôlables, comme cela est le cas pour d'autres métaphores utilisées dans l'enseignement de la physique.

D'une manière générale, les connaissances véhiculées par un modèle physique ne se limitent donc pas aux définitions visibles, mais elles introduisent inévitablement, par le biais des mots employés, d'autres connaissances de type linguistique, extérieures à la théorie physique. Certaines de ces connaissances peuvent jouer un rôle "parasite", et ajouter à la définition donnée des significations non prévues. D'autres sont indispensables, comme celles qui sont liées à d'autres domaines sémantiques des mots utilisés, ainsi qu'à leurs caractéristiques syntaxiques et distributionnelles.

2.4 Point de vue linguistique sur les connaissances référentielles

Pour réaliser les associations entre entités théoriques du modèle et éléments du champ matériel réclamées par la modélisation, l'élève mobilise sa connaissance des objets. Dans une approche sémantique, il semble donc raisonnable d'ajouter au signifié des mots, proche du sens défini par Frege [dans Hacking, 1989], un sens référentiel, tenant compte des propriétés connues des objets dénotés.

Notons encore que du point de vue de l'expérimentation scientifique, deux types de référents pourraient être distingués, selon qu'il s'agit d'une expérience de routine ou d'une expérience innovante et peut être selon le degré d'expertise du manipulateur. Lorsqu'une expérience de routine se déroule sans anicroche, les éléments référentiels invoqués sont des éléments génériques. En revanche, si l'expérience est nouvelle ou ne fonctionne pas, interviennent alors des connaissances liées au référent occurrence. ("C'est cette fameuse pile qui est presque à plat "...)

Notons enfin que le rôle d'expérimentateur incite à un point de vue réaliste à propos des entités théoriques et des phénomènes observés, c'est à dire à considérer que ces entités existent vraiment, ce qui conduit généralement à leur attribuer un pouvoir de causalité.

3. Cadre théorique de l'analyse linguistique

Le point de vue adopté ici, celui de l'influence du fonctionnement propre de la langue sur les mécanismes cognitifs par l'entremise de la production d'énoncés, doit s'intéresser à trois aspects : l'énonciation initiale d'une

hypothèse, l'acceptation (ou le rejet) de cette hypothèse par l'allocutaire, et enfin sa résistance aux éventuelles objections ultérieures.

3.1 *La dicibilité, condition de l'énonciation d'une hypothèse*

Il faut donc analyser la capacité d'une hypothèse à surmonter chacun de ces obstacles en privilégiant la notion de « dicibilité » en langue, indépendamment de la valeur de vérité et des connaissances relatives à la discipline. Cette notion recouvre la possibilité d'énoncer l'hypothèse, et la plausibilité de l'énoncé obtenu au regard des règles de la langue. Il paraît raisonnable d'avancer que dans les trois aspects mentionnés, la dicibilité constitue sinon une condition suffisante, tout au moins une condition nécessaire ; il est clair en effet qu'une hypothèse incapable de se plier aux contraintes de la langue semblera a priori fort peu plausible.

Un premier niveau de cohérence de l'énoncé constituant une hypothèse peut alors être cerné à l'aide du concept d'isotopie. L'isotopie, ou "continuité sémantique à travers une séquence" est conçue, selon Rastier, comme la condition de toute lisibilité ; elle est une condition de la dicibilité nécessaire à la cohésion et à la cohérence, dont la nécessité a été reconnue plus haut⁴.

3.2 *L'analyse sémique est nécessaire à la recherche d'isotopies*

La recherche des divers niveaux d'isotopie sous-tendant la cohérence des protocoles constitue donc la pierre angulaire de cette étude. Dans la situation que nous étudions, la détection et la classification des isotopies s'appuiera sur une description sémique non seulement précise, mais caractéristique des apprenants observés et non pas seulement théorique⁵.

Afin de rendre les plus rigoureuses possibles la recherche des isotopies afférentes, et l'analyse de l'influence de normes socialisées, je propose ici une représentation dynamique de la description sémique. Cette représentation utilise un cadre théorique laissant place aux phénomènes de subduction et de polysémie, et permet de préciser la notion d'équivalence entre sémèmes. Ce cadre théorique, emprunté à J. Picoche, s'appuie sur les travaux de G. Guillaume, et introduit les concepts de *signifié de puissance* et de *cinétisme*, bien adaptés à la description dynamique recherchée.

⁴Berrendonner assimile quant à lui le concept d'isotopie à la "bonne formation sémantique d'une séquence".

⁵[Rastier, 1987, p 102].

3.3 Analyse de syntaxe structurale et recherche d'isotopies

La structure syntaxique des verbes employés dans le texte du modèle, et en particulier leur valence⁶, peut refléter les caractéristiques physiques structurelles et fonctionnelles de certains objets, tandis que les propriétés sémantiques des actants précisent la manière dont se fera l'actualisation. Une étude syntactico-sémantique sera donc indispensable à la recherche des isotopies.

3.4 Vers des règles descriptives

Dans l'approche choisie, c'est la connaissance préalable d'une description sémique et syntactico-sémantique des items linguistiques qui guide la recherche des isotopies. Il convient alors de dégager les éléments systématiques participant à ces isotopies, afin d'établir des règles formelles capables de rendre compte des mécanismes réellement observés, capables aussi d'atteindre un niveau de prédiction dont la validation s'effectuera ultérieurement sur des protocoles différents de ceux qui auront conduit aux premières hypothèses.

4. Méthodologie

Tout d'abord, l'activité langagière manifestée dans les protocoles est l'objet d'une analyse préalable. A partir des hypothèses issues de cette analyse, il devient possible de préciser les données langagières nécessaires à la mise au point des règles recherchées, puis de procéder à la définition d'une enquête destinée à recueillir ces données. L'analyse des résultats de ces enquêtes permet alors d'établir les descriptions sémantiques et syntactico-sémantiques nécessaires. Les règles linguistiques proposées sont fondées sur ces descriptions et sur les hypothèses initiales. Enfin, ces règles reçoivent une validation par application à de nouveaux protocoles.

5. Règles d'association

L'accomplissement de la tâche pouvant être ramenée à un ensemble d'appariements entre entités du modèle et objets ou phénomènes du mode sensible, la distinction la plus macroscopique conduit à distinguer ceux qui sont à l'origine de raisonnements plus élaborés, de ceux qui apparaissent au contraire comme le résultat de raisonnements achevés. Il convient ensuite de préciser quels types d'associations relèvent d'une description linguistique,

⁶ [Tesièrè, 1965].

d'étudier sur quels fonctionnements de la langue elles se fondent, et de déterminer enfin si certaines d'entre elles sont suffisamment simples et stables pour donner lieu à des règles.

Une première analyse des dialogues révèle que pour la plupart des binômes, certains appariements se produisent avant les autres types de raisonnements : ce sont généralement des appariements terme à terme. Elle révèle également la stabilité de ces appariements au cours des raisonnements ultérieurs. Ces associations sont nommées ici : « appariements élémentaires ». Du point de vue linguistique, elles consistent en des opérations de rapprochement ou d'opposition ; elles peuvent être établies à partir de ressemblances sémiques qui les rendent plausibles.

D'autres appariements, qui mettent en jeu plusieurs éléments de chaque monde, sont plus tardifs, et donnent lieu à une importante activité langagière, à des argumentations, à des négociations ; ils peuvent s'appuyer sur la structure syntaxique de l'un des mots du modèle, et sur les contraintes sémantiques qui pèsent sur cette structure dans le fonctionnement usuel de la langue. Il sont considérés ici comme des « appariements de structure ».

Dans les deux cas, les mécanismes associatifs utilisés sont des mécanismes très souples, permettant d'œuvrer à partir de connaissances incertaines et incomplètes, de proposer l'appartenance d'entités à des catégories mal maîtrisées, ne possédant pas encore de référents. Ces mécanismes s'appuient davantage sur des ressemblances que sur des conditions parfaitement définies, et conduisent donc à rechercher un ensemble de règles plausibles dont aucune ne joue le rôle de règle nécessaire et suffisante d'appartenance.

Il existe également des associations qui prennent place à la suite d'une appréhension globale du modèle physique ou de l'expérience réalisée, et qui sont la conséquence de raisonnements plus achevés, de stratégies de vérification, voire de complémentations pragmatiques. Les métaphores susceptibles de traverser la tâche constituent également des mécanismes globaux.

5.1 Les appariements élémentaires

Ce type d'appariement se manifeste par un énoncé associant le lexème qui désigne une entité du modèle physique avec celui qui désigne un objet ou un phénomène physique dans la langue. Un prototype de ces associations est l'appariement [réservoir - pile] des expériences « Pile ampoule » :

Dialogue F & L sur Pile - ampoule

22 F [...] Le réservoir stocke l'énergie, donc c'est la pile. [...]

Je propose l'hypothèse selon laquelle l'origine de ces associations tient à une *parenté sémantique élémentaire* entre le lexème désignant l'entité du modèle et celui désignant l'objet ou le phénomène pressenti pour l'appariement. La description sémantique choisie permet d'introduire une gradation dans la plausibilité des appariements.

L'étude menée permet de proposer la règle suivante pour rendre compte de l'existence des appariements élémentaires :

Dans les tâches de modélisation observées, un appariement élémentaire entre une entité E du modèle physique et un objet ou un phénomène O du monde sensible, pourra se produire si les propriétés sémantiques de O lui confèrent une parenté sémantique élémentaire avec E.

La parenté sémantique étant définie dans le cadre théorique des « signifiés de puissance » et des « cinétismes » de O et E, extraits de l'analyse linguistique.

La règle ainsi formalisée n'est pas déterministe, mais permet de rendre compte des associations possibles et de leur degré de confiance. Elle concorde bien avec la stratégie générale des élèves, qui consiste à envisager les associations plausibles sans rechercher de preuves, et à multiplier les vérifications à l'aide de raisonnements d'un niveau plus global.

5.2 Les appariements de structure

De telles associations se construisent à partir d'entités du modèle théorique possédant une structure syntaxique riche d'au moins deux actants⁷. Elles se manifestent dans les dialogues par une activité de reformulation caractéristique, organisée autour du lexème qui désigne l'entité théorique dont elle explore assez systématiquement les valences.

A titre d'exemple, l'analyse syntactico-sémantique du verbe « transformer » fait apparaître deux places quasi indispensables : un « état initial », et un « état final ». Le court extrait ci-dessous met bien en évidence la nécessité d'une recherche de ces deux états d'une transformation :

Dialogue P & F sur l'expérience du Poids qui monte

⁷ [Tesnière, 1965].

- 385 P Non, mais le transformateur, i(l) faut que ça transforme quelque chose en, en autre chose...
- 386 F Mmm...
- 387 P Donc c'est... ça transforme le déplacement. C'est quoi qui transforme le déplacement ? Eh ?
- 388 F On met déplacement ?
- 389 P C'est la pente ? [...]

Lors de la découverte d'un appariement de structure, on voit se construire des phrases dont l'armature est constituée par l'une des formes syntaxiques possibles pour l'entité du modèle. L'adéquation aux contraintes sémantiques est également recherchée, avec la mise en évidence de la différence entre état initial et état final grâce à l'utilisation de l'archilème "chose". L'activité consiste en fait à rechercher des propriétés des objets ou des phénomènes susceptibles de saturer les places disponibles dans la structure syntaxique d'une entité du modèle.

On observe une convergence parfaite entre les structures associées aux lexèmes du modèle⁸, et les structures des phrases produites lors des activités de reformulation. On peut donc considérer que le mécanisme cognitif mis en œuvre s'appuie en partie sur la langue, et met à profit toute la connaissance portée par les structures syntaxiques et par les relations sémantiques qui contraignent les actants à l'intérieur de ces structures. La nécessité d'une isotopie se traduit alors par la recherche d'un isomorphisme entre la structure en langue d'un terme du modèle, et la structure possible des éléments de l'expérience ; et l'on peut dire que la contrainte de dicibilité guide la structuration du monde sensible (figure 3). On peut donc mettre en parallèle la structure des connaissances langagières et le monde réel où se construit une organisation des relations entre objets et phénomènes.

La connaissance des structures des mots du modèle peut alors être considérée comme un outil de description et de prévision de cette organisation. On peut également faire l'hypothèse que ces structures influent sur les processus cognitifs mis en œuvre. Je propose donc de formaliser comme suit une règle décrivant l'apparition de ce type de processus.

Un appariement structuré se produit entre une entité E du modèle physique présentant une structure syntaxique, et un objet ou un phénomène O du monde sensible, si les relations que peut

⁸ Telles qu'elles apparaissent au travers des enquêtes linguistiques effectuées pour cette étude.

entretenir O avec d'autres entités sensibles en font le nœud d'une structure isomorphe à la structure syntaxique de E.

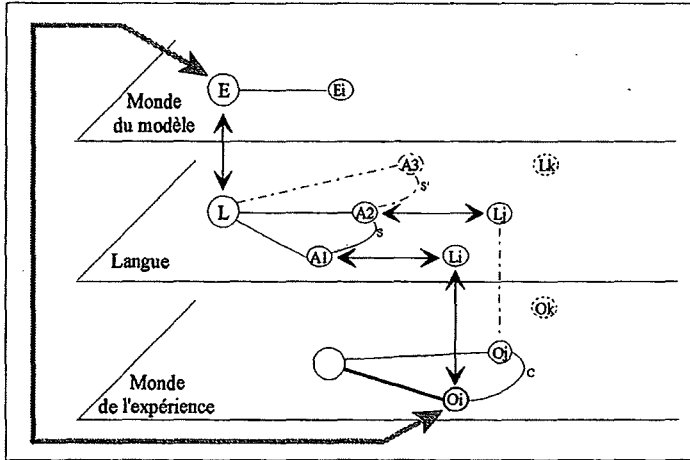


Figure 3 : La structure syntaxique guide la structuration du monde de l'expérience

La flèche en gris figure l'appariement entre l'entité E du modèle, et l'objet Oi de l'expérience, par l'intermédiaire du lexème L dont la structure se projette dans le monde sensible.

L et Li sont les lexèmes désignant respectivement E et Oi.

S et S' figurent les contraintes sémantiques entre les actants A1, A2, etc. Et C une contrainte de cohérence entre objets du monde sensible

5.3 La saturation de structure

Le phénomène que je nommerai « saturation de structure » se développe à partir d'un appariement déjà pressenti, voire établi, dont il constitue un prolongement. Mettant à profit la structure associée à une entité du modèle, il conduit à rattacher de nouveaux éléments du monde sensible à une organisation déjà amorcée. C'est un mécanisme extrêmement puissant, où le fonctionnement de la langue est mis au service de multiples stratégies. Il peut en particulier conduire à de nouveaux appariements, mais il peut également offrir des moyens de vérification de la cohérence globale.

On peut suggérer que les contraintes d'adéquation sont fixées en partie par la langue, qui propose une structure et des relations sémantiques, et par là

même oriente la recherche, et que ce n'est qu'en second lieu que sont prises en compte les contraintes du modèle. Ainsi dans l'extrait suivant, la structure du verbe transformer accompagne une recherche qui s'actualise en lexèmes courants, puis est reformulée de manière à approcher la terminologie du modèle :

Dialogue D & S sur Pile ampoule

157 S Non, le filament, i(l) produit le courant... i(l) reçoit le courant, i(l) produit la lumière.

158 D Voilà, i(l) reçoit le courant, i(l) produit la lumière.
Euh donc i(l) transforme l'énergie en lumière.

159 S Ouais.

Les effets attendus de cette exploration sont multiples et d'une grande importance cognitive.

Le premier de ces effets est un renforcement de l'appariement de structure originel. La saturation peut en effet permettre de départager des appariements concurrents, ou de surmonter l'effet d'une règle non linguistique en augmentant le poids d'une association.

Un second effet est l'organisation du champ de l'expérience dans lequel de nouveaux éléments sont associés aux éléments déjà utilisés et peuvent donc être à leur tour appariés. C'est ainsi que la lampe pressentie pour la catégorie transformateur devient, par la vertu de la structure syntaxique de ce lexème, le noeud d'un réseau où viennent s'agréger l'actant *état initial* : l'électricité qu'elle absorbe, puis un actant *état final* : la lumière qu'elle produit et parfois la chaleur (Figure 4).

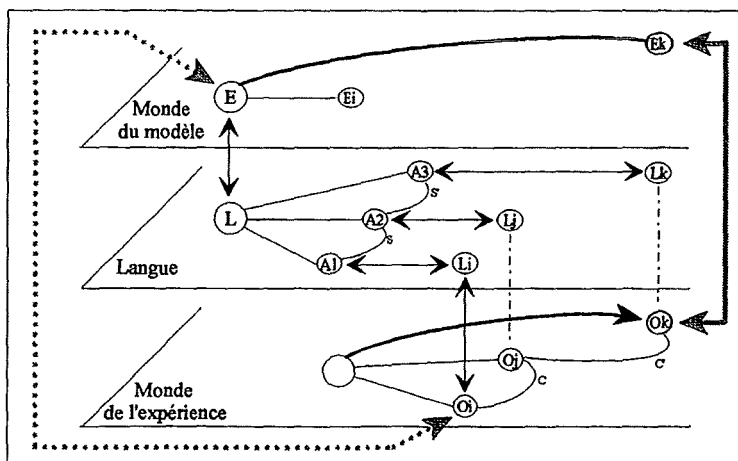


Figure 4 : Découverte d'un appariement par saturation d'un appariement existant

La recherche dans le monde réel d'un objet Ok pouvant jouer le rôle du terme Lk de la structure syntaxique, conduit à un nouvel appariement, figuré par la flèche grisée continue.

Ce processus cognitif est à rapprocher de l'une des qualités dont Jackendoff crédite un système de règles de préférence : toutes les conditions de ces règles n'ont pas à être réunies pour qu'un jugement puisse être porté ; les conditions non utilisées peuvent alors être employées "à l'envers", et conduire aux arguments manquants⁹.

Un troisième effet est la possibilité de contrôle qui découle de l'obtention d'information redondante. Lorsqu'un appariement de structure peut être établi de manière plausible, sans utiliser toutes les valences disponibles, celles-ci peuvent être mises à profit immédiatement ou en temps différé pour contrôler, si besoin est, la valeur de cet appariement, ou pour détecter une éventuelle contradiction à l'intérieur d'une autre association, voire à un niveau global. Le mécanisme de saturation a donc pour effet général d'exploiter à fond la puissance agglomérante et structurante conférée par la langue à certaines entités du modèle. Ce mécanisme participe également à l'évaluation de la plausibilité des appariements de structure déjà établis, et guide éventuellement "l'invention" d'objets ou de *phénomènes périphériques*.

Au terme de l'analyse du fonctionnement et des conséquences de la saturation de structure, il est possible de proposer une règle décrivant ce processus cognitif.

Lorsque le lexème désignant une entité du modèle physique présente une structure syntaxique, il convient de rechercher l'actualisation dans le monde sensible de tous les éléments de cette structure, de manière à parvenir à la forme la plus complète du lexème considéré.

⁹"But one of the virtues of a preference rules system [...] is that one does not have to check *all* the conditions to arrive at a judgment. [...] one can turn around and employ these unused preference rules to supply default values for features of the concepts that have not been established [...]." [Jackendoff, 1990, p 140].

6. Validation des règles

Les règles qui viennent d'être mises au point formalisent, dans le cadre théorique choisi, les hypothèses établies à partir de l'étude d'une première série de dialogues. Ces règles peuvent alors être confrontées à un second groupe de protocoles.

6.1 Un appariement de structure autour du verbe "transformer"

A titre d'exemple, un court extrait du dialogue entre S et M sur la modélisation de l'expérience « pile ampoule » montre que la règle d'appariement de structure peut rendre compte des énoncés observés. La description du verbe transformer a montré que l'actant "état initial", éventuellement fusionné avec l'actant "entité", est nécessaire ; elle a également mis en évidence la nécessité d'une distinction oppositive entre un état initial et un état final de l'entité. Ces propriétés impliquent au minimum la recherche de l'entité susceptible d'être transformée, ou des états attestant sa transformation. Or c'est précisément cette recherche que traduit l'énoncé de S lorsqu'elle déclare :

26 S Le trans... Elle transforme quelque chose,
l'ampoule ?

Les conditions d'application de la règle ne sont donc pas réunies (schéma figure 5), et la proposition est momentanément abandonnée.

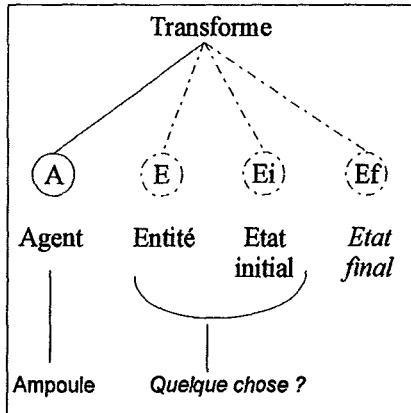


Figure 5 : Les conditions de la règle d'appariement structuré ne sont pas réunies en [26]

M revient nouveau sur cet appariement [40], en proposant cette fois-ci en deux étapes un énoncé qui satisfait maintenant aux conditions nécessaires d'application de la règle d'appariement structuré :

40 M [...] Enfin, l'ampoule transforme l'énergie. L'ampoule transforme l'énergie en lumière. T'es d'accord ?

41 S Ouais.

Les schémas de ces deux énoncés successifs sont représentés sur la figure 6, et montrent bien l'actualisation progressive de la structure.

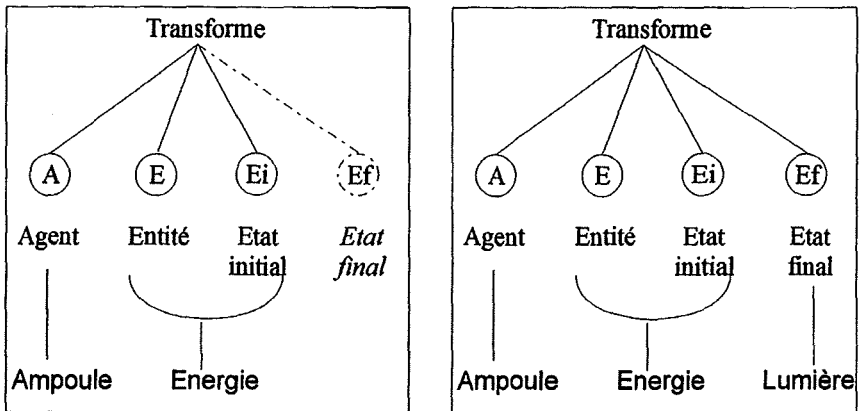


Figure 6 : Les conditions sont réunies en [40]

L'actualisation observée réalise bien un isomorphisme entre la structure syntaxique du lexème qui désigne l'entité du modèle, et un réseau de relations entre objets et phénomènes compatible avec les connaissances de l'élève sur le monde sensible. L'isomorphisme ici établi se traduit à la fois par l'organisation des liens entre les éléments de l'expérience, et par leur nature sémantique qui peut être exprimée à l'aide des primitives sémantiques¹⁰ du verbe transformer.

6.2 La règle de saturation et la découverte de transferts

La mise en place des deux transferts par chaleur et par rayonnement, comme nous allons le voir dans le dialogue de P et M, peut être associée à la saturation du verbe transférer, elle même contrainte par les caractéristiques sémantiques du mot réservoir. Cherchant à utiliser la chaleur et le

¹⁰ L'étude des primitives sémantiques, montre par exemple que "produire" agit comme une primitive de "transformer".

rayonnement, P propose une formulation dans laquelle le mot "environnement" permet de désigner la destination du transfert :

121 P [...] Mais l'énergie utilisée est transformée en chaleur et rayonnement, qui est répandu dans l'environnement.

Pourtant l'environnement ne constitue pas une destination suffisamment précise pour que le déplacement corresponde à un transfert, et la substitution du verbe "répandre" au verbe "transférer" est à cet égard très révélatrice.

Reprenant alors le texte du modèle, M rappelle que "le réservoir initial est différent du réservoir final". L'actant recherché se trouve ainsi précisé et contraint, puisqu'il doit maintenant s'accorder à la signification du mot réservoir. Cette nouvelle contrainte est immédiatement relevée par P, qui déclare, perplexe :

125 P [...] Comment on peut appeler un réservoir qui est l'environnement ou...

L'activité de saturation de la structure syntaxique tend maintenant à découvrir un réservoir plausible. L'acteur recherché, à la fois destination du transfert et réservoir, est d'abord remplacé par l'archilèxème : "quelque part", puis par : "toutes les directions", dans des énoncés simplifiés qui utilisent la primitive "aller" du verbe transférer, puisque l'origine, quant à elle est connue :

127 P [...] Mais l'énergie, le courant électrique est changé en énergie de chaleur et de rayonnement qui va quelque part [...]

[...]

129 P La chaleur, mais la chaleur et la lumière va dans toutes les directions. Mais comment on appelle, on appelle ça.

La contrainte de parenté sémantique de l'actant cherché avec un réservoir apparaît bien dans l'intervention suivante de P, qui recherche désespérément où peut bien "rentrer" l'énergie :

130 P [...] Et celle là elle rentre, elle rentre pas dans la batterie, elle va quelque part.

M propose alors "l'air" comme destination de la chaleur et du rayonnement, mais se rétracte aussitôt. Constatons que "l'air" ne possède

aucun des sèmes qui le désigneraient comme **réservoir**. P tente alors de substituer à "l'air" "l'espace", puis à nouveau "l'environnement", la dicibilité des énoncés obtenus pouvant fournir l'indication d'une meilleure adéquation. La dernière proposition est acceptée avec enthousiasme, et c'est cette fois-ci un effet de contrat didactique qui vient renforcer l'appariement¹¹, comme le montre de manière humoristique l'intervention de M :

146 M Ouais, ah ben oui, tu sais, tu sais en plus ils nous ont donné un questionnaire sur l'environnement. Elle va être contente...

7. L'apport à l'étude des mécanismes de modélisation

On constate donc que l'approche soutenue ici permet de proposer un point de vue linguistique sur certaines des forces qui sont à l'origine des processus cognitifs mobilisés dans ce type tâche de modélisation.

Les règles d'appariement décrivent les bases sémantiques des associations, et montrent le rôle que jouent les caractéristiques sémantiques des termes du modèle.

Il s'est en outre avéré que la connaissance précise de l'organisation sémique des lexèmes permet de rendre compte de l'origine d'un certain nombre "d'échecs" au cours de la tâche de modélisation. C'est ainsi que l'analyse linguistique a montré quelles contraintes langagières rendent difficile l'appariement de la chaleur et du rayonnement, qui ne pourraient donner lieu à des énoncés acceptables.

De la même manière, l'approche linguistique a montré que des contraintes propres à la langue, telles que le respect de l'isotopie du nombre, et l'éclatement des connexions métaphoriques peuvent peser sur les processus cognitifs mobilisés lors d'une tâche de modélisation. Enfin et surtout, il est apparu que ce type de contraintes de nature essentiellement langagières est à l'origine de mécanismes du plus haut intérêt cognitif, qui sous-tendent l'activité heuristique, et participent à la structuration du champ de l'expérience puis à la découverte ou à l'invention.

L'analyse sémantique et syntactico-sémantique qui est proposée dans cette étude pourrait donc constituer le soubassement d'une "description cognitive", en permettant d'explicitier la partie cachée de certains raisonnements, celle qui est la plus proche de la langue, en mettant en

¹¹L'analyse a montré que la plausibilité de l'appariement **réservoir** – **environnement** est faible, mais pas nulle. M indique, en précisant : "en plus", que l'effet de contrat ne fait que confirmer un appariement possible.

lumière les bases langagières de certaines hypothèses qui ne sont pas explicitement argumentées.

Conclusion

Cette étude propose à la didactique un point de vue fondé sur des concepts de linguistique, afin de contribuer à l'analyse du fonctionnement d'une situation de modélisation. Au delà de la situation étudiée, la méthodologie proposée rejoint des préoccupations plus générales des sciences cognitives, et apporte sa pierre à la description de certains aspects du fonctionnement du discours scientifique.

Les concepts d'analyse sémantique employés donnent accès à une description du sens d'un modèle physique qui intègre les connaissances langagières ; ils en modélisent la souplesse et les potentialités. Ils fournissent enfin des éléments qui contribuent au débat sur la similarité en permettant d'aborder de manière homogène des mécanismes qui vont de la similarité de simple apparence jusqu'à la métaphore¹².

Le concept d'homogénéité sémantique, pour sa part, permet de fonder théoriquement la notion de dicibilité en langue d'un énoncé. Insistons aussi sur la fonction heuristique, dont l'importance est bien entendu de premier ordre d'un point de vue cognitif. Selon Gentilhomme, la notion, attachée à un terme et portée par la langue, en constitue le moteur. Or cette notion, versant linguistique d'une connaissance préalable¹³, peut être cernée à l'aide des descriptions sémantiques proposées dans cette étude. L'activité langagière peut alors être considérée au minimum comme un miroir de l'activité cognitive, mais on peut raisonnablement suggérer que des règles propres au fonctionnement de la langue influent directement sur les hypothèses avancées par les apprenants, ainsi que sur les solutions qu'ils retiennent. On a vu en particulier comment certaines stratégies heuristiques¹⁴, peuvent être en partie révélées, voire partiellement prédites, sur la base d'une approche linguistique.

Il apparaît ensuite que ce travail de recherche met en lumière la place qu'occupe la langue dans la construction des concepts scientifiques. L'approche linguistique rejoint ici la position didactique, et montre de quelle manière le concept scientifique se construit à l'aide de la notion préalable véhiculée par un terme. Loin de considérer que la notion entache le concept

¹² [Gentner, 1989].

¹³ [Johsua et Dupin, 1993, p 122].

¹⁴ [Gentilhomme, 1994, P 20].

d'un flou originel, qu'elle est en somme un "mal nécessaire", on peut considérer qu'elle constitue non seulement un passage obligé, mais bien davantage la partie motrice de la signification. Le champ de cette étude peut alors être élargi, puisque ce mode de fonctionnement du modèle dans le discours scientifique n'est pas particulier à l'apprenant débutant, mais concerne jusqu'au spécialiste lui même, une science étant toujours en cours d'élaboration.

Bibliographie

- Bogacki, K., Bogustawska, G., Kreisberg, A., Lewicka, H., Lozinska, M., Thieme, M., Wójcik, L., Zielinska, T. (1983) *Dictionnaire sémantique et syntaxique des verbes français*, dir. Halina Lewicka et Krzysztof Bogacki, Pánstwowe Wydawnictwo Naukowe, Warszawa.
- Cordier, F. (1994) *Représentation cognitive et langage: une conquête progressive*, Armand Colin, Paris.
- Gaudin, F. (1993), *Pour une socioterminologie*, Presses Universitaires de Rouen, Rouen.
- Gentilhomme, Y. (1994) L'éclatement du signifié dans les discours technoscientifiques, dans *Cahiers de lexicologie* n° 64, 1994-1, Didier érudition, Paris.
- Gentner, D. (1989) The mechanism of analogical learning, dans *Similarity and analogical reasoning*, dir. Štella Vósniadou & Andrew Ortony, Cambridge University Press.
- Guilbert, L. (1973) La spécificité du terme scientifique et technique, dans *Langue française* N° 17.
- Hacking, I. (1989) *Concevoir et expérimenter*, Bourgois, Paris.
- Halbwachs, F. (1971) L'histoire de l'explication en physique, dans *Les théories de la causalité*, PUF, Paris.
- Jackendoff, R. (1990) *Semantics and cognition*, The MIT Press, Cambridge, Massachussets.
- Johsua, S. et Dupin, J.J. (1993) *Introduction à la didactique des sciences et des mathématiques*, PUF, Paris.
- Kocourek, R. (1991) *La langue française de la technique et de la science*, Brandstetter Verlag, Wiesbaden.
- Picoche, J. (1986) *Structures sémantiques du lexique français*, Nathan, Paris.
- Rastier, F. (1987) *Sémantique interprétative*, PUF, Paris.
- Rastier, F. (1991) *Sémantique et recherches cognitives*, PUF, Paris.
- Tesnière, L. (1965) *Eléments de syntaxe structurale*, Klincksieck, Paris
- Tiberghien, A. & Megalakaki, O. (1994) Contribution to a characterisation of a modelling activity, case of a first qualitative approach of energy concept, *European Journal of Psychology in Education*.
- Gérard COLLET, UMR GRIC-COAST, Université Lyon 2, Bat C, ét. 1
5, Av Pierre Mendès France CP 11 69676 BRON CEDEX 11

Modélisation et pratique en psychopathologie neurologique

C. Fayada

*Centre du Langage, Service Agid & Lyon-Caen
Hôpital La Salpêtrière,
47, Bd de l'Hôpital, 75013 PARIS - France*

1. INTRODUCTION

Les recherches en neuropsychologie mettent en oeuvre des modèles liant effets des lésions cérébrales sur les processus psychologiques, et troubles des comportements. Ces modèles s'inscrivent dans une perspective des bases dites « biologiques » et/ou « psychologiques » des comportements anormaux. Dans le cadre de ces modèles, notre réflexion est issue des nécessités d'une action quotidienne du praticien (neurologue ou psychologue) auprès des patients neurologiques, pour un diagnostic prédictif de leur dynamique d'adaptation et pour leur prise en charge thérapeutique.

Des sujets avec certaines lésions cérébrales présentent des troubles du comportement et de l'adaptation à la vie quotidienne, et des relations avec autrui, qui sont mal « évalués » par les tests neuropsychologiques usuels, et même difficiles à simplement mettre en évidence dans les situations expérimentales classiques. Ces tests et situations, issus des modèles neuropsychologiques classiques (qui constituent un peu un kaléidoscope d'approches éclatées, dont les fondements épistémologiques sont peu explicités), ne sont donc pas suffisamment adéquats. Les raisons de ces difficultés sont liées au « modèle » objet de l'étude (le sujet humain cérébrolésé, c'est à dire autrui, comme à la fois différent et identique à soi - au contraire de l'animal de laboratoire des neurosciences ou de la psychologie animale) et à l'histoire de la neuropsychologie (concepts et méthodes d'exploration du cerveau et des processus psychologiques). En fait, la neuropsychologie s'est développée dans une perspective explicative biologique ou psychologique des comportements, et peu dans celle des nécessités cliniques et thérapeutiques, à dimension sociales, posées par ces patients. De cette "praxis" émerge la nécessité de modèles intégratifs des comportements, susceptibles de contribuer, par leur cohérence et leur

“forme”, à orienter l'action du praticien vis à vis des troubles de comportement des patients.

Nous allons alors essayer de préciser en quoi la « praxis » de la clinique psychopathologique (notamment, la gestion des situations de crise de l'intégration sociale des patients) invite à “revisiter” les modèles classiques de la neuropsychologie. Pour ce faire, nous les résumerons globalement. Puis nous envisagerons comment la considération des plaintes des patients et de leur famille, les enjeux de la clinique (en particulier sur un plan qui n'a rien de spéculatif, mais qui est d'ordre thérapeutique et psychosocial), et enfin les intentions de communication de l'observateur-thérapeute (le praticien neurologue ou psychologue), amènent à renouveler les perspectives théoriques elles-mêmes et les pratiques.

2. MODÈLES EN NEUROPSYCHOLOGIE

2. 1. Anomalies des processus psychologiques

Certains patients avec lésion cérébrale présentent classiquement des anomalies psychologiques dans les registres du langage, de l'activité visuospatiale, de l'attention, de la mémoire ou de l'action (dans la perspective des activités générales), ou du raisonnement et de l'activité intellectuelle. Ces anomalies psychologiques sont diversement associées et regroupées en « syndromes » en fonction de l'étendue et de la topographie des lésions cérébrales. Elles sont étudiées au moyen de tests neuropsychologiques classiques (« mesurant » ces activités psychologiques catégorisées) ou de situations expérimentales contraintes explorant plus analytiquement ces mêmes processus psychologiques.

Ces anomalies ont l'avantage d'indiquer des processus psychologiques intermédiaires sous-jacents à une activité mentale « normale » et de contribuer à la connaissance des bases psychologiques des comportements (Luria, 66, Baddeley, 89, Grafman, 89, Shallice, 93, Jeannerod, 94). En revanche, de manière critique, on peut considérer que précisément cette perspective « découpe » l'activité mentale. Elle n'associe pas les différentes anomalies des processus, ne prend pas non plus en compte la dimension sociale de cette activité, ni les motivations et les caractéristiques spatio-temporelles des contextes du quotidien, ni encore les spécificités individuelles des patients. Par ailleurs, sur le plan psychopathologique, persiste une ambiguïté d'interprétation de l'activité

psychologique (« déficitaire » ou « compensatoire » ?) après lésion cérébrale.

2. 2. Les syndromes topographiques : les corrélations anatomocliniques

Classiquement, la neuropsychologie est dans la tradition neurologique et vise à établir des corrélations entre lésions cérébrales et anomalies neuropsychologiques. Ces dernières sont regroupées selon différents critères topographiques des lésions, comme par exemple les « syndromes frontaux » (troubles de l'organisation de l'action, troubles particuliers de l'attention, de la mémoire ...) ou bien leurs latéralités hémisphériques (troubles du langage à gauche, ou troubles visuo-spatiaux à droite), ou enfin leur topographie corticale ou sous-corticale (Botez 87).

Ces corrélations ont l'avantage de documenter les bases biologiques des comportements et la localisation de certains « noeuds » des processus psychologiques par rapport à l'activité distribuée des réseaux de neurones. En revanche, elles sont limitées par les spécificités des techniques utilisées (scanner, IRM, IRM fonctionnelle, potentiels évoqués). Également, les composantes métaboliques (neurotransmetteurs, hormonales ...) sont peu souvent intégrées (sauf dans les « modèles » de la maladie de Parkinson ou de la Chorée de Huntington).

2. 3. Application des syndromes psychiatriques (troubles émotionnels) en psychopathologie neurologique

Pour certaines pathologies neurologiques (démences, traumatismes crâniens, épilepsie), les troubles émotionnels sont « cliniquement » visibles (Damasio, 95) et posent des problèmes de prise en charge. Sont alors importés des syndromes psychiatriques et des critères « DSM 4 » (DSM 4, 1996) (démences, troubles de la personnalité, dépression...). De nouvelles échelles dérivées spécifiques de ces mêmes pathologies peuvent être construites.

Ces syndromes ont l'avantage de prendre en compte la dynamique globale des patients, en lui associant leur personnalité antérieure. Également, des dimensions intermédiaires (comme le sens du soi, par exemple) sont incluses. Des repères (en particulier fonction de systèmes biologiques) peuvent ainsi être déterminés pour prescrire des médicaments psychotropes. En revanche, l'ontologie des troubles psychologiques est radicalement différente de ces syndromes. Par ailleurs certains items de définition de ces syndromes (notamment dans la dépression ou l'anxiété) sont peu pertinents

sur le plan phénoménologique, en raison des troubles d'attention, de mémoire et de langage.

En résumé, la démarche neuropsychologique classique s'attache à considérer les bases biologiques et/ou psychologiques des comportements anormaux des patients, dans une démarche peu intégrative, étant donné d'une part, la tradition de cette démarche, et de l'autre, la nature concrète, repérable, focalement identifiable des lésions cérébrales (sur lesquelles on peut donc immédiatement tabler comme "cause" des phénomènes psychopathologiques) Pourtant, les nécessités pratiques de la clinique psychopathologique invitent à adopter des perspectives plus intégratives. En effet, la «praxis» neurologique contraint à réinterroger les limites des modèles eux-mêmes et les conditions de leur utilisation.

3. PSYCHOPATHOLOGIE NEUROLOGIQUE : QUELS MODÈLES ?

3. 1. Les troubles de comportement en psychopathologie neurologique : les plaintes

Notre réflexion s'est développée dans le cadre d'une action quotidienne avec des patients porteurs de lésions cérébrales fronto-temporales, qui présentent des troubles spécifiques du comportement et de l'adaptation (Messulam, 86). Ces troubles sont particulièrement envahissants dans la vie quotidienne, et sont à l'origine de plaintes formulées par le patient et/ou la famille, s'exprimant essentiellement selon trois registres :

- handicap dans les actions (en particulier dans les activités socioprofessionnelles antérieures, en termes de motivation ou de régularité, et selon la contrainte de ces actions),
- difficultés relationnelles (en termes de communication avec autrui - retrait ou approche exagérée -, avec intolérance éventuelle à la frustration ...).
- changement de personnalité (propos agressifs ou bizarres, changement de l'humeur, en termes d'éléments de discours caractérisant le soi, d'expérience émotionnelle transmissible, ou de profils d'interaction du patient avec sa famille),

Ces plaintes reflètent une limitation de l'autonomie du patient et une intolérance familiale et sociale. Elles présentent des caractéristiques intéressantes pour notre propos :

- elles s'expriment en termes de handicap adaptatif, avec anomalies du comportement global regroupant des anomalies de comportement classiquement considérés comme indépendants : cognition et action, d'une part (anomalies de l'organisation de l'action, de la représentation du soi, ou de l'interaction avec autrui, dans la communication en particulier), émotions et actions, d'autre part (motivation, nature de relations avec autrui, personnalité ...). La question posée est celle de l'articulation des mécanismes sous-tendant ces comportements permettant de rendre compte du trouble de l'adaptation dans sa globalité.

- elles témoignent d'une différence avec les comportements antérieurs, quels que soient la personnalité ou le style cognitif du patient. La question posée est celle de la pondération des facteurs de personnalité dans les troubles de comportement.

- elles sont relatives aux conditions d'adaptation aux activités socioprofessionnelles. Ces dernières, au contraire des situations expérimentales, sont de nature "écologique" (en particulier sociales), en mobilisant un sens, une motivation et des intentions d'actions, qui sont en règle générale très différentes de ces situations, ou des tests neuropsychologiques. La question posée est celle des conditions d'une observation et d'une « mesure » pertinente des troubles de l'adaptation à la vie courante.

- leur expression est spécifique du système "patient-milieu", en particulier social, avec des facteurs de régulation émotionnelle et des comportements propres à un patient donné, et une valeur spécifique des « signaux » émotionnels qu'il fournit à son entourage. La question posée est celle de la spécificité psychopathologique de ces signaux, en fonction du milieu du patient.

- les plaintes s'expriment en termes d'intensité (en comparaison avec l'état antérieur et les problématiques de deuil), de souffrance globale du système patient-milieu, et demande d'aide thérapeutique. La question posée est celle des indices permettant d'assurer efficacement cette aide.

En résumé, les plaintes concernant les comportements des patients amènent à revisiter et articuler les modèles utiles en neuropsychologie pour « comprendre » les anomalies de l'action et du comportement des patients, dans une démarche intégrative. Outre les impératifs de la situation clinique, trois dimensions paraissent susceptibles d'organiser les fondements d'une telle démarche : « globalité et heuristique », « interaction avec un contexte "écologique" (en particulier social) », et « spécificité individuelle » (relative à l'ontologie et au milieu socioprofessionnel).

3. 2. Les exigences de la situation clinique

Les plaintes des patients et de la famille émergent dans l'entretien clinique traditionnel, avec des demandes d'intervention très concrètes. Les enjeux sont certes thérapeutiques, en particulier avec la nécessité fréquente de gérer les situations de crise (liée aux difficultés d'intégration sociale), mais aussi sociaux et financiers (chances de réinsertion professionnelle, réparation juridique et financière d'un préjudice subi, responsabilité civile ...). Ils concernent les modes d'évolution du système patient-famille (au domicile, en institution ...). L'évaluation de ces patients se réalise à partir d'informations, subjectives ou objectives, fournies par l'entretien clinique, les tests neuropsychologiques, et les documents neurophysiologiques ou radiologiques. Ni corrélations anatomo-cliniques, ni DSM IV, ne sont suffisants pour véritablement prédire les capacités d'adaptation des individus ou les prendre en charge. Les tests classiques de laboratoire prédisent mal l'adaptation à la vie quotidienne, et réciproquement, en ce qui concerne le DSM IV, la phénoménologie de la dynamique comportementale des patients est fortement marquée par des anomalies attentionnelles, mnésiques, et par le langage Pourtant, l'expérience du clinicien averti, avec ses connaissances implicites et explicites et sa subjectivité au moment de l'entretien, lui permet souvent de s'orienter globalement et de formuler des conseils thérapeutiques. Compte tenu des enjeux, pour être efficace, il faut s'orienter dans la diversité des modèles susceptibles d'être évoqués, tant dans la description des troubles du comportement, que des processus psychologiques en oeuvre, ou des contextes d'expression et d'observation. Pour ce faire, il est souhaitable de s'entourer de moyens de réflexions complémentaires, c'est à dire de situations d'observation des troubles de comportement plus spécifiques, en contextualisant ces comportements dans un bilan médico-psychologique intégratif. Celui-ci sera susceptible de répondre à plusieurs questions :

- sur le plan pratique, comment concrétiser la réalité et l'intensité du handicap comportemental évoqué ?

- quelles sont les spécificités de cet handicap liées à l'individu (caractéristiques développementales, familiales, socioprofessionnelles)

- quelles sont celles plus spécifiquement liées au contexte (avec des fluctuations du comportement souvent notées par la famille), et celles liées à la pathologie ?

- de manière fiable, comment formuler une prédiction quant aux possibilités d'adaptation ultérieure du patient dans son milieu socioprofessionnel (ce qui est fondamental, en particulier en termes

d'expertise avec les notions de déficiences, incapacités, situations de handicap, etc.).

- enfin, comment traiter le patient ?

En résumé, la considération des plaintes de comportement avait invité à renouveler la place et l'articulation des modèles neuropsychologiques classiques. Les enjeux importants de la psychopathologie neurologique amènent à « mesurer » l'adaptation des patients (étant donné la nécessité d'une mise en forme et d'une lisibilité comparative et prédictible), en termes intégratifs. Par ailleurs, la « praxis » neurologique met en jeu une interaction entre les intentions d'action et de communication de l'observateur thérapeute et de son patient, qui est à prendre en compte par les modèles neuropsychologiques.

3. 3. Les intentions de communication de l'observateur-thérapeute

Dans l'activité quotidienne, le praticien (neurologue, psychologue) est un observateur-thérapeute, avec en fait deux catégories de missions ou d'intentions qui marquent profondément ses actions et ses représentations :

- une première catégorie, à l'origine d'une action de « simulation » de la psychopathologie du patient, pour « comprendre » et prédire son adaptation à la vie quotidienne.

- une deuxième catégorie, à l'origine d'une « co-action » avec ce même patient, dans les situations de communication et de médiation thérapeutique.

Pour agir, le praticien « expert » développe des représentations de son patient, issues à la fois de son expérience implicite (issue elle-même des connaissances et pratiques propres à l'apprentissage explicite antérieur, et des ajustements quotidiens aux patients), et d'une action plus explicite, orientée fortement par les « modèles d'autrui » neuropsychologiques, et actualisée par les tests effectués sur ces patients. Les représentations du praticien, développées sur les bases biologiques et psychologiques des comportements, sont sujettes à un "flottement" chronique de ses intentions explicatives et causales. Il se traduit soit par une cristallisation autour des corrélations anatomo-cliniques, ou anatomo-psychologiques (évacuant les processus et la dynamique psychologique au sens psychiatrique, DSM IV par exemple), soit par le rejet de toute corrélation de ce type. Dans la perspective qui nous concerne, qui est d'ordre pragmatique et non seulement spéculative (c'est à dire ancrée dans les problèmes du patient qui peut ou non reprendre son travail, et pour lequel la famille attend une prise en charge efficace), il est essentiel d'essayer d'éviter un tel flottement.

Le concept de représentation de l'action et de l'interaction avec autrui permet d'étudier les conditions de l'efficacité de l'action du praticien, dans sa « simulation » du patient et ses "co-actions" avec ce dernier :

- Dans la « simulation », le praticien-observateur a pour objectif de prédire l'adaptation du patient à la vie quotidienne, en essayant de comparer activités mentales et capacités d'adaptation supposées de ce patient, à l'aide d'évaluations spécifiques. Ces évaluations sont issues soit du modèle neuropsychologique, soit du contexte socioprofessionnel que le praticien peut inférer (à l'aide d'une représentation très approximative). L'application du modèle neuropsychologique à l'homme au travail est censée permettre de formuler un avis « expert », et de prédire le devenir du sujet cérébrolésé dans son monde socioprofessionnel (prédiction à l'évidence lourde de conséquences pour le patient).

- Dans la co-action et la médiation thérapeutique, l'objectif du praticien-observateur est beaucoup plus dynamique, et nécessite une lecture du comportement et des intentions d'autrui différente du cas de la "simulation" précédente. Le praticien-observateur, engagé dans une action commune avec le patient et une « co-contruction », doit ajuster ses représentations aux anomalies de comportement du patient dans la communication, et exercer une médiation thérapeutique pragmatique.

Ces réflexions peuvent contribuer d'ailleurs à illustrer les difficultés des neurosciences et des sciences humaines à construire des modèles communs : l'intention et l'action de l'observateur vis à vis de son objet d'observation - autrui - est souvent fluctuante, autrui instrumentalisé comme un objet physique, ou en co-action avec l'observateur.

En résumé, la question posée par la « praxis » de l'observateur-thérapeute est bien celle encore une fois des conditions d'utilisation des modèles, de manière complémentaire aux paragraphes précédents. Au delà des nécessités d'une démarche intégrative pour articuler des niveaux opératoires au sein d'une modélisation de type « neuropsychologique » (ancrée dans les bases cérébrales des comportements), la référence aux intentions d'action du praticien implique de nouveaux modèles, qui sont de fait implicitement pratiqués, dans la mesure où le modèle neuropsychologique classique n'autorise pas à prédire une adaptation socioprofessionnelle, sans référence à un modèle (en général implicite) d'aptitudes de l'homme au travail. C'est ainsi qu'on peut se demander si le modèle neuropsychologique est compatible avec ceux de communication avec autrui, dans l'interaction thérapeutique avec le patient, et si il est légitime de passer d'un modèle à l'autre, non seulement sur le plan théorique mais aussi pour la pratique (espaces de validité des tests proposés).

4. QUELQUES PROPOSITIONS

4. 1. Quelques références complémentaires utiles

Trois perspectives convergentes, « extra-neuropsychologique » fédèrent des notions clés susceptibles de contribuer à un support conceptuel pluridisciplinaire : les neurosciences, l'ergonomie cognitive et la systémique, incluant le domaine psychosocial. Les "noeuds" de modélisation et des paramètres analogiquement comparables, caractérisant un système dynamique, peuvent être alors être repérés.

- les neurosciences

Fuster (1995), à partir de travaux d'électrophysiologie chez le singe et d'imagerie fonctionnelle chez l'homme normal, évoque la structure et la dynamique de l'activité neuronale (dans l'espace et le temps) qui sous tend les comportements observés. Des analogies directes sont possibles entre l'activité neuronale et les différents processus psychologiques d'attention, de mémoire, de perception et d'action motrice, en fonction de contextes expérimentaux restreints. En effet, l'intention, le déroulement d'un comportement, ou l'utilisation des mots, de la syntaxe, etc., dans un discours, correspondent à une activité neuronale multidimensionnelle, articulant les canaux d'*entrée* sensoriels aux *sorties* motrices observables.

Des autocorrections (boucles de rétroaction, copies d'efférences ...) se produisent à tous les niveaux - du neurone aux comportements - en fonction des contextes internes et externes (qui ne sont pas observables pour la plupart, ni par le sujet, ni par l'expérimentateur). Par ailleurs, les patterns d'activation neuronale sont aussi fonction de l'anticipation et des contextes d'interaction (l'interlocuteur, la situation d'interaction) fortement prédictibles par le sujet, ou non, et concernant le soi (en termes opérationnels de routine comportementale, ou des diverses mémoires ...).

Fuster décrit particulièrement la spécificité du réseau fronto-temporal (cycle perception-action), fondamental pour l'adaptation fine, au quotidien, des actions et des comportements. Il insiste sur la nécessité d'employer des notions opératoires, identiques pour différents domaines psychologiques, en se basant sur des faits expérimentaux aussi simples et irréductibles que possible, et en faisant appel à un minimum d'hypothèses et d'inférences. La référence neuronale liée à l'objet de représentation « cerveau » amène alors

à affiner l'objet «dynamique-psychologique», et à définir des critères plus précis d'observation des comportements adaptatifs dans un contexte donné.

- la psycho-ergonomie

Hoc (1996), à partir de travaux concernant les postes de travail et le diagnostic de panne dans des actions quotidiennes complexes, décrit l'évolution des actions et de l'adaptation de l'homme à son poste selon deux caractéristiques de traitement de l'information : un niveau statique et un niveau dynamique, introduisant une notion de temps liée aux actions entreprises (indiciage des prises d'information dans le contexte externe), et non liée au temps chronologique, dans le contexte du travail quotidien. Cet auteur illustre la nécessité de considérer que les conditions d'interaction du contexte interne avec le contexte externe sont déterminantes, s'il s'agit de conditions de déroulement de l'action en situation écologique, mais pas s'il s'agit de tests de laboratoire.

Par ailleurs, sur le plan des aptitudes au travail et à l'adaptation, il souligne, en collaboration avec Amalberti (1996), l'importance des capacités d'autonomie, d'apprentissage en toutes situations (transfert de connaissances), de flexibilité de stratégies, de rétroaction et du sens du soi, qualités comportementales qui font justement défaut aux patients avec troubles du comportement, et qui sont encore peu considérées de manière intégrative en neuropsychologie.

Également, dans le domaine du stress (Bruchon-Schweitzer 94) et de la psychopathologie du travail (Dejours 93), les notions de "coping" et de représentations sociales communes paraissent particulièrement pertinentes, compte tenu des conflits et des crises d'intégration sociale observés dans la population des patients cérébro-lésés.

- la systémique

La systémique (Andreewsky 91; Shannon 93), dans une perspective phylogénétique, ontogénétique et heuristique, introduit les notions de système ouvert-fermé, d'homéostasie et de rétroaction d'un système. Elle légitime des modélisations globales, en articulant facteurs biologiques et facteurs psychosociaux. Centrée sur les enjeux complexes du comportement de l'individu et de la société, elle implique le contexte, les intentions et les motivations : métaconnaissances mises en jeu, actions avec autrui, et élaboration de l'identité de l'individu dans un groupe social complexe. Elle rend compte de l'importance des comportements globaux étudiés en éthologie, avec en particulier le rôle et le stéréotype caractérisant les

communications sociales. Les contextes implicites (ceux des situations et de la représentation d'autrui) et explicites (notamment linguistiques) peuvent ainsi être articulés. La systémique fait le lien entre les deux perspectives précédentes. Elle justifie une démarche intégrative, soulignant le sens des actions complexes, en particulier sociales, leurs dimensions spécifiques (en particulier en liaison avec la notion d'identité et de permanence des systèmes), mais aussi la limitation des systèmes en question.

4. 2. L'adaptation du patient cérébrolésé à son milieu socioprofessionnel : sur le plan théorique, modèle neuropsychologique ou celui de l'homme au travail ?

Les considérations précédentes nous permettent de situer les comportements comme l'expression de processus d'adaptation « ajustée » aux contextes internes et externes, dans une dimension complexe, ontologique et bio-psycho-sociale, des individus. L'ontologie (phylogénétique et ontogénétique) témoigne de l'évolution des représentations selon les conditions de survie et d'adaptation des individus. L'intégration biologique et psychologique des comportements témoigne de la complexité de l'organisation (hiérarchisée et parallèle) des processus d'adaptation, du biologique au psychosocial, et au comportement "phénoménologiquement" visible ou « mesurable ». Le comportement peut alors être considéré dans une perspective dynamique d'émergence liée au contexte, en termes d'adaptation à ce contexte :

- Il existe différents « sens » possibles du comportement, en relation avec la complexité de l'adaptation, et l'action heuristique selon le contexte ; le « sens » d'un comportement correspond à la nécessité d'un signal émergent, spécifique du système organisme vivant et autrui (par l'interaction et la communication sociale), en fonction des situations d'adaptation, et selon des intentions d'action qui vont de la simple survie à des nécessités très élaborées.

- Les conditions d'expression comportementale de la représentation des actions sont étroitement liées aux caractéristiques fines des contextes d'action et d'interaction. Toutes les catégories de la cognition sont par là-même articulées dans une dynamique psychologique : perception focalisée sur des cibles pertinentes pour l'action et le contexte interne, activation sémantique et langagière, en rapport avec des informations mnésiques, activation de scripts d'action prototypiques, abstraction, conceptualisation, et surtout processus attentionnels. A l'instant « t », le comportement « visible » de l'individu dans une situation donnée, dépend de la spécificité

très transitoire de son interaction (attentes et anticipations fonctions de scripts antérieurs, et des résolutions de problème situationnels).

- L'adaptation et la dynamique temporelle des comportements. Pour faire face aux contraintes exercées par le milieu extérieur et intérieur (conditions d'homéostasie de l'organisme), et construire en permanence les conditions d'autonomie du sujet, les actions se déroulent dans un contexte donné, selon des processus ascendants et descendants.

- Plan pratique : contextes intégratifs pour l'évaluation - les enjeux :

Les enjeux de l'évaluation des patients sont liés à la mise en oeuvre d'une vision heuristique des troubles de l'adaptation de ces patients, et aux problèmes posés par la psychopathologie neurologique. Ils se définissent selon quatre préalables :

- description « structurée » et liens de causalité en termes adaptatifs (c'est à dire organisée et hiérarchisée) et opératoire (c'est à dire pertinente pour les problèmes d'adaptation posés au quotidien par ces patients), selon une trame de réflexion théorique intégrative (des neurosciences à la psychologie du travail ou à la systémique) des capacités d'adaptation du patient. Cette description doit permettre au thérapeute de disposer d'informations d'ordre objectif et subjectif, non seulement pour « comprendre » la dynamique psychologique du sujet à l'instant « t » (au delà de l'empathie et d'une expérience implicite), mais encore pour prédire ses capacités d'adaptation dans son milieu, et pour conseiller des orientations thérapeutiques plus pertinentes.

- valeur relative et non absolue, selon le contexte socioprofessionnel antérieur, de la description finale des capacités d'adaptation du patient.

- pas de préjugé initial de l'origine « organique » ou « fonctionnelle » des troubles observés. A posteriori seulement, des arguments particuliers, pas toujours spécifiques, peuvent plaider en faveur de l'une ou l'autre, au moins partiellement.

- pas de diagnostic psychiatrique, neuropsychologique ou neurologique, mais « psychologique », selon les dimensions clés de l'adaptation repérées (« noeuds » psychologiques pour prendre l'analogie avec les réseaux d'information), synthétisées en termes de dynamique et de profil psychologiques.

Les contextes

L'objectif est de contextualiser les troubles du comportement et de l'adaptation des patients, en intégrant dans une perspective systémique les

anomalies de l'attention, de la mémoire et du langage, et les caractéristiques des processus inférées en fonction de la topographie apparente des lésions cérébrales. Un modèle intermédiaire « bio-psychologique » doit être susceptible d'aider à anticiper la dynamique d'adaptation du patient, dans un "va et vient" entre les mesures et les concepts, qui consiste à hiérarchiser les informations recueillies et à préciser le niveau des processus psychologiques concernés.

En important des caractéristiques de l'adaptation d'un sujet donné dans ce modèle intermédiaire (en particulier en ce qui concerne les activités quotidiennes et l'interaction sociale), et en les associant aux caractéristiques des comportements usuels des patients avec lésion cérébrale, il est possible de situer les informations utiles à recueillir, et de proposer des contextes permettant des comparaisons utiles :

- d'abord, recueil d'informations liées au contexte de vie quotidienne du sujet, avant la maladie neurologique (informations d'ordre factuel concernant l'ontologie des événements de vie, les orientations professionnelles, mais aussi informations d'ordre subjectif, concernant le style adaptatif), et depuis cette maladie (au moyen d'échelles standardisées d'autonomie et de comportement social)

- ensuite, plusieurs contextes peuvent être proposés : constructions spatiales (consignes progressives, forme de l'action avec les objets), tâches standardisées de communication avec autrui (forme de l'action avec autrui), tâches concernant la représentation d'autrui, tâches avec ou sans contrainte de temps (modification de l'action selon les conditions de travail), tâches d'interprétation, selon la polysémie des contextes (détection de signal, quelle représentation pour quelle action), tâches indiquant la régularité des comportements ...

La méthode doit bien préciser les modes de présentation des consignes (auditives, verbales, surprises ou non ...) et les modes de mesure de l'expression comportementale (expression gestuelle, temps de réaction, performances, éléments de discours), en limitant les inférences. Les questions de fond concernant l'articulation des référents dits « cognitifs » et ceux qui sont plus psychosociologiques peuvent se résumer ou s'illustrer selon la question suivante : en quoi les troubles de l'adaptation de l'action dans les activités quotidiennes et dans l'interaction sociale (opérationnalisées dans une situation de communication) sont-ils différents ou comparables, quels facteurs attentionnels susceptibles de mobiliser des changements d'action et de comportement peuvent être identifiés ?

4. 3. L'interaction thérapeutique avec le patient cérébrolésé en clinique psychopathologique. Modèle neuropsychologique ou modèle de communication ?

- les enjeux

L'enjeu est celui d'une communication thérapeutique avec un patient qui présente des troubles du comportement dans l'interaction sociale, et chez qui souvent des anomalies du sens du soi ou de l'état affectif ou émotionnel, sont constatées. Pour ces patients, la situation de communication avec le praticien représente une forme de contextualisation sociale.

Les troubles de l'interaction sociale de ces patients sont mal connus sur le plan phénoménologique ou sur celui des anomalies des processus psychologiques en cause. Leur expression est hétérogène, et phénoménologiquement dépendante des troubles de l'attention, de la mémoire ou du langage, mais sans corrélation, semble-t-il, avec les scores mesurant ces mêmes anomalies dans les situations de tests neuropsychologiques classiques. Pour certains patients, à l'aide d'un matériel mis au point chez des patients schizophrènes, il est possible aussi d'observer des anomalies de la représentation des intentions d'autrui (Fayada, 96). Certains patients présentent des anomalies de la communication avec autrui, susceptibles d'être interprétés soit dans le registre de la pragmatique (Trognon, 93), soit dans un registre plus intégratif et systémique, articulant les données bio-psycho-sociales, par exemple en considérant que les troubles de la communication sont l'expression d'un trouble de l'interaction sociale (dans une double perspective, d'abord dans celle d'un comportement de communication spontanée anormale (au temps t_0) et ensuite dans celle d'un ajustement dynamique anormal à l'interaction en cours (t_i) (Fayada, 96).

Sur le plan des anomalies du sens du soi, et de la dynamique émotionnelle interne, les "contenus de pensée" sont décrits ponctuellement, et là encore avec des pondérations mal connues, dépendantes des troubles « cognitifs » ponctuels, mais sur le plan clinique « intuitif » nécessitant une prise en charge globale.

- ébauche de proposition

Les troubles de l'interaction sociale et de la dynamique émotionnelle interne du patient sont à l'origine d'une souffrance globale du système patient-famille, et il est souhaitable de proposer des actions thérapeutiques.

La question est alors d'essayer d'identifier des modes d'adaptation pertinents du patient à l'observateur-thérapeute, pour développer une co-construction dont les objectifs sont à identifier, en particulier dans une dynamique de reconstruction de l'identité, pour ces patients qui présentent des anomalies fréquentes du sens du soi. L'observateur-thérapeute ajuste sa « co-action » en fonction à la fois des modèles neuropsychologiques qu'il actualise, mais aussi, et surtout, il opère en temps réel avec le patient, une telle interaction n'étant pas comparable aux situations de thérapie classiques. Peu de données existent sur ce sujet actuellement.

5. CONCLUSION

La démarche neuropsychologique classique s'attache à considérer les bases biologiques et/ou psychologiques des comportements anormaux des patients, dans une perspective peu intégrative. Une telle perspective est liée, d'une part, à la tradition de cette démarche, et de l'autre, à la nature concrète, repérable et focalement identifiable des lésions cérébrales ("causes" prêtes à penser des phénomènes psychopathologiques). Pourtant, la clinique psychopathologique invite à adopter une perspective plus intégrative ; en effet, la « praxis » neurologique contraint à réinterroger les limites des modèles, et les conditions de leur utilisation. Ces réflexions peuvent contribuer par ailleurs à illustrer les difficultés des neurosciences et des sciences humaines à construire en commun un modèle des intentions et actions de l'observateur par rapport à son objet d'observation - autrui.

BIBLIOGRAPHIE

- AMALBERTI R., 1996, La conduite de systèmes à risque, le modèle comme outils de base de la gestion cognitive, Presses Universitaires de France.
- ANDREEWSKY E., 1991, Systémique et cognition, Afcet systèmes, Dunod édition.
- BADDELEY A., 1986, Working memory, Oxford Clarendon Press.
- BOTEZ M.I., 1987, Neuropsychologie clinique et neurologie du comportement, Presses Universitaires de Montréal.
- BRUCHON-SCHWEITZER M., DANTZER R., 1994, Introduction à la psychologie de la santé, Presses Universitaires de France.
- DAMASIO A. R., 1995, L'erreur de Descartes, la raison des émotions, Editions Odile Jacob.
- DEJOURS C., 1993, Travail et usure mentale, Bayard éditions.
- DSM4, 1996, Manuel diagnostique et statistique des troubles mentaux, Masson.

- FAYADA C., VOLPE L., OUSS L. et BAKCHINE S., 1996, Patients neurologiques et approche systémique de la communication, Third europeean systems science congress, Rome, Oct 1996.
- FAYADA C., SARFATY Y., VOLPE L. and BAKCHINE S. (1996), Meta-representational deficit in patients with fronto-temporal lesions, Second congress of the European federation of neurological societies, Rome, Nov 1996.
- FUSTER J.M., 1995, Memory in the Cerebral Cortex, MIT Press.
- GRAFMAN J., 1989, Plans, actions and mentals sets : Managerial Knowledge Units in the Frontal Lobes. In E. Perecman (Ed), integrating theory and practice in clinical neuropsychology, (pp 93-138), Hillsdale, NJ : Lawrence Erlbaum Associates.
- JEANNEROD M., 1994, The representing brain : Neural correlates of motor intention and imagery, Behavioral and Brain Sciences 17:187-245.
- HOC J.M., 1996, Supervision et contrôle de processus, la cognition en situation dynamique, Presses Universitaires de Grenoble.
- LURIA A.R., 1966, Higher cortical functions in man, Consultants Bureau New York
- MESSULAM M., 1986, Frontal cortex and behavior, Ann. Neurol. ;19, 4 : 320-325.
- SHALLICE T., BURGESS P.W., 1993, Supervisory control of thought and action. In A. Baddeley, L. Weiskrantz (Eds), Attention: selection, Awareness and Control : a tribute to D. Broadbent, Oxford University Press.
- SHANON B., 1993, The Representational and Presentational, An Assay on Cognition and the Study of Mind, Harvester Wheatsheaf.
- TROGNON A., GHIGLIONE R., 1993, Où va la pragmatique?, De la pragmatique à la psychologie sociale, Presses Universitaires de Grenoble.

La notion de modèle et son utilisation en Psychologie

Jean Pierre ROSSI

*Université Paris-Sud
Centre National de la Recherche Scientifique
Groupe Cognition Humaine
LIMSI¹*

Le but de cet article est d'analyser la fonction du modèle dans la recherche en Psychologie. Une analyse historique montrerait probablement que le rôle du modèle dans la recherche croit avec le développement de la science. L'histoire de la physique illustre bien l'importance de la modélisation dans le développement d'une discipline. Le succès des modèles a entraîné une augmentation sensible de leur nombre, de leurs rôles et de leurs propriétés de sorte que progressivement le modèle devient un système suffisamment précis pour déboucher sur des simulations. C'est pourquoi, la définition de la notion est complétée par la description des

différentes catégories modèles et l'analyse de leurs propriétés réciproques. Cette analyse permet de montrer que les modèles formels permettant des simulations constituent des outils fondamentaux de la recherche actuelle. Les propriétés de ces simulations et leurs rôles dans la recherche seront donc analysées. Mais qu'il s'agisse de simulations ou de modèles généraux, les systèmes proposés ne prendront toutes leurs valeurs que dans la mesure où ils pourront être validés. Que signifie valider un modèle? Quelles sont les limites de ces validations? Telles sont les questions qui seront abordées dans la dernière partie du texte.

¹ LIMSI BP 133, 91 403 ORSAY Cedex, France. Tél : 01 69 85 80 16 Fax : 01 69 85 80 88, Email : Rossi@limsi.fr

1. Définition et classifications

1.1. Caractéristiques communes à l'ensemble des modèles

Les modèles élaborés par les chercheurs sont des représentations manipulables et validables. Le terme de représentation établit la distance entre le modèle et son objet. Le modèle est une description (pouvant être formelle) des propriétés ou qualités de l'objet qu'il représente. Mais cette représentation a pour propriété essentielle d'être manipulable et validable. Manipulable signifie que l'on peut modifier en vue d'expériences scientifiques. Le modèle comporte donc des éléments dont les propriétés peuvent être analysées, étudiées et modifiées. Au mieux, le modèle sera constitué de paramètres qui pourront être modifiés et adaptés en fonction des objectifs du chercheur. Cette manipulation est possible dans la mesure où le modèle est une forme simplifiée et généralement opérationnelle de l'objet qu'il représente. En effet le modèle est une simplification schématique du phénomène étudié. Il ne s'applique qu'à un domaine très spécifique, localisé et isolé. En ce sens, le modèle se présente souvent comme une réduction et une imitation simplifiée de la réalité qu'il est censé représenter.

Par ce qu'il est manipulable, le modèle pourra être l'objet non seulement de modifications et d'ajustements mais aussi de tests, de mesures ou de calculs qui ne sont pas commodément applicables à son objet. Il permet des prédictions qui pourront être validées par comparaison aux données réelles. Par validation on entend le processus qui permet d'établir la preuve qu'un mécanisme produit bien ce qu'il est censé produire ou qu'un instrument mesure ce qu'il est censé mesurer. Cette validation est le but de la démarche expérimentale.

Généralement le modèle est censé s'appliquer à un ensemble de données comme une structure ou un schéma général. En effet, même lorsqu'il est local le modèle décrit des structures ou des fonctionnements s'appliquant à des contenus différents. Les réseaux sémantiques (Quillian, 1967) sont construits pour décrire l'organisation du sens de tout ou partie du vocabulaire. Pour la catégorie de mots qu'il est censé représenter, la formalisation est indépendante des exemplaires ou instances. Si le réseau sémantique est limité à la représentation du contenu des mots concrets, les principes de son organisation seront valables pour l'ensemble des mots concrets.

1.2. De l'utilité des modèles dans la recherche en Psychologie

Si comme l'affirme Johnson-Layrd (1983) la construction de modèles de fonctionnement de l'esprit est le but de la recherche en Psychologie, l'importance des modèles en Psychologie n'est plus à démontrer. Au delà de cette affirmation, il est possible d'analyser les fonctions des modèles dans la recherche en distinguant les fonctions communes à toutes les catégories de modèles et les fonctions spécifiques de chaque catégorie. Deux fonctions générales peuvent être attribuées aux modèles : une fonction de synthèse et une fonction d'opérationnalisation.

La fonction de synthèse renvoie au fait que le modèle permet de rendre compte d'un ensemble de données. Le chercheur fait l'hypothèse qu'un ensemble d'observations est produit par des structures ou des fonctionnements communs. Il tente alors d'imaginer ces structures et ces fonctionnements. Le modèle constitue un cadre unificateur possédant une cohérence interne. Cette cohérence sera d'ailleurs l'objet de différentes validations.

L'opérationnalisation est une autre propriété des modèles. Elle consiste à donner une forme matérielle à des concepts. Cette opérationnalisation va permettre la manipulation, le test ou le calcul. Elle va amener à introduire des paramètres qui pourront être ajustés, modifiés, adaptés. Les conséquences de l'introduction de ces paramètres pourront être testés et évalués. Mais, surtout la matérialisation permet des prédictions sur lesquelles repose la validation du modèle.

La nécessité de donner une forme matérielle à des énoncés va, de plus, obliger le chercheur à être précis et donc à affiner sa conceptualisation au point de paramétrer les différents composants du modèle. Le niveau de matérialisation et donc de précision dépendra évidemment du type de modèle construit.

Le modèle est enfin un instrument de prédiction et de simulation. En effet le modèle ne se contente pas de rendre compte de certaines données il doit permettre de faire des prédictions sur ce qui doit se produire lorsque certaines circonstances sont rencontrées. Cette fonction de prédiction prend tout son sens lorsque le modèle est à l'origine de simulations.

1.3. Classification des modèles

Bien que tous les phénomènes pouvant être analysés et décomposés peuvent être l'objet d'un modèle (Tuggle, 1978) la variété des modèles reste limitée.

Les classifications varient selon que l'on considère l'objet modélisé, le but et la forme de la modélisation. Pour ce qui est des buts Tuggle (1978) distingue l'exploratoire du prédictif, descriptif du normatif, le modèle scientifique de modèle d'entraînement (essentiellement le jeu). pour ce qui est de la forme il oppose le discret du continu, le déterministe du probabiliste. En Psychologie, l'objet de la modélisation peut être de décrire une relation, une structure, un fonctionnement, ou même un ensemble constitué de structures et de processus.

Le but de la modélisation peut être de fournir un cadre (modèles-schémas) ou de permettre une simulation (modèles-simulations). Il peut être de rendre compte aussi fidèlement que possible des propriétés des objets d'intégrer le phénomène étudié dans un phénomène plus général.

La forme de la modélisation renvoie aussi bien au type d'analogie qui est choisi qu'à la forme des expressions qui sont utilisées. L'analogie peut renvoyer à un système physique : le coeur décrit comme une pompe ou le nerf comme un circuit électrique (Mouloud, 1995); à un système biologique : l'apprentissage est décrit comme une modification du réseau neuronal. Ces descriptions peuvent prendre la forme de dessins, de schémas ou même de dessins animés (Mouloud, 1995).

Les modèles peuvent aussi être formels. La représentation de structures et de fonctionnements psychologiques au moyen d'un langage logico-mathématiques constitue le niveau le plus élevé de la modélisation.

Le type d'unité utilisé dans le modèle a amené les Psychologues à distinguer l'approche symbolique de l'approche connexioniste. Dans un cas le modèle utilise des symboles dans l'autre il fait appel à des unités formelles non symboliques. Si cette distinction était pertinente au début du développement des modèles connexionistes elle tend à s'estomper dans la mesure où nombre de modèles connexionistes utilisés en Psychologie Cognitive associent des unités neuronales à des concepts. Le problème reste alors de définir ce qui fait l'originalité des modèles connexionistes. Rappelons que peut être qualifié de connexioniste tout modèle composé d'unités élémentaires recevant et propageant de l'activation par le biais d'un réseau de connexions.

L'outil de modélisation peut servir de critère de classification. Dans ce cadre, les modèles connexionistes constitueraient un bon exemple d'une classe

dé modèles. Une grande variété de modèles connexionistes sont actuellement développés. Cette diversification est liée tant à l'objet de la modélisation qu'à sa réalisation concrète. Des modèles connexionistes sont développés dans des disciplines aussi différentes que la Psychologie et l'économie. La diversité des réseaux est liée aussi bien au système de propagation de l'influx qu'à la fonction des neurones formels. En effet, les fonctions des unités élémentaires sont très diverses elles peuvent soit servir de simple relais soit procéder à des calculs complexes. Mais, tous ces modèles ont une structure de base commune : ils sont composés d'unités élémentaires interconnectées. La caractérisation de structures communes pourraient être un principe de classification des modèles.

Le principe d'organisation des modèles permet aussi de distinguer les systèmes modulaires des systèmes non modulaires (Fodor, 1983). Les modèles modulaires sont constitués d'entités autonomes fonctionnant selon des règles qui leur sont propres.

En référence aux pratiques de la recherche en Psychologie (Estes, 1991), il s'avère pertinent de reprendre la distinction entre "modèle-cadre" et "modèle hypothétique" proposée par Rouanet (1983). "Le "modèle-cadre" correspond à un dispositif, une maquette, que l'on construit aux fins de rendre opérationnelles les questions que l'on se pose sur les phénomènes étudiés. Le modèle hypothétique correspond à un mécanisme plus ou moins précis, qui exprime le fait qu'un certain processus psychologique se passe d'une certaine façon (et pas d'une autre)" (Lecoutre, Rouanet et Denhiere, 1991, p. 384). Dans l'inférence statistique le "modèle-cadre" est le modèle d'échantillonnage. "Les procédures statistiques seront centrées sur l'évaluation du modèle hypothétique" (Lecoutre, Rouanet et Denhiere, 1991, p. 385). Dans la recherche expérimentale le modèle cadre est le modèle décrit par le chercheur (modèle de fonctionnement...), le modèle hypothétique est le modèle d'analyse des données. Cette distinction est importante car elle permet de souligner que la recherche expérimentale consiste à recueillir des données qui permettent de valider le modèle d'hypothèse mais le modèle-cadre n'est en fait jamais directement validé, il est validé de façon indirecte. Ainsi, par exemple le chercheur qui travaille sur la compréhension du langage postule l'existence d'un lexique mental chez tous les individus. Il construit ensuite un modèle de ce lexique puis sur la base de ce modèle élabore des prédictions qui seront l'objet d'hypothèses statistiques et de validation. En réalité, seules ces hypothèses sont testées, le modèle ne l'est que par le biais des prédictions du chercheur quant à l'existence même de l'entité « lexique » elle reste hypothétique.

Si le modèle se présente comme une description, il est alors possible d'identifier plusieurs niveaux de description. L'exemple des modèles de tâches permet d'illustrer comment un même événement peut-être l'objet de différentes descriptions. Un modèle de tâche se définit comme étant une description de la tâche effectuée par des opérateurs. Cette description peut être limitée à l'énumération du but et des sous-but de la tâche (Sebillotte, 1991). Un niveau supérieur peut être atteint par la description des actions permettant de réaliser chacun de ces buts. Psychologues et Ergonomes ont élaboré différentes méthodes de description (Richard et Poitrenaud, 1991; Scapin et Pierret-Golbreich, 1991). Le choix d'une méthode définit le système de description. Un troisième niveau correspond à la description des opérations mentales mis en oeuvre pour réaliser chacun des sous tâches. Cette description peut être plus ou moins précise. Elle peut aboutir à l'écriture d'un algorithme décrivant avec précision les différentes procédures mis en oeuvre pour réaliser la tâche. Elle peut enfin déboucher sur une simulation de type informatique. Il s'agit alors de réaliser des programmes simulant tout ou partie de la tâche. Dans ce cas, chaque élément de l'algorithme est traduit sous la forme d'un programme informatique. Le choix du langage permettant la programmation est une nouvelle source de variation. Cet exemple illustre bien la variété des représentations qui sont qualifiées de modèles et la difficulté d'en proposer une classification utile.

En Psychologie Cognitive, la notion de modèle est utilisée dans quatre acceptions apparemment différentes mais en fait complémentaires. Schématiquement, le terme de modèle désigne l'une des quatre entités suivantes:

- une forme idéale ou forme standard (Chaplin, 1985);
- une représentation théorique : schéma ou architecture;
- une description utilisant un langage formel;
- un système concret de simulation.

La description de chacun de ces modèles permet une analyse de son rôle dans la recherche.

2. Le modèle, forme idéale ou standard

2.1. Description

Dans cette première acception, proche du sens commun, le modèle est une référence généralement normative qui peut être si ce n'est copiée du moins transposée. L'effet de l'apprentissage sur le réseau neuronal de l'*Aplysia Californica* (Richelle, 1991) est utilisé comme modèle des transformations neuronales déclenchées par un apprentissage. Le système neuronal de cet invertébré étant très simple, ses neurones étant peu nombreux et de grande taille il s'est avéré être un support idéal pour analyser les transformations induites par l'apprentissage et décrire ainsi un modèle de l'effet de l'apprentissage sur l'organisation neuronale. Ce modèle va servir de référence dans la mesure où il est représentatif d'un phénomène.

2.2. Rôle des modèles « forme idéale ou standard » dans la recherche

La forme idéale ou standard se présente comme une référence ou comme une norme. la référence est un système exemplaire qui s'applique à différentes situations. Ainsi, par exemple, le modèle homéostatique sert de référence dans de nombreux domaines d'étude en Psychologie. On peut ainsi considérer qu'un besoin est créé par un déséquilibre (la soif est déterminée, entre autre, par une différence de concentration en sodium entre le milieu intra cellulaire et le milieu extra cellulaire), ce déséquilibre crée une tension qui, étant difficilement supportable, doit être réduite par un comportement appropriée, dans le cas de la soif une prise d'eau. Ce modèle tend à être appliqué à toutes les motivations qu'elles soient physiologiques que psychologique. Le système homéostatique sert de modèle dans l'étude de la motivation. Dans ce cadre, le chercheur tente d'identifier l'origine, les manifestations et les causes de déséquilibre puis il décrit les comportements qui tentent à réduire la tension créée par le déséquilibre et le cas échéant les conditions de retour à l'équilibre. Le modèle sert de cadre et de système d'analyse pour l'étude des différentes motivations. Cette fonction de référence est évidemment utilisée pour découvrir les variations par rapport au modèle ou les situations où le modèle ne s'applique pas. La référence peut non seulement servir de cadre mais devenir une norme dans certaine conditions. C'est ainsi que les modèles décrivant le développement de l'enfant peuvent avoir un caractère normatif : description

d'un modèle de développement par rapport auquel le développement individuel sera apprécié.

3. Modèle schéma ou architecture

3.1. Description

Le modèle schéma est une représentation théorique. Comme toute représentation il décrit au moyen d'images ou de mots tout ou partie des caractéristiques de l'objet représenté. Leur but est de présenter sous forme d'un schéma architectures² du système étudié. Ils se présentent comme une suite de structures réunies par des flèches qui indiquent le sens de circulation des informations. Des boucles signalent les processus de contrôle et de feed-back. Les structures représentées peuvent être de natures différentes. Il peut s'agir de structures biologiques, de modules informatiques ou de structures psychologiques hypothétiques : lexique, mémoire de travail, mémoire à court terme...

3.2. Rôle des modèles schéma ou architectures dans la recherche

Dans l'histoire de la Psychologie, les modèles-schémas ont tenu une place importante dans les années 1960. Dès cette époque, et à la suite du développement de la théorie de l'information (Shannon et Weaver, 1949) le fonctionnement mental était conçu par les cognitivistes comme un système de traitement de l'information constitué par une série de sous-systèmes qui reçoivent des informations, les traitent parfois même les transforment et finalement fournissent une réponse qui elle-même peut servir d'entrée à un autre sous-système. Ces sous-systèmes ont souvent été présentés sous forme de boîtes, d'où le nom de boxologie pour désigner ce type de représentation, reliées par des systèmes de flèches comprenant parfois des boucles (feed-back). Ces représentations ont souvent été qualifiées de modèles. Leur propriété essentielle est de représenter l'ensemble des structures étudiées en soulignant leurs articulations. Les positions spatiales des structures renseignent sur l'ordre dans lequel elles interviennent. Les multiples flèches qui les unissent indiquent

² Le terme d'architecture est utilisé ici en référence à l'informatique où l'architecture décrit la structure du système.

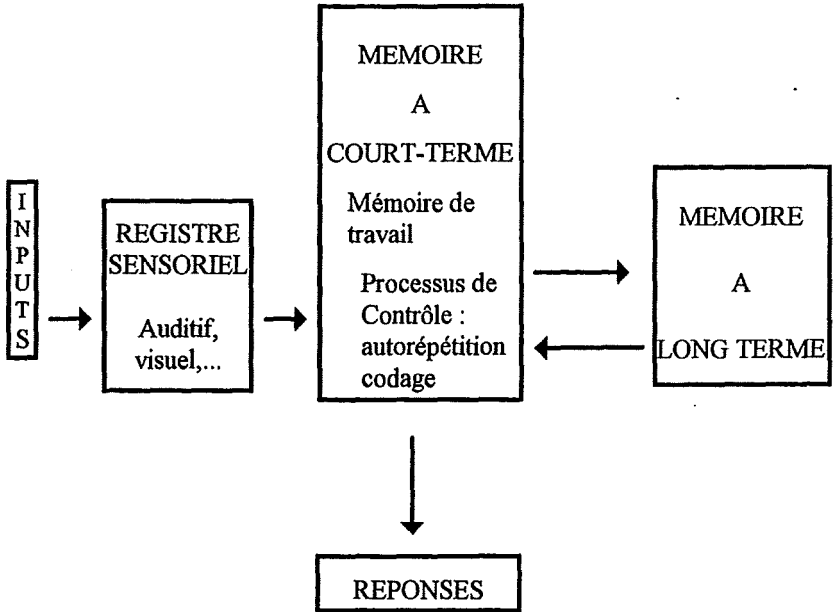


Figure N°1 : Modèle de la mémoire selon Atkinson et Shiffrin (1971)

le sens et les systèmes de contrôle mis en oeuvre. La fonction schématique des modèles de ce type est illustrée dans le schéma reproduit dans la figure 1 qui décrit une conception de l'organisation des différentes mémoires. Ce modèle d'organisation peut permettre des prédictions qui pourront être validées au moyen de données expérimentales. Ces prédictions portent aussi bien sur l'existence des structures décrites, leurs fonctions ou leur mode d'organisation que leur position réciproque. On peut ainsi s'interroger sur l'existence des registres sensoriels, sur le rôle de l'autorépétition ou sur l'action de la mémoire à long terme sur la mémoire de travail. Le cadre ainsi tracé peut aussi être à l'origine de différentes simulations. C'est ainsi, que le modèles-schéma est parfois utilisé pour représenter l'architecture d'un système informatique complexe. La figure N°2 présente un exemple tiré d'un article de Masson (1995) dans lequel l'auteur présente une simulation connexionniste de l'identification des mots. Le schéma est utilisé pour indiquer comment l'ensemble est organisé. Trois modules sont prévus, l'un traite du sens tandis que les deux autres analysent les propriétés perceptives du stimulus visuel ou auditif. Les flèches indiquent que l'aces au sens peut se faire selon les deux modalités (visuelles, auditives) et que les deux systèmes sont interconnectés. La différenciation entre les trois structures est établie expérimentalement (Rossi, 1981; McRae, de Sa et Seidenberg, 1997) de même que la possibilité d'un accès direct de l'écrit au sens sans codage phonologique a été l'objet de nombreuses recherches (Rossi, 1981).

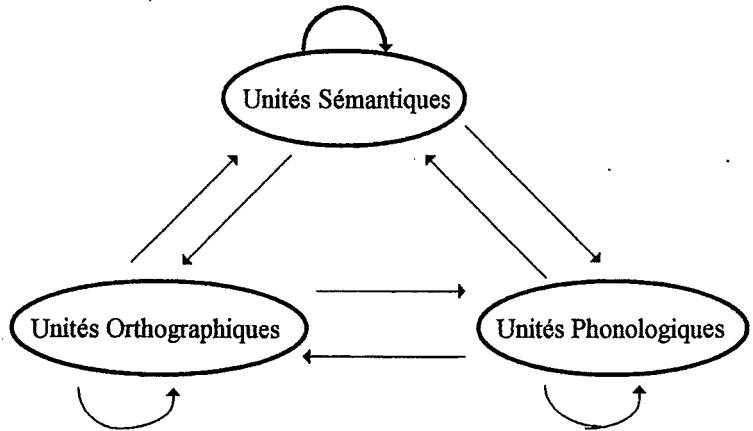


Figure N°2 : organisation des modules de traitement dans le modèle de dénomination des mots (Masson, 1995)

4. Modèles formels

4.1. description

Un système formel est un langage composé de formules et de règles de démonstration. Ses propriétés essentielles sont d'être non ambiguë et de permettre l'élaboration d'expressions applicables à différents contenus. L'équation $y=ax+b$ définit une droite quelque soient les valeurs de a , de x et de y . De même, une grammaire formelle décrit une syntaxe applicable à différents contenus (Grise, 1991).

La plupart des modèles formels utilisent le langage mathématiques. Ce langage sert soit pour décrire une relation au moyen d'une équation soit pour réaliser des simulations complexes comprenant différentes composantes. Ce dernier cas sera traité dans le cadre des simulations. Les modèles mathématiques réduits à l'expression d'une simple équation ont été développés en Psychologie dès la naissance de la Psychophysique. La loi de Fechner

(1860) selon laquelle la sensation est proportionnelle au logarithme de la stimulation est une description mathématique de la relation entre stimulation et sensation. Elle s'exprime sous la forme de l'équation : $y = K \log x + h$ dans laquelle y correspond à l'intensité de la sensation, x l'intensité de la stimulation tandis que K et h sont des constantes de proportionnalité. La forme de cette description est mathématique, c'est un modèle mathématique.

4.2. Rôles des modèles formels dans la recherche

Le rôle de la simulation informatique faisant l'objet d'un chapitre particulier, nous nous limiterons aux formalisations qui se traduisent par la mise en relation entre différents indicateurs dont la loi de Fechner est un bon exemple.

L'établissement d'une loi de ce type va générer des recherches qui permettront d'évaluer sa validité et les conditions de son application. La loi de Fechner est-elle aussi valable pour les intensités faibles que pour les intensités moyennes ou fortes? Cette loi s'applique-t-elle à l'ensemble des sensations : vision, ouïe, goût... Aude là de l'étude de la validité de la relation la loi exprimée sous la forme d'une équation présente en Psychologie deux intérêts principaux : a) permettre des prédictions; b) donner un cadre, pour l'étude des structures et des fonctionnements. La fonction de prédiction est évidemment la plus simple puisqu'il est possible de calculer la valeur d'une inconnue lorsque l'on connaît les valeurs des autres termes. La seconde fonction est plus complexe puisqu'il s'agit de découvrir au niveau de l'organisation des structures et des fonctionnements les implications des relations décrites par l'équation. Dans l'exemple de la loi de Fechner il s'agit de savoir pour quelles raisons l'intensité de la sensation croît comme le logarithme de la stimulation. Qu'est ce que cela implique au niveau du traitement de l'information et des structures qui produisent la sensation. dans ce cadre l'établissement de la loi n'est qu'une étape dans le processus d'analyse du phénomène étudié.

5. Validation des modèles et simulations

5.1. La notion de validation

Le modèle comme la théorie est une construction hypothétique et donc susceptible de validation. Valider consiste à déterminer si le modèle décrit

correctement ce qu'il est censé décrire : le modèle de la mémoire décrit-il correctement la mémoire? La notion de validité renvoie donc au système de description qui peut être physique, physiologique, psychologique... En Psychologie Cognitive la validation des modèles consiste à déterminer si les structures et les processus décrits par le modèle rendent compte de la façon dont l'information est traitée par le système cognitif. Dans la mesure où le cerveau est le support biologique de la pensée, le chercheur en Psychologie peut avoir pour ambition de mettre au point des modèles ayant une validité psychologique et biologique c'est à dire rendant compte des processus psychologique et respectant les règles de fonctionnement du cerveau. Le développement des modèles connexionistes qui prennent des apparences biologiques dans la mesure où ils reposent sur l'organisation de réseaux dits "neuronaux" pourraient permettre de réaliser un tel projet. En fait, le fonctionnement des neurones formels utilisés dans ces modèles est souvent très éloigné de celui des neurones "biologiques".

Quelque soit le système de description choisi, il est nécessaire de distinguer une validité que l'on peut qualifier « d'interne » dans la mesure où elle est liée à la cohérence de la description et une validité qui peut être qualifiée « d'externe » ou d'empirique dans la mesure où elle fait appel à l'utilisation de données empiriques. La validité interne comprend deux composantes. La première concerne la cohérence de la description et le respect des règles du langage adopté. Si le langage est mathématique, le non respect d'une règle de ce langage permettra de réfuter le modèle. La seconde a trait au respect des connaissances acquises : le modèle ou certains aspects de la description contredisent-ils des faits établis et si c'est le cas comment le chercheur justifie-t-il ces écarts?

La validation externe renvoie à la confrontation du modèle aux données empiriques. La référence à ces données a déjà été évoqué à propos de la conformité du modèle aux faits établis mais cette fois il s'agit de prédire les conséquences empiriques du modèle et de ses composantes. Le raisonnement qui est développé est de type hypothético-déductif, il prend la forme suivante « Si le modèle est correct alors dans telle situation il doit se passer tel phénomène ». Le phénomène doit se traduire en observables. L'expérimentation permet de produire des observables qui seront comparés aux prédictions. La puissance du modèle dépend de sa capacité à permettre des prédictions non ambiguës, non équivoques. Les prédictions concernent aussi bien la totalité du modèle qu'une de ses composantes. Soulignons que la

validation du modèle passe par un raisonnement débouchant sur des prédictions, ce raisonnement nécessite parfois l'énoncé d'hypothèses intermédiaires qui viennent compliquer la validation.

Dans le cas des modèles schéma ou architecture les prédictions portent soit sur les structures qui sont décrites, soit sur l'organisation du système. Pour ce qui est des structures on peut tenter de valider leur existence, leurs fonctions et leurs organisations. Pour ce qui est de l'architecture générale on peut s'intéresser à la position des différentes structures et aux relations qui les unissent.

Comme il était souligné plus haut les prédictions seront d'autant plus aisées à élaborer que le modèle sera précis. Les descriptions opérationnellement les plus précises devraient pouvoir permettre la réalisation de simulations. C'est pourquoi la simulation est un outil de validation particulièrement riche.

5. 2. *Simulations*

5.2.1. Description

Simuler consiste à reproduire au moyen de systèmes physiques (généralement informatiques) l

es descriptions développées dans le modèle théorique. En ce sens, simuler revient à construire un système permettant d'opérationnaliser un modèle théorique.

Les premières simulations furent des automates qui évitaient les obstacles lors de leurs déplacements. Ils reproduisaient ainsi la programmation du déplacement en fonction des informations visuelles. Ils étaient constitués d'un système de détection des obstacles qui transmettait des informations aux appareils qui géraient les déplacements. Ils furent ainsi les ancêtres des simulateurs de la perception et du déplacement dont le « Perceptron » (Minsky, 1969) est l'exemple le plus connu tandis que « Darwin III » (Edelman, 1987) est la forme la plus évoluées. « Darwin III » est un automate de reconnaissance possédant un oeil, un bras mobile sensible au toucher et un système neuronale qui gère à la fois les déplacements de l'oeil et l'adaptation du bras en fonction de la taille, du poids et de la nature des objets.

La simulation ne se réduit pas à la construction d'automates. Le développement de l'informatique a mis à la disposition du chercheur en

Psychologie un outil de simulation très puissant qui lui permet d'aborder des domaines aussi complexes que le traitement du langage.

La Psychologie Cognitive se situant dans le cadre de l' « information processing » dont l'intelligence artificielle est l'héritière directe considère que les processus cognitifs peuvent être réduits à des systèmes de traitement de l'information. Dans ce cadre les modèles sont constitués de structures et de procédures (Jackendoff, 1987). Les simulations reviennent donc à construire des systèmes informatiques qui reproduisent les structures et procédures théoriques décrites dans le modèle. Le chercheur fournit des informations au système qui produit des réponses. La simulation met en oeuvre des procédures de transformation des « inputs » en « outputs », elle produit des réponses. Ainsi, par exemple, McClelland et Rumelhart (1985) ont construit un système qui simule la lecture c'est à dire reconnaît les mots écrits et active les significations. Le développement des études en intelligence artificielle témoigne du succès de ces approches. La mise au point de systèmes susceptibles d'apprentissage (réseaux de neurones) a augmenté l'intérêt que le chercheur en Psychologie peut porter aux simulations informatiques.

Toute simulation informatique utilise des fonctions mathématiques. Le problème se pose alors de savoir si une équation qui peut avoir le statut de modèle formel peut être considérée comme une simulation. La loi de Fechner qui exprime sous forme d'une équation la relation entre l'intensité de la stimulation et l'intensité de la sensation est elle une simulation? Plus généralement une loi peut-elle être considérée comme une simulation? En Psychologie Cognitive, il est préférable de réserver le terme de simulation pour décrire des systèmes physiques ou informatiques comportant des structures et des procédures tandis que la loi est une formule établissant une relation entre deux termes. Cette loi pouvant être validée expérimentalement.

Le passage du modèle théorique à la construction d'un système de fonctionnement n'est possible que si la description développée dans le modèle est précise et univoque. Un modèle imprécis pourra être à l'origine de simulations contradictoires. La construction d'une simulation n'est possible que si les structures et les traitements sont décrits avec précision.

L'opérationnalisation du modèle va permettre au chercheur d'identifier les passages obligés c'est à dire les opérations ou les structures indispensables aux traitements qui doivent être effectués et ce n'est pas le moindre avantage de la simulation. C'est à ce stade que le chercheur va rencontrer de nouveaux

problèmes l'amenant souvent à rajouter des postulats ou des hypothèses qui rendront d'autant plus difficile la validation.

Le principal avantage de la simulation est d'être une maquette dont chaque constituant pourra être l'objet de calculs, de tests, de mesures. Le chercheur va pouvoir procéder à des essais, analyser des résultats puis ajuster les systèmes ou les paramètres. Cette possibilité de modification, de manipulation et en définitive d'ajustement est évidemment essentielle dans la mesure où le système cognitif n'est pas directement manipulable. Il n'est pas possible de manipuler directement le lexique mental d'une personne en revanche, la simulation peut être l'objet de différents tests. Le chercheur va pouvoir tester différentes solutions et observer les conséquences pratiques de chacun de ses choix. Dans la simulation le chercheur trouve un moyen d'opérationnaliser ses conceptions et de les valider.

5.2.2. Validation des simulations

Les principes de validation développés précédemment s'appliquent évidemment à la simulation. La validation dite externe ou empirique concerne aussi bien le résultat final du système ou «ouput» que les résultats intermédiaires. Cette validation consiste à comparer le résultat produit par le système à des observables soit recueillis directement soit produit au moyen d'une expérience. Cette comparaison suppose l'équivalence des deux mesures : la mesure produite par la simulation et la mesure produite par l'expérimentation. Or, il n'est pas rare que les unités de mesure ne soient pas équivalentes et que le système de simulation produise une mesure correspondant à une échelle d'intervalle tandis que les données produites par l'expérimentation correspondent à une échelle ordinale. Ce problème est illustré par l'étude du modèle "intégration-construction" de Kintsch (1991). Ce modèle a pour ambition de prédire les informations importantes du texte. Pour ce faire un système néo-connexioniste est construit dans le but d'associer une valeur numérique à chaque proposition. Ces valeurs donnent le poids de chaque proposition. Ce poids peut être considéré comme un indicateur de l'importance de la proposition. Il permet de hiérarchiser les propositions sur la base de leur importance. Le système produit un continuum numérique or les études expérimentales indiquent que le lecteur a tendance à constituer des classes correspondant à des niveaux d'importance : informations très importantes,

informations moyennes importantes, informations secondaires, informations peu importantes (Rossi, Bert-Erboul, 1989). La simulation fournit un continuum numérique correspondant à une échelle d'intervalle tandis que le lecteur classe en catégories ordinales. Il s'agit alors de savoir quelles valeurs numériques peuvent être considérées comme équivalentes. La difficulté est d'autant plus grande que l'échelon différentiel varie selon la zone où l'on se situe dans le continuum. Cet exemple illustre les difficultés liées à la comparaison entre le résultat produit par le système de simulation et le résultat observé.

Lorsque la simulation est un système complexe comportant de nombreuses étapes il est souvent possible de valider les traitements intermédiaires c'est à dire de comparer le résultat d'une de ces transformations au résultat observé dans une situation où le sujet a pour seule tâche d'effectuer cette transformation. La validation de ces résultats pose les problèmes déjà évoqués mais s'y ajoute les questions liées au fait d'isoler des traitements. A supposer que les traitements intermédiaires soient correctement identifiées, encore faut-il s'assurer que le traitement étudié est effectué de façon identique lorsqu'il est isolé ou lorsqu'il est intégré dans une tâche complexe. Prenons l'exemple d'une tâche de décision faisant intervenir dans une première étape une classification des informations. Peut-on faire l'hypothèse que la classification s'effectue de la même façon lorsqu'elle est effectuée pour elle-même (c'est à dire lorsqu'elle est la tâche) que lorsqu'elle est intégrée dans la tâche complexe? Cette question pose le problème de la modularité du système, de l'autonomie des modules de traitement et donc de la possibilité de valider indépendamment chaque module.

La correspondance entre le résultat produit par la simulation et ceux produits par l'expérimentation n'est pas un garant de la validité psychologique du modèle c'est à dire du fait que le sujet effectue mentalement les opérations décrites par le modèle. Dans la mesure où un même résultat peut être produit par des procédures différentes il est clair que la simulation peut produire des résultats similaires à ceux enregistrés par le chercheur sans utiliser les mêmes procédures. Sans oublier que des sujets peuvent produire des résultats identiques en utilisant des procédures différentes (Hollman, 1990) et compte tenu des richesses et la flexibilité du système cognitif un même sujet peut produire des réponses identiques en mettant en oeuvre des procédures différentes. La vicariance inter ou intra individuelle (Reuchlin et Bacher 1989), c'est à dire la possibilité de réaliser une même tâche selon des procédures différentes est un thème de recherche qui mériterait d'être plus développé en Psychologie.

Les critères permettant d'apprécier la validité psychologique des procédures mises en oeuvre par la simulation n'ont rien de mathématiques. Il s'agit essentiellement d'analyser la conformité du fonctionnement du système aux connaissances en psychologie. Dans l'exemple du modèle « intégration-construction » le chercheur est amené à se demander si le système cognitif produit des calculs similaires de ceux décrits dans la simulation. Il est clair, par exemple, qu'il est peu probable que le système cognitif multiplie les itérations.

L'analyse des formalisations mathématiques utilisées dans les modèles de simulation informatique est au centre de la validation interne du modèle. Dans ce cadre quatre nouvelles questions peuvent être posées : 1) Quelles propriétés mathématiques possède le système? 2) Quels paramètres ont été fixés? 3) quelles sont les conséquences des choix effectués? 4) Quels sont les corrélats psychologiques des propriétés analysées précédemment? L'intérêt de ces questions peut être illustré par l'étude du modèle « construction-intégration ». L'exemple d'une propriété mathématique peut être trouvé dans l'analyse de la matrice de connexion. Pour attribuer des coefficients à chaque proposition le système doit procéder à des calculs par itération jusqu'au moment de la stabilisation. Le problème est de savoir s'il existe des conditions dans lesquelles le système ne se stabilise pas. L'analyse mathématique a permis de mettre en évidence que dans certaines conditions le système ne se stabilisait pas et donc itérait indéfiniment. Ce constat est en soi intéressant si l'état de non stabilisation a un corrélat psychologique, dans le cas présent cela correspondrait à l'impossibilité de hiérarchiser les propositions. Or, pratiquement un lecteur peut toujours hiérarchiser les informations qu'il lit. Les propriétés mathématiques qui déterminent la non stabilisation du système n'ont pas de corrélats psychologiques.

L'analyse des paramètres peut s'avérer tout aussi intéressant. Dans la matrice de connexion proposée par Kintsch le lien entre deux propositions qui partagent un argument est exprimé par une valeur numérique. La première question est de savoir ce qui justifie le choix de cette valeur. La seconde concerne les conséquences de ce choix. Ainsi, dans l'exemple précédent, le fait que les valeurs soient positives rend le système non orienté de sorte que deux liens opposés se trouvent codés de la même façon. Il est possible de proposer une modification du paramètre par l'introduction de valeurs négatives. Les conséquences de ces modifications sur l'ensemble du système doivent être alors réévaluées.

Tous système de traitement de l'information repose sur la transformation d'unités que l'on qualifie d'unités de traitement. Le problème de la validité des unités de traitement est alors posé. Il peut être illustré par les quatre questions suivantes : 1) Quelles propriétés du matériel sont reflétées par les unités construites? Il s'agit de savoir quels aspects sont pris en compte et quels aspects sont négligés. 2) Quels sont les règles de construction de ces unités? 3) Ces règles de constructions peuvent-elles être mises en oeuvre par un système psychologique? 4) L'unité ou les unités utilisées sont-elles des unités psychologiques?. L'étude du langage permet d'illustrer la portée pratique de ces questions. Le mot est utilisé comme unité de traitement dans les premiers réseaux sémantiques (Quillian, 1967). La proposition, unité constituée d'un prédicat et de ses arguments a été proposée comme unité supérieure. En 1972, Rumelhart et al. propose de constituer des entités complexes à savoir l'événement et l'épisode. Si le système cognitif travaille à ces quatre niveaux (mot, proposition, événement et épisode) il est nécessaire de décrire avec précision les propriétés de ces unités et leurs règles de construction (l'identification de l'unité « mot » n'est pas toujours simple). Avant même la description le chercheur doit s'interroger sur la validité psychologique de ces unités et présenter les arguments qui autorisent à affirmer que le système cognitif les utilise. La discussion sur le fait que les représentations des connaissances soient de type propositionnelle n'est à ce jour pas close (Guha et Rossi, 1997).

A supposer que ces différentes validations aient abouti positivement, c'est à dire que le modèle soit provisoirement validé, il reste à préciser ce qu'il nous apporte pour la connaissance du phénomène étudié. Un système qui simule correctement un phénomène est un système qui met en oeuvre, des structures, des processus et des systèmes de liaisons pertinentes. De ce fait, le modèle constitue un cadre qui peut être à l'origine de nouvelles hypothèses et de nouvelles recherches permettant de l'améliorer soit en augmentant la précision des prédictions qu'il peut générer, soit en intégrant de nouveaux composants soit enfin en élargissant son champ d'action. On ne saurait pourtant oublier qu'un modèle peut généralement être opérationnalisé de différentes façons et que le choix d'une simulation particulière amène à formuler de nouvelles hypothèses ou à poser de nouveaux postulats.

Conclusion

L'analyse critique qui vient d'être développée à propos de la validation des simulations pourrait amener à douter de l'intérêt de la modélisation. Or, la modélisation est un outil déterminant du développement de la recherche. En Psychologie, l'élaboration de ce que nous avons nommé les modèles-schémas a permis de passer de l'étude des variables à l'étude des fonctionnements et des structures. La construction de simulations a permis d'opérationnaliser concrètement les modèles-schémas et donc à passer du stade des représentations verbales à celui de la réalisation concrète de systèmes de production d'«outputs». Ce double mouvement a des conséquences épistémologiques évidentes. L'élaboration des modèles consacre un changement radical des perspectives de recherche tandis que la simulation fournit un moyen de concrétisation indispensable à la validation de conceptions théoriques. Le modèle devient donc une source d'hypothèse et de production de données. En Psychologie cognitive, la modélisation va permettre d'organiser la recherche autour de thèmes unificateurs que sont l'organisation et le fonctionnement des structures cognitives. Puisse le développement des outils de simulations nous permettre d'améliorer les approximations rudimentaires que sont les modèles actuels!

Bibliographie

Atkinson, R.C. et Shiffrin, R.M. (1971), The control of short term memory, *Scientific American*, 225, 82-90.

Chaplin, J.P., (1985), *Dictionary of Psychology*, New-York : Bantam Doubleday Dell Publishing.

Craik, K. J. W., (1943), *The nature of explanation*, Cambridge : Cambridge University Press.

Edelman, G.M., (1987), *Neural Darwinism : The Theory of Neural Group Selection*, New-York : Basic Books.

Estes, W. K., (1991), *Statistical models in behavioral research*. Lawrence Erlbaum Associates : Hillsdale, New-jersey.

Fechner, G., (1860), *Elemente der Psychophysik*. Leipzig : Breitkopf und Hartel.

Fodor, J. A., (1983), *The modularity of mind*. Cambridge Ma. : The MIT Press.

- Grize, J. B., (1991), Formel in Roland Doron et Françoise Parot, Dictionnaire de Psychologie, Paris : PUF
- Guha, A. et Rossi, J. P. (1997). *A comment of « construction-integration » model*. Document polycopié, LIMSI.
- Jackendorff, R. (1987). *Consciousness and the computational Mind*. Cambridge University Press.
- Johnson-Laird, P.N. (1983) - *Mentals Models*, Cambridge University Press.
- Kintsch, W. (1991) - The role of knowledge in discourse comprehension : a construction-integration model in G. Denhières et J.P. Rossi (Eds.) *Text and text processing*. Amsterdam : Elsevier Publishers.
- Lecoutre, B., Rouanet, H. et Denhière, G. (1991). L'inférence statistique comme instrument de validation des modèles in J.P. Caverni, C. Bastien, Psychologie. Mendelsohn et G. Tiberghien (Eds.), *Psychologie Cognitive Modèles et Méthodes*, Presses Universitaires de Grenoble.
- McRae, K, de Sa, V. R. et Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology : General*, 126, 2, 99-130.
- Masson, M. E. J., (1995). A distributed Memory Model of Semantic priming, *Journal of experimental Psychology : Learning, Memory and Cognition*, 21-1, 3-23.
- Minsky, M., (1969). *Perceptrons*, Cambridge, MA : MIT Press.
- Mouloud, N., (1995). Modèles, in *Encyclopédia Universalis*. Paris : Encyclopédia Universalis.
- McClelland, J.L. and Rumelhart, D.E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology : general*, 15, 114-162.
- Ohlman, T. (1990). Les systèmes perceptifs in M. Reuchlin, J. Lautrey, C. Marendaz et T. Ohlmann, *Cognition : l'individuel et l'universel*. Paris : PUF.
- Putman, H., (1975), *Minds and machines*, in H. Putman (ed.), Cambridge : Cambridge University Press.
- Quillian, R. (1967). Words concepts: a theory and simulation of some basic semantic capabilities, *Behavioral Science*, 12, 403-417.
- Reuchlin, M. & Bacher, F. (1990). Les différences individuelles dans le développement cognitif de l'enfant. Paris : PUF.
- Richard, J.F. et Poitrenaud, S. (1991), Problématique de l'analyse des protocoles individuels d'observations comportementales in J.P. Caverni, C.

Bastien, Psychologie. Mendelsohn et G. Tiberghien (Eds.), *Psychologie Cognitive Modèles et Méthodes*, Presses Universitaires de Grenoble.

Richelle, M., (1991), Modèle in Roland Doron et Françoise Parot, *Dictionnaire de Psychologie*, Paris : PUF

Robert, P., (1995), *Dictionnaire de la langue française*. SNL Le Robert: Paris.

Rossi, J.P. (1981), *Les mécanismes de la lecture*. Paris : Sorbonné.

Rossi, J. P. (1997), *La recherche expérimentale en Psychologie*. Paris : Dunod.

Rossi, J.P. & Bert-Erboul, A. (1991). Sélection des informations importantes et compréhension de textes. *Psychologie Française*, 36, 2, 135-142.

Rouanet, H. (1983). Modèles en tous genres et pratiques statistiques. In R. Giglione (Ed.), *Comprendre l'homme construire des modèles*. Paris : Editions du CNRS, *Comportements*, 4, 113-124.

Scapin, D.L. et Pierret-Golbreich, C. (1990). Towards a method for task description : MAD, in L. Berlinguet et D. Berthelette (Eds.), *Work with display units 89*, Elsevier Sciences Publishers, North-Holland.

Sebillotte, S. (1991). Décrire des tâches selon les objectifs des opérateurs, de l'interview à la formalisation. *Le Travail Humain*, 54-3, 193-223.

Shannon, C.E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana : University of Illinois Press.

Tuggle, F. D., (1978). Theory content and explanatory power of simulation models. *Behavioral Sciences*, 23, 271- 290.

Les modèles de la Vie Artificielle et le(s) concept(s) d'émergence : de la métaphore à la réalisation

Miriam Edith Scaglione
Institut de hautes études en administration publique*
MiriamEdith.Scaglione@idheap.unil.ch

7 janvier 1998

Résumé

L'objectif de ce texte est tout d'abord de montrer que les démarches qui considèrent les propriétés émergentes comme non-déductibles d'aucune des propriétés sous-jacentes du système aboutissent à une conception superflue de l'émergence. Ensuite, discuter de certains usages du terme émergence dans la vie artificielle pour finalement, attirer l'attention sur le fait que, dans le cadre de la vie artificielle, certaines définitions de l'émergence sont davantage un outil pour mesurer l'ampleur d'une métaphore qu'un concept ayant un contenu philosophique précise.

1 L'émergentisme et les propriétés relationnelles

Un des héritages de l'émergentisme britannique est de considérer que les propriétés ou entités émergentes ne sont pas déductibles voire prévisibles des propriétés des parties. Dans [Scaglione, 1996] je montre que cette caractéristique, qui s'avérait pourtant centrale pour les émergentistes britanniques du début du siècle, devient irrationnelle à la lumière des progrès scientifiques de nos jours.

Je montre dans mon texte que tous les essais que l'on a faits pour rendre compatible le concept de réduction de C. D. Broad avec différents concepts

*Fondation associée à l'Université de Lausanne et à l'École polytechnique fédérale de Lausanne (Suisse).

techniques de réduction comme celui de Ernest Nagel, de Robert Causey et de Robert Cummins, doivent aussi être récusés.¹

Dans ce texte, en revanche, je vais à mettre en rapport le concept d'émergence avec celui de non-déductible *en prenant ce dernier de manière intuitive*.

Le texte de Paul Teller² est une source intéressante de réflexion à ce sujet.

1.1 Sur l'existence des propriétés non relationnelles

Dans quelle mesure est-il possible de donner une définition de propriété émergente pour une propriété globale d'un système comme étant non déductibles des propriétés relationnelles/non relationnelles³ de ses parties ? Prenons la première définition suivante :

DEFINITION 1 (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.

On peut aussi énoncer la définition alternative suivante de propriété émergente :

DEFINITION 2 (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*. Essayons de comparer ces deux définitions à la lumière de différents exemples.

¹L'idée de telles démarches était la suivante : premièrement, montrer la compatibilité de la réduction selon Broad avec la réduction des auteurs cités ci-dessus, deuxièmement montrer que, puisque les concepts de réduction coïncident, il y a encore lieu de parler des phénomènes émergents comme irréductibles à la Broad, dans la philosophie des sciences de nos jours. Ces démarches ont échoué soit parce qu'elles négligent la puissance du concept d'explication ou réduction chez Broad (par exemple dans le cas de Nagel), soit parce qu'elles rendent encore plus strict le concept de réduction de Broad (par exemple dans le cas de Causey). Or, il n'y a pas de parallélisme entre la réduction de Broad et une autre conception de réduction actuelle qui laisserait une plage aux phénomènes émergents à la Broad et justifierait ainsi leur caractère irréductible.

²[Teller, 1992]

³Rappelons-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 de". Les propriétés non relationnelles sont celles qui peuvent être décrites comme des propositions d'ordre égal à 1, par exemple "être rouge", "être un électron". Ces dernières propriétés sont aussi appelées *intrinsèques*.

Emergence des propriétés globales et relationnelles de la totalité :
Prenons l'exemple suivant :

EXEMPLE 1

Soit une boîte à crayons ; considérons la propriété "être le crayon le plus long de la boîte". [Teller, 1992, cf. page 141]

La propriété "être le crayon le plus long de la boîte" est bien une propriété globale du système ("la boîte de crayons") et elle est bien une propriété relative à la totalité. Selon l'énoncé 2, cette propriété relationnelle ("être le crayon le plus long de la boîte") est émergente puisque le fait de satisfaire la propriété "d'être le plus long" ne peut pas être déduit à partir des seules propriétés des crayons prises de façon isolée.

Cependant, elle ne vérifie pas la condition d'émergence selon l'énoncé 1. 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.⁴ Analysons maintenant la portée des deux définitions. Le premier énoncé exige aux propriétés émergentes le caractère irréductible à tout type des propriétés soient-elles relationnelles ou non relationnelles des parties du système. Le second énoncé, en revanche, demande que cette irréductibilité ne se vérifie que pour les propriétés non-relationnelles des parties. Je me propose d'entreprendre l'analyse comparative à partir de deux perspectives différentes. La première sera l'analyse de la richesse des concept d'émergence que chacune de ces deux définitions sous-entend, la seconde est, le caractère opérationnel de ces deux définitions.

La première perspective d'analyse : Selon l'énoncé 2, une propriété émergente est une propriété d'une totalité qui n'appartient à aucune des parties prises de manière isolée. Autrement dit, une propriété émergente n'est point identique à une des propriétés non relationnelles des parties. Selon cette définition, une propriété émergente sera une propriété globale non-héritée des propriétés des parties. Les propriétés d'une totalité héritées des propriétés non relationnelles des parties étaient appelées, par les émergentistes britanniques, propriétés *résultantes* du système. Il en découle que l'énoncé 2 met l'accent sur le caractère global ou gestaltiste des propriétés émergentes. Il faut bien noter que selon le second énoncé, rien n'empêche qu'une propriété émergente soit explicable à partir des propriétés *relationnelles* des parties, comme le montre bien l'exemple de notre boîte à crayon.

⁴En 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.

Dans le premier énoncé, en revanche, l'accent est mis sur le caractère irréductible des propriétés émergentes à partir de toutes les propriétés des parties soient-elles relationnelles ou non. Cet énoncé souligne l'impossibilité absolue d'explication des propriétés émergentes.

La seconde perspective d'analyse : L'exemple de mon analyse, notre boîte à crayons est un système si simple que l'on peut trouver toutes les propriétés qui la caractérisent. Cependant dans des systèmes plus complexes la condition d'application de l'énoncé 1 se complique comme les sections suivantes le montrent. Dans quelle mesure peut-on dire qu'une propriété est émergente au sens d'irréductible ou inexplicable des propriétés du système hors du cadre d'une théorie ? En effet, comment déterminer de manière intuitive l'ensemble de toutes les propriétés qui caractérisent un système donné ? Une réponse pourrait être que l'on peut exiger d'une propriété émergente qu'elle ne soit pas explicable à partir des propriétés des parties connues ou explicites dans le cadre des toutes les théories que l'on connaît jusqu'à présent. Cependant, ceci n'est qu'une fuite en avant, en effet, l'émergence ainsi conçue devient la salle d'attente des propriétés que l'on n'a pas pu expliquer jusqu'à présent. Si une propriété globale d'un système est émergente, elle doit l'être indépendamment de l'état où se trouve la connaissance scientifique. Il me semble que cette conception de l'émergence est totalement stérile et doit être récusée.

Essayons maintenant de caractériser une propriété émergente comme irréductible à toutes les propriétés d'un système donné dans le cadre d'une théorie. Ceci signifie que ladite propriété n'appartient point au domaine d'explication de cette théorie. Est-il pertinent d'utiliser le terme d'émergence pour des propriétés étrangères aux explications d'une théorie donnée ? Il me semble que non, parce qu'alors on est en train de vider le concept de sa substance et de le rendre soit totalement inutile, soit inerte dans toute démarche explicative.

Le second énoncé me semble pertinent parce que l'émergence sert à l'enrichissement des théories des systèmes. En effet, selon cette définition il y a dans tout système quelques propriétés qui sont émergentes parce qu'elles sont globales ou gestaltistes. Le fait que ces propriétés puissent être expliquées à partir des propriétés de ses composantes ou des couplages de ces dernières, permet d'avantage de connaissance sur le système. Cette dernière conception est l'opposé de l'énoncé 1 dont le caractère d'irréductible va à l'encontre du progrès scientifique.

Le reste de l'analyse va se faire en adoptant l'énoncé 2.

Emergence des propriétés non relationnelles et survenantes de la totalité : Soit la propriété non-relationnelle (en principe) "d'être une calculatrice" ou "une clé".

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 cet objet peut être réalisé avec divers matériaux et même sous différentes formes, comme celle de l'abaque par exemple. En outre, elle est survenante⁵ des propriétés non-relationnelles de ses parties car, si l'on copie une machine à calculer atome par atome, le résultat sera aussi une machine à calculer.

Analysons la caractérisation non-relationnelle que l'on donne à la dite propriété. L'ensemble des possibilités de machines qui méritent la définition de calculatrice est très vaste et dépend du contexte social où l'on se trouve, 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. 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. Or, certaines propriétés que l'on considère comme non relationnelles ne sont que des propriétés relationnelles que l'on a simplifiées.⁶

Teller va encore plus loin, il dit que ce caractère implicite reconnu dans les propriétés non-relationnelles survenantes d'une totalité s'avère pertinent dans le cas des propriétés relationnelles non survenantes.⁷

⁵Les caractéristiques de la relation de survenance peuvent être déclinées comme suit : d'abord, il y a une covariation entre les propriétés de base et les propriétés du niveau supérieur, ensuite il y a une dépendance des propriétés du niveau supérieur en fonction du niveau de base et enfin, les propriétés du niveau supérieur ne sont pas réductibles à celles du niveau de base. Lorsque je dis *corrélation dépendante* il faut comprendre que ce que l'on cherche est qu'une fois établies les propriétés de base, les propriétés du niveau supérieur de celles-ci sont aussi établies. Ainsi tout changement du niveau supérieur implique qu'il existe un changement au niveau de base sans que pour autant la converse s'avère forcément vrai. Dans le cas de la calculatrice une fois que les propriétés sont établies, alors les propriétés du niveau supérieur comme le fonctionnel (dans notre cas "servir pour calculer") le sont aussi.

⁶Je vais revenir sur cette constatation lors de la discussion du concept d'émergence dans la vie artificielle.

⁷En effet, prenons la propriété relationnelle "être le frère 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

Teller nous soumet une troisième définition qui prend en compte la possibilité d'émergence pour les propriétés fonctionnelles :

DEFINITION 3 (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 relat. [Teller, 1992, cf. page 146]

Cependant, Teller lui même récuse cet énoncé car il estime que le concept de fonction n'a pas encore le poids analytique nécessaire pour être utilisé comme facteur d'individualisation des propriétés de ce type. En effet, il pense [Teller, 1992, cf. page 146] que les propriétés fonctionnelles ont un tel niveau de complexité que cela relève souvent des caractéristiques de l'environnement social en infirmant la définition de l'énoncé 3. Face aux éléments qui sont considérés comme implicitement définis, la caractérisation d'une propriété comme émergente devient non une classification nette 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 selon l'énoncé 3, on peut se demander s'il vaut la peine d'en tenir compte dans la caractérisation de l'émergence. Il me semble que non.

Neanmoins, l'utilité du concept d'émergence dans le cadre d'une théorie donnée peut encore être préservée. Il faut être moins libéral des éléments intervenants dans sa définition. L'analyse de Teller de la propriété "d'être une calculatrice" montre que l'on considère cette propriété comme non-relationnelle parce que la définition du système dit émergent est flou. Il faudrait, donc, renoncer à la caractérisation d'une propriété émergente sur la simple base intuitive dans la détermination des parties d'une totalité. L'émergence d'une propriété globale dans le sens de l'énoncé 2 nécessite une définition claire des niveaux de la totalité en question. Ceci n'est possible que si l'on possède une théorie de systèmes comme une métathéorie. Dans ce cas seulement, les cas triviaux des propriétés émergentes pourront être décelés.

fait n'ont pas de parent du 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 .

Il découle de tout ceci que le concept d'émergence le plus riche et utile est loin d'être intuitif lors de son application à une totalité donnée.

1.2 La propriété de *fitness*

La propriété de *fitness* pourrait être considérée, selon Teller, comme une contrepartie de la propriété "d'être une calculatrice" dans le domaine des sciences de la vie.⁸ Cette propriété est, comme chacun sait, la capacité d'un organisme à s'adapter à son environnement. Cette capacité est le fait d'un système complexe spécialisé dans les relations avec le milieu. Afin d'illustrer la complexité de la relation pour continuer à être vivant, un organisme doit se nourrir mais en mangeant de la nourriture qui lui soit appropriée.

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 mêmes intuitions incitent de nos jours la communauté scientifique de la vie artificielle à parler d'émergence.

2 L'émergence dans la vie artificielle

La vie artificielle (VA) veut se différencier foncièrement des disciplines de l'IA. Un des éléments de cette différenciation 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, et ce, en 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 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

⁸[Teller, 1992, cf. page 148]

entre eux. En ce faisant, le comportement observé du système se révèle doté de capacités d'adaptation d'une souplesse remarquable.

Selon les chercheurs de la VA ces deux modèles constituent des théories des phénomènes naturels, mais les modèles de l'IA seraient des *simulations* tandis que ceux de la VA seraient des *réalisations*,⁹ ce qui ne me semble pas justifié. 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 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 mutatis 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 selon les 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 bien adaptés à leur milieu. Il sera en effet difficile voire impossible de trouver des propriétés physiques communes. Il en va de même pour les propriétés 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 le résultat d'un amalgame des deux domaines distincts : le domaine cognitif et le domaine biologique.

⁹Selon la caractérisation de H. Patee cf. [Patee, 1989]

2.1 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'une 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. 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. Comme Sober signale :

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 et ceci de façon à ce qu'elles constituent une réalisation ? L'unique vérification possible consiste en l'observation de la performance mais cela 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 mathématique 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 ¹⁰ 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 des gènes A et q celles de a (où $p + q = 1$) alors, dans les circonstances citées ci-dessus, les fréquences 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 de chaussures des deux teintes du pied droit.

Le patron de la chaîne envisage de trier les paires en utilisant une machine aléatoire. 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* (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, la différence fondamentale 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* selon 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.

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) subject

¹⁰La définition est la suivante : "diploid : an organism or a cell having double the basic or haploid number of chromosomes." Selon le Webster's Encyclopaedic Unabridged Dictionary (1989)

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].

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 computationnel se révèle vivant à l'avenir.¹¹

Le problème est qu'il est très difficile de trouver des critères objectifs¹² pour déterminer jusqu'à quel point un système mérite la qualification de réalisation des phénomènes de la vie. 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.

2.2 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 qui 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 précédemment 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.¹³

¹¹Pour une discussion des postulats de la vie artificielle maintenant que le processus de la vie est purement physique voir [Rasmussen, 1991]

¹²Le terme *objectif* veut simplement éliminer toute possibilité, pour ces critères, d'être une simple projection de nos capacités comme être vivants.

¹³Je 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

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 catégories sont à l'origine des changements de comportements.

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 abus de langage.¹⁴ 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 comme *methodologie 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 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.

par la métaphore suivante : "l'agent devient l'auteur de . . . [la] connaissance [des relations avec l'environnement]"

¹⁴Par exemple dans une modélisation donnée voir [Rodriguez, 1994, page 108]

2.2.1 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.¹⁵

Dans le domaine de la VA proprement dite ces modèles sont utilisés par Langton et Toffoli¹⁶. 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 émergent ne change en rien la qualité d'émergence de ce dernier. L'idée ancienne et fautive 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 les bases de l'irréductibilité de ces phénomènes.

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 en 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 thermo-

¹⁵Pour un description détaillée voir [Smolensky, 1994].

¹⁶cf. [Langton, 1986] et [Toffoli, 1982]

dynamiques 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'explications le concept d'*attracteur*¹⁷ joue un 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, entre autres, I. Prigogine et J. Petitot.¹⁸

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.

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 rencontrés, l'observateur apparaît comme extérieur à la programmation de l'agent et le comportement est défini comme 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 émergente 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 une 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

¹⁷Je fais ici référence au concept d'*attracteur* dans le cadre strict de la théorie morphodynamique.

¹⁸cf. [Prigogine, 1980], [Petitot, 1992], [Nicolis and Prigogine, 1985].

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 new 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 complètement 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.

3 Conclusion : Les modèles de la VA comme simulations de processus biologiques

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. Le fait de découvrir dans ces observations des implications plus ou moins directes pour les sciences de la vie ne me semble pas justifié. De plus, cette idée me paraît malheureuse car la VA veut être considérée comme une 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 qu'éveille 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.

Références

- [Assad and Packard, 1991] Assad, A. and Packard, N. (1991). *Emergent colonization in an artificial ecology*, pages 143-152. In [Varela and Bourguine, 1991].
- [Beckermann et al., 1992] Beckermann, A., Flohr, H., and Kim, J. (1992). *Emergence or Reduction ?* Walter de Gruyter, Berlin-New York.
- [Cariani, 1991] Cariani, P. (1991). *Emergence and Artificial Life*, pages 775-797. In [Langton et al., 1991].
- [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 overview*. MIT Press, Cambridge, Massachusetts.
- [Langton et al., 1991] Langton, G. C., Taylor, C., Farmer, J. D., and Rasmussen, S. (1991). *Artificial Life III. Proceedings of the workshop on artificial intelligence held february, 1990 in Santa Fe, New Mexico*. Addison Wesley, Redwood City, California.
- [Macdonald and Macdonald, 1994] Macdonald, C. and Macdonald, G., editors (1994). *The Philosophy of Psychology : Debates on Psychological Explanation*. Basil Blackwell, Oxford.
- [Nicolis and Prigogine, 1985] Nicolis, G. and Prigogine, I. (1985). *Exploring complexity*. Piper, Muchen.
- [Pattee, 1989] Pattee, H. (1989). *Simulation, realisation and theories of life*, pages 63–77. In [Langton, 1989].
- [Petitot, 1992] Petitot, J. (1992). *Physique du sens*. Edition du CNRS, Paris.
- [Prigogine, 1980] Prigogine, I. (1980). *From Being to Becoming*. Freeman, San Francisco.
- [Rasmussen, 1991] Rasmussen, S. (1991). *Aspect of Information, life, reality and physics*, pages 767–773. In [Langton et al., 1991].
- [Rodriguez, 1994] Rodriguez, M. (1994). *Approche Constructiviste de l'architecture de contrôle et de la représentation des connaissances*. PhD thesis, IIIA. Université de Neuchâtel.
- [Scaglione, 1996] Scaglione, M. (1996). *L'intentionnalité et les modèles artificiels. Émergence et réalisation : deux côtés de la même pièce*. PhD thesis, Université de Neuchâtel.
- [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].
- [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].
- [Toffoli, 1982] Toffoli, T. (1982). *Physics and computation. Int. J. Theor. Phys.*, 21(3–4) :165–175.

[Varela and Bourguine, 1991] Varela, F. and Bourguine, P. (1991). *Toward a practice of autonomous systems. Proceedings of the first European Conference on Artificial Intelligence*. MIT Press, Cambridge, MA.

LE PARADOXE DE L'ÉTHOLOGIE ET DE LA PSYCHOLOGIE : DE L'USAGE D'UN RÉFÉRENTIEL À LA CONSTRUCTION DU RÉFÉRENTIEL DU SUJET D'ETUDE

Fabienne DE GAULEJAC & Alain GALLO

*Laboratoire de Neurobiologie et Comportement. Université Paul Sabatier
118 route de Narbonne. 31062 Toulouse cedex
Tel : 05 61 55 69 12. Fax : 05 61 55 84 44
E-Mail : gaulejac@scico.u-bordeaux2.fr & gallo@cict.fr*

1. Introduction

Qu'il existe de nombreux modèles du même réel ne surprend pas ceux qui distinguent réel et réalité. " *Un même phénomène peut permettre des modélisations sans rapport les unes avec les autres* " (Dupuy, 1994) et donc donner naissance à des artefacts différents. Dès lors, on peut s'interroger sur la distance entre un artefact qui mime le comportement d'un animal et le comportement même de cet animal. Dans l'optique d'une "psychologie du décalage", il ne s'agira pas, en fait, de se demander si le modèle du déplacement d'un animal peut ou non se rapprocher du modèle de déplacement d'un mobile physique mais de s'interroger sur les possibilités d'en inventer un autre, radicalement différent (Gallo & de Gaulejac, sous presse). Il s'agit ici d'imaginer, de construire, non pas un modèle du comportement animal plus ou moins analogue à celui que nous pensons être le nôtre, mais plus véritablement un modèle de fonctionnement cognitif qui soit *a priori* différent du nôtre, une psychologie animale qui ne soit pas presque humaine. Si notre objet d'étude, contrairement à un objet physique quelconque, est lui-même un sujet de connaissance, cela ne signifie pas qu'il faille lui prêter le fonctionnement de l'intelligence humaine.

Selon Varela (1989), nous ne pouvons pas sortir de notre domaine cognitif, de notre mode de fonctionnement. Comment, dans ces conditions, pouvons-nous avoir plus d'un système unitaire de référence à prêter à des systèmes cognitifs que l'on suppose différents des nôtres ? Si "*dans son activité modélisatrice, le savant projette son esprit dans le monde des choses*" (Dupuy, 1994), et si "*chaque sujet vit dans un monde où il n'y a que des réalités subjectives et où les milieux mêmes ne représentent que*

des réalités subjectives” (von Uexküll, 1956), n'a-t-il pas, en effet, quelque précaution épistémologique à prendre quand il étudie l'animal ? Que cette précaution soit d'ordinaire oubliée ne signe pas la validité d'un raccourci où l'autre étudié risque fort de perdre sa spécificité.

Cette précaution devrait conduire à une typologie des modèles, tous n'étant pas, d'un point de vue épistémologique constitutif, adaptés à la réalité définie. Au long des exemples développés, nous essayerons d'illustrer la nécessité d'un effort de relativisation qui nous paraît fondée par la conception de l'environnement animal. On peut imaginer que les artefacts qui pourraient résulter de modèles différents pourraient obéir à des principes différents.

2. Comment construire un modèle de comportement animal ?

Le problème ici posé est celui des fondements épistémologiques, théoriques et méthodologiques de l'étude du comportement animal. Si, en éthologie, tout comportement est l'expression de la relation (continuité) du sujet au monde, la nature de cette relation est déterminée par le référentiel dans lequel le chercheur inscrit son objet d'étude (Gallo & Huet, 1984). La construction d'un modèle de comportement différent du modèle humain peut alors passer par la construction d'un principe de fonctionnement psychologique singulier, régissant la relation de l'animal à son monde et déterminant ses comportements. Une façon de procéder consiste à raisonner par simple différence, en énonçant une rupture avec le principe qui unit le chercheur à son propre monde. Ce postulat d'un fonctionnement psychique différent a pour conséquence la distinction entre le système de description de l'observateur et le "point de vue" de l'être étudié.

Une telle démarche consiste en l'abandon de propositions théoriques postulant, entre autre chose, le principe d'action à distance (à travers le vide) et donc l'existence d'un monde d'objets pour l'être humain. Au modèle cognitiviste axé sur le schème de la permanence de l'objet pour l'homme serait ainsi substituée la proposition hypothétique d'un monde composé de *non-objet pour le non-humain*.

Selon von Uexküll (1956), *"l'animal ne rentre jamais en contact avec un objet comme tel"*. Cette formulation est négative. Pour la compléter, il nous faudrait dire ce qu'est pour l'animal ce que nous, humains, appelons "objet". En ce point, on peut trouver un appui dans la thèse de Gibson (1979) qui considère que les unités naturelles de l'environnement que l'animal construit ne doivent pas être confondues avec les unités métriques conventionnelles de l'espace et du temps. En d'autres

termes, la description physique de l'objet serait à distinguer de sa description écologique, laquelle est à relier étroitement aux comportements des animaux. Dans ce contexte, ce qui est objet pour l'homme est possibilité d'action pour l'animal (*affordance*).

Nous ajouterons à cette thèse la proposition selon laquelle l'objet neutre, sans valeur biologique, n'a pas d'*affordance* en tant que telle, pas de signification comportementale : il reçoit celle du lieu où il est placé. Ce principe de dépendance entre l'acte, l'endroit où il se déroule et l'objet sur lequel il s'applique, nous l'avons baptisé principe d'*acto-spatialité* (de Gaulejac & Gallo, 1996).

Il en résulte que la description comportementale que réalise l'éthologiste doit, si l'on accepte ce principe, être *contextualisée*, en ce sens qu'elle doit respecter l'unité du comportement et du lieu dans lequel il se déroule. En effet, la rupture qui isole traditionnellement, dans la description éthologique, le comportement du lieu de sa réalisation, s'inscrit dans une théorie de l'objet et une théorie géométrique de l'espace (de Gaulejac, 1997).

A l'acceptation de la partition sujet/monde, sujet/objet et à la définition de l'animal comme Système de Traitement des Informations est préférée, dans ces conditions, la conception d'un Système de Production de Significations (de Gaulejac & Gallo, 1997).

3. L'opérationnalisation

L'opérationnalisation de la thèse que nous proposons implique la prise en compte des caractéristiques de l'objet chez l'homme et l'énoncé, pour chaque caractère, d'un caractère différent pour l'animal. L'objet perçu comme *unité* et tombant sous le coup de la catégorie logique de l'*identité* ($A = A$ dans l'espace et le temps) dispose chez l'être humain d'une *permanence* (de Gaulejac & Gallo, soumis).

Selon le principe énoncé d'une cognition autre - celle de l'animal -, on se doit donc de tester les couples d'opposition suivants : unité / non-unité de l'objet ; identité / non-identité ; permanence / non-permanence. Dans l'ordre donc, il faut examiner, en se référant aux données empiriques, si l'objet constitue ou non une unité pour l'animal, puis s'il garde son identité dans le temps et l'espace et enfin, éventuellement, s'il obéit au principe de permanence. Il en résultera le choix entre ces deux modèles de la cognition. Ne nous y trompons pas, "*cette pratique d'observation du comportement animal est une pratique accomplie par un observateur théorique et non pas*

cette activité définie comme un regard attentif s'attachant à un objet” (Gallo, 1988).

Le découpage du monde en *unités* différentes repose sur une conception intégrée, donc unitaire, de tous les caractères de l'objet. Chez l'animal, il semble que ce découpage n'obéisse pas aux contraintes qui existent chez l'être humain puisqu'il peut être modifié, particulièrement en fonction de l'état de l'individu et de la signification psychologique qu'il accorde à son emplacement (cf. par exemple, de Gaulejac & Gallo, 1996).

Une fois ce monde d'unités séparées constitué, se pose le problème de l'*identité* des objets. L'étude des loups captifs montre qu'un objet ne conserve pas son identité dans le système spatio-temporel (de Gaulejac, 1997). L'absence d'identité (A¹ A) des réactions au même objet selon son emplacement fonctionne tout autant pour des objets sans valeur biologique que pour des objets à forte valeur biologique comme la nourriture. En ce dernier cas, la signification biologique de l'objet pouvant être différente de la signification de son lieu de dépôt (par exemple manger et se coucher), on observe généralement un passage d'un comportement à l'autre, révélant une rupture dans l'identité de l'objet. Seules les aires significatives du domaine vital semblent donc perçues comme unitaires et obéissant au principe d'identité. Mais des études longitudinales permettent également de reconsidérer l'identité de l'objet et du territoire en fonction du temps. Reconnu comme zones alimentaire, de repos, de jeu, de miction... à une période définie de l'année, le territoire subit généralement, chez les mammifères, lors de la saison suivante, des changements notables de significations, voire de limites.

Enfin, l'examen de la *permanence* de l'objet pose la question de l'existence de la capacité à dépasser le *hic et nunc* des données perceptives. La recherche d'un objet disparu derrière un écran signifie qu'il reste présent pour le sujet sans aucun indice de cette présence et que la logique de ses déplacements est comprise par l'animal. Il apparaît que la permanence n'est pas parfaite, sauf peut être chez les singes - encore que l'on puisse se demander si les stades ultérieurs du développement que ces animaux n'atteignent pas ne rétroagissent pas chez l'enfant en modifiant le schème de permanence de l'objet.

4. D'un modèle l'autre, à un autre modèle : d'autres systèmes de référence ?

Le modèle ainsi construit par différence est-il spécifique de la réalité pour laquelle il a été construit ou peut-il s'appliquer à d'autres êtres que les animaux ? En fait, seule la confrontation avec les observations telles que nous les avons définies est à même de répondre à cette question. Car imaginer un monde composé de non-objets ne nous renseigne pas sur la nature de l'être qui pourrait éventuellement y vivre. En l'occurrence, nous nous sommes demandés si ce modèle partiel (relation aux objets) que nous avons développé chez l'animal ne serait pas, par exemple, applicable, en psychopathologie humaine, à des êtres que l'on suppose fonctionner différemment, par exemple à des personnes autistes ? Seule l'observation de leurs comportements vis-à-vis des objets pourra nous assurer que le modèle construit s'applique à la réalité étudiée et permettra un point d'ancrage de la réflexion sur les raisons pour lesquelles ces personnes se comportent comme elles le font.

Malgré les réticences et la résistance à questionner l'approche classique, un certain nombre de chercheurs reconsidèrent actuellement la vision jusque là dominante de l'autisme. Se pourrait-il que ce qui fait sens pour nous soit totalement obscur pour une personne autiste ? que ses actes répondent à une logique qui ne soit pas celle utilisée par le plus grand nombre ? ... Ne peut-on envisager que l'autisme ne se ramène pas à une maladie (Peeters, 1996) mais constitue un mode particulier et spécifique de fonctionnement psychologique, une cognition *autre* ?

"Être atteint d'autisme ne signifie pas être inhumain, mais plutôt étranger. Ce qui est normal pour les autres ne l'est pas pour moi, et ce qui est normal pour moi ne l'est pas pour les autres. (...) Ma manière d'être n'est pas une version dégénérée de la vôtre. Remettez-vous en question (...)" Sinclair (1992)

4.1. La relation objectale des enfants autistes

Le rapport de l'autiste à l'objet semble problématique :

"Un jour, j'ai été obligée de lui apprendre les différentes catégories de personnes et d'objets, (...). Il ne voyait pas ce qui correspondait à telle ou telle catégorie. Il mettait par exemple des images de bananes avec des images de personnes, des vaches avec des voitures, des personnes avec des animaux..." (de Clercq, 1997).

Il en va des objets inanimés, semble-t-il, comme des objets vivants :

"Les objets sont angoissants. Le fait d'être confronté à des objets mobiles est pire à cause de la difficulté supplémentaire provoquée par le mouvement lui-même. Les objets qui font du bruit sont bien pires (...). Mais les êtres humains sont les plus angoissants de tous, parce qu'ils bougent lorsqu'on ne s'y attend pas" (Joliffe et al., 1992).

Référence peut ici être faite à la reconnaissance d'un type particulier d'objets utilisé par l'enfant autiste : les *objets autistiques et confusionnels* (Tustin, 1986) par opposition à l'*objet transitionnel* (Winnicott, 1958), "*première possession 'non-moi' de l'enfant*", constituant un pont vers la mère différenciée et permettant de supporter son absence.

Les objets autistiques sont utilisés comme s'ils faisaient partie du corps ("*possession-moi*", Tustin, 1986), sur un mode idiosyncrasique, obsessionnel ne correspondant pas à leur fonction. Ils sont particuliers à chaque enfant et sont peu différenciés. "*Si son objet autistique disparaît, l'enfant sera affligé comme s'il avait perdu une partie de son corps, mais il le remplacera bientôt par un autre qu'il ne différenciera pas du premier*" (Tustin, 1986). Ces objets ne sont pas, comme l'est l'objet transitionnel, des objets de compagnie. Ils constituent pas des *substituts* d'une personne, ils *sont* cette personne parce qu'ils leur donne la sensation qu'ils désirent. Par la procuration d'une satisfaction instantanée, ces objets suppriment le délai entre l'anticipation et la réalisation. Ils constituent alors une barrière séparant l'enfant de ses parents. En recourant à eux, l'enfant autiste évite d'éprouver un sentiment de besoin et d'attente.

A mi-chemin entre les objets transitionnels et autistiques, les objets confusionnels "*sont une confusion d'éléments du 'moi' et d'éléments du 'non-moi'*" (Tustin, 1986). Interchangeables, utilisés de manière idiosyncrasique dans un but d'évitement, ces objets se distinguent cependant des objets autistiques en ce que l'enfant a vaguement conscience qu'ils sont séparés de son corps.

4.2. Un troisième style cognitif ?

Pour aborder ce monde particulier de l'autisme, nous disposons désormais de deux modèles traduisant le monde des objets dans lequel nous vivons et un monde d'affordances dans lequel semblent vivre les animaux. Eventuellement, comment construire un troisième modèle pour l'autisme qui ne renvoie ni à l'objet, ni au non-objet, mais vraisemblablement à autre chose qu'un objet ? L'autiste, ici, est interrogé de deux lieux : celui de l'objet

et celui du non-objet. Que l'enfant autiste soit, comme l'animal, confronté aux objets ne signifie pas une équivalence d'opérations, une identité de présence au monde.

En fait, tout se passe comme si nous devions trouver une voie moyenne entre deux mondes pour en inventer un troisième. La tâche est différente de celle que nous avons accomplie antérieurement et qui consistait à construire un modèle à partir d'un autre modèle. De la même façon que pour la recherche en éthologie, nous nous fierons aux observations comportementales qui sont, répétons-le, les seules à même de nous indiquer que le modèle que nous aurons construit a quelque chance de correspondre à la nature du monde dans lequel vivent les autistes.

Différentes expérimentations, actuellement en cours de réalisation, permettront peut-être de préciser le mode spécifique de présence au monde des enfants autistes.

Les épreuves d'appariement à un modèle, également appelé perception intermodale, sont généralement utilisées afin de vérifier le caractère unitaire de l'objet. Dans l'activité perceptive, l'objet tombe sous le coup de modalités sensorielles différentes. La perception intermodale est un phénomène qui désigne les capacités d'extraction et de transfert des informations entre différentes modalités sensorielles (visuelle et tactile, visuelle et olfactive, visuelle et gustative...). Trois objets sont, par exemple, présentés à un sujet qui ne peut en voir que deux et toucher le troisième. Un des objets visuellement perceptible est identique à l'objet exploré par voie tactile, tandis que l'autre est nettement différent. La tâche consiste à choisir entre les deux objets vus celui qui est identique à l'objet touché. Si le choix est correct, le sujet reçoit une récompense. Quand cet appariement est parfaitement réalisé, on soumet le sujet à de nombreux problèmes, présentés une seule fois chacun, en recourant chaque fois à des objets nouveaux.

Par ailleurs, un objet n'est jamais donné dans la perception sous tous ses angles en même temps (problème phénoménologique de l'objet transcendantal). On peut donc se demander si le sujet est capable de reconnaître le même objet vu sous des angles différents. Une rotation de 180 ° de l'objet autour de son axe entraînera des réactions très différentes selon que ce mouvement change ou ne change pas pour le sujet la nature de l'objet.

A supposer que les enfants autistes soient capables de reconnaître les unités du monde, cela ne signifie pas que, pour eux, l'objet A reste, dans le temps et l'espace, identique à lui-même. Ils peuvent fonctionner en

constituant des unités, mais le contenu de ces unités peut se modifier, de sorte qu'une unité considérée à un instant t peut, à un instant $t+1$, se transformer, par exemple, en de multiples unités. On dira alors que l'identité de l'objet n'est pas respectée.

Il est donc possible de mettre en évidence des altérations du principe d'identité ($A = A$) d'une unité en montrant qu'un même objet, pour l'observateur, est la cible de comportements différents de la part de l'enfant, sans qu'il y ait de modification physique de l'objet mais simple déplacement ou variation d'orientation dans l'espace.

Enfin, on pourra se demander si les réactions à deux objets appartenant à la même catégorie (identité catégorielle) sont différentes des réactions à deux objets appartenant à des catégories différentes.

Si unité et identité des objets sont présents, la permanence de l'objet pourra, dans un dernier temps, être testée. Cette permanence ne s'éprouve pas dans le fait que l'objet reste ce qu'il est dans la durée mais dans l'expérience de sa redécouverte à l'endroit où on l'a caché. Il est présent là où on l'a placé, même s'il n'est plus perceptible, ce qui signifie qu'il est représenté. Pour ce faire, différentes situations classiques de "cache" seront mises en œuvre.

5. Conclusion

L'insistance sur le vécu est ici la marque d'une épistémologie du vivant qui pose comme principe la substitution du point de vue de l'observateur par le point de vue du sujet étudié (cf. par exemple, pour un autre domaine, la critique wallonienne de l'adultocentrisme). On retrouve en ce point la difficulté d'une étude de type psychologique qui ne peut se contenter d'un regard extérieur. Le problème méthodologique posé est celui d'un système cognitif étudiant un autre système cognitif. Si la relation du chercheur au comportement de la personne autiste ou de l'animal était totalement étrangère à celle de ces individus, il n'y aurait pas d'étude possible de leur comportement. En conséquence, le mode de relation de l'homme inclut celui de ses sujets d'étude. L'évolution pourrait alors être conçue par l'homme comme apparition historique de relations, intégrées finalement en l'homme sous la forme d'une relation incluant toutes les autres et permettant leur connaissance (Gallo & Huet, 1984).

Quelle que soit la nature de la relation au monde de l'être étudié, son Umwelt psychologique, en tant qu'il est la cible de notre investigation, fait donc partie intégrante de notre propre monde de signification. De ce fait, nous ne faisons donc, en dépit de nos ambitions, qu'exprimer notre propre

Umwelt en tentant d'exprimer celui de l'animal ou de tout autre individu. Cela ne signifie cependant pas qu'un seul référentiel, absolu, soit disponible. Plus que le contenu des systèmes cognitifs proposés, c'est donc davantage le projet de cette recherche que nous souhaitons défendre.

6. Références bibliographiques

- de Clercq, H. (1997). *Dit zijn geen rozen, 't zijn witten. Over een jongen met autisme*. Antwerpen, Hadewijch.
- de Gaulejac, F. (1997). *Effets du contexte sur la réaction comportementale à la nouveauté chez l'animal : d'un système cognitif à un autre ?* Thèse de troisième cycle. Université Paul Sabatier, Toulouse.
- de Gaulejac, F. & Gallo, A. (1996). L'animal et la nouveauté : Représentation ou acto-spatialité. *Intellectica*, 1: 22, 169-184.
- de Gaulejac, F. & Gallo, A. (1996). Les comportements inter-individuels des animaux sont-ils des comportements sociaux ? *Actes des Journées de Rochebrune : Du collectif au social*, ENST, Paris, 1996, pp. 71-83.
- de Gaulejac, F. & Gallo, A. - Des interactions entre l'animal et le monde à l'enaction d'un monde propre. In G. Théraulaz et F. Spitz (Eds.), *Auto-organisation et comportement*, Paris, Hermès, pp. 61-75, 1997.
- de Gaulejac, F. & Gallo, A. (soumis). L'identité pour l'animal : entre identification individuelle et identification catégorielle.
- Dupuy, J.P. (1994). *Aux origines des sciences cognitives*. La Découverte, Paris.
- Gallo, A. (1988). *Pour une approche psycho-éthologique du comportement animal*. Thèse d'état. Université Paul Sabatier, Toulouse.
- Gallo, A. & Huet, C. (1984). L'évolution de la relation de l'être vivant au monde et les conditions d'une connaissance scientifique du comportement animal. *Bulletin Intérieur SFECA*, 1, 105-117.
- Gallo, A. & de Gaulejac, F. (sous presse). Modèle physicaliste de l'environnement et modèle écologique de l'Umwelt. In , P. Lestage, B. N'Kaoua & B. Leblanc (Eds.), *Second colloque Sciences Cognitives - Facade Atlantique, Cognitives*, vol. 2.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Houghton Mifflin, Boston.
- Jolliffe, T., Landsdown, R. & Robinson, C. (1992). Autism: a personal account. *Communication*, 26 (3).
- Peeters, T. (1996). *L'autisme : de la compréhension à l'intervention*. Dunod, Paris.

- Sinclair, J. - Bridging the gaps: an inside-out view of autism. In E. Schopler & G. Mesibov (Eds.), *High functioning individuals with autism*, Plenum Press, New York, 1992.
- Tustin, F. (1986). *Les états autistiques chez l'enfant*. Seuil, Paris.
- Varela, F.J. (1989). *Autonomie et connaissance : Essai sur le vivant*. Seuil, Paris.
- von Uexküll, J. (1956) *Mondes animaux et monde humain*. Gonthier, Hambourg.
- Wallon, H. (1941). *L'évolution psychologique de l'enfant*. Armand Colin, Paris.

A model for serendipity

by Pek van Andel (m.v.van.andel@med.rug.nl)

(Laboratory for cell biology and electron microscopy, Medical Faculty, Oostersingel 69-2, 9713 EZ, Groningen, Holland)

SUMMARY

In French the English-American word 'serendipity' doesn't exist. And dictionaries do not transcribe it as *serendipité*, but translate it as '*l'art de faire des trouvailles*'. I define *serendipity* as the art of making an unsought finding and collected about two thousand cases of serendipitous discoveries, inventions and creations in science, technology and art. In this chaos of so called 'accidental' or 'chance' findings I found an order of twenty-one 'patterns of serendipity'. My model of serendipity patterns introduces a new and possibly stimulating perspective on the old subject of serendipity. Knowledge of these 'serendipity patterns' might help in expecting also the unexpected and finding also the unsought. Each pattern describes how an unanticipated (stochastic) observation, that was essential for an unsought finding, presented it self to the finder who elaborated it into a discovery, invention or creation.

INTRODUCTION

With a background as medical researcher, I began my research on serendipity - as a '*Spielerei nebenbei*' - by collecting examples of serendipity. In this 'serendipity forest' I found, till now, twenty-one clearly recognizable patterns by which unsought findings cropped up and/or were dugged up. The existence of these patterns and these patterns were surprising because 'chance', 'accidental' or 'fortuitous' observations are conventionally seen as 'random' and therefore excluding any 'order in the chaos' of serendipity examples. But 'order in chaos' is exactly what I found. 'Chance', 'accidental' and 'fortuitous' means in a case of serendipity not 'random' but only that an 'accidental' observation 'falls' to you (*ac-cido = ad-cedo = to fall to*), which you are or were not looking (on that moment or before). As Claude Bernard wrote :

'Il n'y a rien d'accidentel, et ce qui pour nous est accident n'est qu'un fait inconnu qui peut devenir, si on l'explique, l'occasion d'une découverte plus ou moins importante.'](Cl. Bernard, *Introduction à l'étude de la médecine expérimentale*, Paris, 1865)

We cannot go only logically from the old to the new, or from the known to the unknown. If we should be able to do so, the result shouldn't be really unknown or new. To discover something really new or unknown, we need also a stochastic (unpredictable) element. In patent laws an invention is therefore defined as something surprising, that did not logically evolve from the known or the old. An 'invention' that evolved only logically from existing knowledge, techniques or insights, it is therefore not patentable. A true new discovery, invention (*in-veno* = I come on [something]), creation or thought is never purely logically. Nor is it purely stochastic. It is always a combination and interaction of logic thinking and/or behavior and one or more stochastic elements. This is not only true at a laboratory bench, for example, but also outside a laboratory and even in our own head. As Henri Poincaré wrote, we sometimes need the unexpected combinations permitted by the disorder of our subconsciousness to solve a question that we couldn't answer logically :

'Dans le moi subliminal règne, au contraire, ce que j'appellerais la liberté, si l'on pouvait donner ce nom à la simple absence de discipline et au désordre né du hasard. Seulement, ce désordre même permet des accouplements inattendus.' (H. Poincaré, *Science et Méthode*, Flammarion, Paris, 1908, Chapitre III: *L'invention mathématique*, pp. 43-63. This text of the celebrated lecture of Poincaré at the *Société de Psychologie de Paris*, was also published in the *Bulletin de l'Institut Général de Psychologie*, No 3, 8e année, 1908, p. 137-151)

Serendipity is by definition a finding that has not been anticipated - with an hypothesis for example - by the finder. This non-anticipation is also the reason, why the finder is surprised by his serendipitous observation. The American sociologist of science Robert Merton (1910-), stated that empirical facts aid in the initiation of theory, and he gave a precise definition of the serendipity ::

(..) The serendipity pattern refers to the fairly common experience of observing an *unanticipated, anomalous and strategic* datum which becomes the occasion for developing a new theory or for extending an existing theory. Each of these elements of the pattern can be readily described. The datum is, first of all, unanticipated. A research directed to the test of one hypothesis yields a fortuitous by-product, an unexpected observation, which bears upon theories not in question when the research was begun.

Secondly, the observation is anomalous, surprising, either because it seems

inconsistent with prevailing theory or with other established facts. In either case, the seeming inconsistency provokes curiosity; it stimulates the investigator to 'make sense of the datum', to fit it into a broader frame of knowledge. He explores further. He makes fresh observations. He draws inferences from the observations, inferences depending largely, of course, upon his general theoretical orientation. The more he is steeped in the data, the greater the likelihood that he will hit upon a fruitful direction of inquiry. In the fortunate circumstance that his new hunch proves justified, the anomalous datum leads ultimately to a new or extended theory. The curiosity stimulated by the anomalous datum is temporarily appeased.

And thirdly, in noting that the unexpected fact must be strategic, *i.e.*, that it must permit of implications which bear upon generalized theory, we are, of course, referring rather to what the observer brings to the datum than to the datum itself. For it obviously requires a theoretically sensitized observer to detect the universal in the particular. After all, men had for centuries noticed such 'trivial' occurrences as slips of the tongue, slips of the pen, typographical errors, and lapses of memory, but it required the theoretical sensitivity of a Freud to see these as strategic data through which he could extend his theory of repression and symptomatic acts.

The serendipity pattern, then, involves the unanticipated, anomalous and strategic datum which exerts pressure upon the investigator for a new direction of inquiry which extends theory. [...] Freud [...] [observed] during the last war [i. e. the first World War, in which he had two sons at the front] that he had mis-read a newspaper headline, 'Die *Feinde* vor Görz' [The *Enemy* before Görz], as the 'Der *Friede* von Görz [The peace of Görz]. Freud took a trivial incident and converted it into a strategic fact. [...]

This is perhaps sufficient to illustrate the operation of the serendipity pattern: an unexpected and anomalous finding elicited the investigator's curiosity, and conducted him along an unpremeditated by-path which led to a fresh hypothesis.' (R. K. Merton, *Social Theory and Social Structure*, Revised and Enlarged Ed., The Free Press, Glencoe, Illinois, 4th Printing, 1957, p. 104-8).

Merton described serendipity as a surprising observation followed by a correct abduction (see for 'abduction' Charles Peirce). In simple words: the searcher does a surprising observation, then tries to invent an hypothesis to explain the unanticipated observation. The hypothesis has to be testable and tested independently from the serendipitous observation. When the hypothesis is falsified, in a experiment for example, the observer will mostly try to invent another hypothesis, to explain the amazing observation. Then he will tests this hypothesis, also independent from the original observation,

and so on, till he has found an explanation for the surprising phenomenon, that withstands one or more scientific tests. If he then publishes about this positive finding, in most cases, if not all, he starts by first mentioning his last hypothesis, that was successfully verified, and doesn't say a word on how this hypothesis was born. He only writes about how this hypothesis was tested, and the results and the conclusions of that test. The reader may then get the idea, that the growth of knowledge goes directly from hypothesis to thesis, from fiction to non-fiction, from fantasy to fact, from theory to practice, in stead of also the other way around. That is a pity, because the reader is then not informed about how the hypothesis was born and developed. So the article he read, does not give the whole truth, but only part of it, even without mentioning that at all. It has been written in such a way that the reader can repeat the experiment, if he wants, to see whether he gets the same results and can draw the same conclusion(s) as the author(s) of the article. The intended advantage of the conventional form of such scientific article is evident. Anybody can check the reproducibility of the experiments, methods, results and conclusions. The unintended disadvantage of this tradition of publishing is also evident. An important part of the scientific research, namely the heuristical part - the context of discovery - is not published and therefore lost forever if not mentioned elsewhere by the author or somebody else. Then nobody can learn anything from the heuristical talents and methods of the author in question. But happily enough there are exceptions: there is a long and old tradition of a minority of (re)searchers who did and do write about the heuristic aspects of their work. Specially in chemistry this is the case. Chemistry is a big field, and chemists think in processes, and so sometimes they also write about the context and process of their discovery, and also about any possibly serendipitous (stochastic) elements their work.

Most texts on serendipity are written by chemists. But also in other disciplines there is a small tradition of writing about the serendipitous aspects of research and its findings. In this article I try to approach serendipity, empirically and systematically, by treating its domains, traditions, appearances and patterns. It is not, as most realises on serendipity, just a lucky bag of serendipities.

There are four main domains of serendipity : science, technique, art and daily life.

And six different traditions : fairy tales, and apocryph, exaggerated, underestimated, unknown and authentic examples of serendipity.

The three appearances of serendipity are : pure serendipity (a surprising observation followed by a correct explanation, e.g. the discovery of 'X-rays' by Röntgen (X is the mathematical symbol for the unknown), negative

serendipity (a surprising observation is made, but not followed by a correct abduction, remember Columbus' finding, we still speak about 'Indians', and pseudo-serendipity (a sought findings done by 'accident', Fleming was, as he said himself, was always on the outlook of what we now call antibiotics, when he discovered penicillin 'accidentally' on a plate that he had made for an other reason. Fleming :

'It was (..) fortunate that (..) I was always on the lookout for new bacterial inhibitors, and when I noticed on a culture plate that the staphylococcal colonies in the neighborhood of the mold faded away, I was sufficiently interested in the antibacterial substance produced by the mold to pursue the subject.' (I couldn't find the source of this quote this time, sorry!)

At last, originally without expecting it or looking for it at all, I found twenty-one specific patterns of serendipity. I will present them in the same order as I discovered them. As far as I know, these patterns were never described before. And I was warned, for example, by the Dutch physicist Dr. Cees le Pair, who is still director of the biggest Dutch foundation for technical sciences, that collecting serendipities could not produce anything systematic, because serendipity is by definition something that happens 'at random', so how could I expect any system at all in 'accidental findings', as I had found?

The solution for this question as mentioned shortly above was simple. When we speak about an 'accidental' observation, 'accidental' can have two different meanings. The first connotation is: a 'random' observation. The second connotation is: an observation that falls to you without expecting it. When we deal with serendipity, we have to do with the second connotation. It has nothing to do with the physicist's notion of 'randomness'. 'Accidental' has in a case of serendipity a psychological, subjective notion. The finder, who observes, discovers, invents, creates or imagines something, which he is or was not in quest of. Therefore he calls it 'accidental', meaning of 'unsought by the finder'.

Knowledge of cases and patterns of serendipity will help a searcher in missing less serendipitous observations. So far the didactical aspects.

Now the historical and philosophical aspects. By knowing the corpus of the more and less important cases and patterns serendipity, we learn to look also at the stochastic elements in the daily practice of scientists and how they dealt with them. Too often the 'context of justification' of a finding is regarded, conscious or not, as a 'context of discovery' also. The majority of the philosophers and historians of science do not like the role of 'accidental', personal and unique happenings. They want to see the great lines. But, especially in science, technique and art, details play an essential secondary

role in new findings, developments and applications. 'God lives in the detail', as Voltaire wrote.

After formulating the hypothesis one can derive testable statements about other facts than the surprising fact. The testing of these statements tells how probable or true the hypothesis is. According to the defenders of 'the hypothetical-deductive method', the last step of testing is the only one that can be logically analysed. But according to Peirce, and much later Hanson, also the process of forming the hypothesis (with the structure of an abduction) can be logically analysed and is even the most interesting part of a scientific process :

'What leads to the initial formation of Hypotheses - the 'click', intuition, hunch, insight, perception is a matter of psychology. But many hypotheses flash through the investigators mind only to be rejected on sight. Some are proposed for serious consideration, however, and with good reasons. (...) we are discussing the rationale behind the proposal of hypotheses as possible *explicantia*. Hypothesis-Deduction theorists never raise the problem at all.' (N. R. Hanson, *Patterns of discovery*, Cambridge University Press, 1958, p. 200)

Four years later Hanson about 'retroductive inference', strange enough without making any reference to 'serendipity' :

[..] From Mill through Braithwaite, scientists were seen to proceed by the hypothetico-deductive method (henceforth 'HD'). Still others have espoused something called 'retroduction' (RD) as the analysis of how scientists reason. Peirce, and Aristotle himself, opted for this view.

[..] the differences between these philosophical arguments are not 'merely psychological', but genuinely conceptual. Perhaps neither the HD nor the RD account has any application for the analysis of scientific argumentation. Or maybe both apply, but never at once. However this may be, it is not the case that the HD and the RD accounts constitute conceptually *equivalent* characterizations of any one scientific argument. [..]

The HD account is concerned not only with *conclusion deducing*, but with *hypothesis testing*. [..]

[..] Whenever the extension of a partially confirmed theory is in question, one generates further observational consequences of the theory and checks them against the facts. [..] The very identification of an event as 'anomalous' depends of this HD elaborations of familiar premise clusters. The HD theorist attends thus to the scientist's inferences from contingent premise clusters to observationally vulnerable conclusions. The RD account focuses rather on

the explanation of anomalies. RD enthusiasts think scientific argumentation to consist first in the recognition of anomalies, and then in the hunt for some premise which, if confirmed, would explain the anomaly. [...] So while the HD account pictures the scientist with a ready-made theory and a store of initial conditions at hand, generating from these testable observation statements, the RD account pictures him as possessing only the initial conditions and an upsetting anomaly, by reflections upon which he seeks an hypothesis to explain the anomaly and to found a new theory. Again, the HD account focuses on *hypothesis testing*; the RD account is concerned with *anomaly explaining*.

Some signal events in history have involved reasoning if the RD kind. The discovery of Neptune, and of the neutrino, are characterizable thus. Just as the discovery of Pluto, and of the antiproton, seem better described in HD terms. Here one runs out the consequences of an accepted theory and tests them. In the RD case, some facts surprisingly fail to confirm the consequences of an accepted theory; one then argues from these to some new hypothesis which may resolve the anomaly.

HD and RD enthusiasts both recognize that their formal criteria for success in argument are *precisely the same*. [...]

Now this is a moment in the history of the 19th century science which, I submit, could be misunderstood by one who minimized the *conceptual* differences between the HD descriptions of scientific discoveries and the RD description. These are not all-or-nothing exclusive accounts. You cannot recognize anomalies without the HD unpacking of accepted hypotheses. And you can't get new hypotheses for future HD unpacking without an RD type hunt for new hypotheses.' ('Retrodutive Inference' by Norwood Russell Hanson in *Philosophy of Science*, The Delaware Seminar, Edited by Bernard Baumrin, Interscience Publishers, Joh Wiley, New York, Vol 1, 1961-62, p. 21-37)

The arguing in retroductive inference (RD) is the inversion of the arguing in the hypothetico-deductive method (HD). Abduction can be regarded as an 'inverted deduction'. To conclude I refer here to a good summary, written by Kuang Tih Fann: *Peirce's Theory of abduction*, originally a M.A. thesis at the University of Illinois in 1963 (Martinus Nijhoff, The Hague, 1970, pp. 62).

Umberto Eco verified the four different kinds of abduction with the text of the third chapter of Zadic. Voltaire (1694-1778) wrote Zadic in 1747 and was inspired by a version of the *Perigrinaggio of the three Princes of Serendip* (possibly *Les Soirées bretonnes*, a free French translation by T.S. Gueulette in 1712). It was also a French version of the *Perigrinaggio* that

inspired Horace Walpole to coin the word 'serendipity'. Horace Walpole (1717-1797), a British author, member of the parlement, but much more interested in genealogy, antiquities, writing letters and stories than in politics (his father was the first great Prime Minister of England). This 'genial dilettant' is especially famous by his many lively and witty letters. In one of his eighteenthundred letters to Horace Mann, the British Envoy in Florence (that he knew from his *grand tour* through France and Italy around 1740), he derived and defined the word 'serendipity' in 1754 :

'Lett. 251. Arlington Street [London], Jan. 28, 1754. [...] This discovery, indeed, is almost of that kind which I call Serendipity, a very expressive word, which, as I have nothing better to tell you, I shall endeavour to explain to you: you will understand it better by the derivation than by the definition. I once read a silly fairy tale, called the three Princes of Serendip: as their Highnesses travelled, they were always making discoveries, by accidents & sagacity, of things which they were not in quest of: for instance, one of them discovered that a mule blind of the right eye had travelled the same road lately, because the grass was eaten only on the left side, where it was worse than on the right - now do you understand Serendipity? one of the most remarkable instances of this accidental sagacity, (for you must observe that no discovery of a thing you are looking for comes under this description) was [...]' (T. G. Remer, and A. G. & T. Borselli, *Serendipity and the Three Princes, from the Peregrinaggio of 1557*, Norman, University of Oklahoma Press, 1965, p. 4-5)

So far the introduction of serendipity and its model. Now the model.

MY MODEL OF SERENDIPITY PATTERNS

A general model for serendipitous findings is simply an enumeration of twenty-one different 'patterns of serendipity'. For didactic reasons I illustrated each pattern with a French example of serendipity (if available in my files):

1. Analogy or metaphor

Stethoscope

Hippocrates of Cos (460-377 B.C.), the first great physician of antiquity, was the first to place his ear against a patients's chest in order to perform 'direct auscultation'.

René Laennec (1781-1826) invented his revolutionary 'stethoscope' after seeing playing children. They scratched with pins on one end of a piece of wood and listened with their ears on the other end. In the introduction of the

second part *Diagnose* of his *Treatise*, the musical and flute playing Laennec wrote :

‘In 1816, I was consulted by a young person presenting general symptoms of a disease of the heart, and with whom the application of the hand and the percussion gave little result on account of the degree of fatness. Because the age and the sex of the patient made it impossible for me to do the kind of examination just mentioned [direct auscultation], I happened to recollect a well-known phenomenon of acoustics, if one applies the ear to one end of a beam of wood one hears very distinctly a pin-prick given at the other side. I fancied at the same time, that one could probably use this property of solids in this case. I took a quire of paper, made a strongly tight roll of it from which I applied one end to the region before the heart and putting the ear on the other end, I was as surprised as satisfied to hear the beating of the heart in a manner much more clear and distinct than I had ever been able to do by the immediate application of the ear.

I presumed from this moment that this means could become a useful method, and applicable, not only for the study of the beating of the heart, but also for the examination of the movements which can make noise in the chest cavity, and consequently for the exploration of the respiration, of the voice, of the rattle, and may be even of the fluctuation of a liquid discharged in the pleura or the pericardium.

In this conviction, I started, on the spot, in the Necker hospital [in Paris], a series of observations, that gave me as results new, sure signs, for the greatest part easy to grasp, and appropriate to produce the diagnostics of almost all diseases of the lungs, the pleura and the heart, more certain and more elaborate probably than the surgical diagnostics established with the help of the probe or the introduction of the finger.’ (R. T. H. Laennec, *Traité de l'auscultation médiate et des maladies des poumons et du coeur*, Paris, Asselin et cie, Librairies de la Faculté de Médecine, 1879, Entièrement conforme à la seconde édition publié en 1826 par Laennec, p. 5 of the introduction)

1.2. When the metaphor is an organism, it gives an example of ‘bionics’

The wasp teaches us

The use of wood as a material from which to make paper was first suggested in the Occident by René Antoine Ferchault de Réaumur (1683-1757), a celebrated naturalist and physicist residing in France. Réaumur had observed the habits of the wasp and concluded that the wood filaments that were used by these insects to construct their nests so resembling paper could also be

used in the actual process of papermaking. Réaumur's observation regarding the habits of the wasp were laid down before the French Royal Academy in a treatise dated November 15, 1719. Réaumur wrote :

'The American wasps form very fine paper, like ours; they extract the fibres of common wood of the countries where they live. They teach us that paper can be made from the fibres of plants without the use of rags and linen, and seem to invite us to try whether we cannot make fine and good paper from the use of certain woods. If we had woods similar to those used by the American wasps for their paper, we could make the whitest paper, for this material is very white. By a further beating and breaking of the fibres that the wasps make and using the thin paste that comes from them, a very fine paper may be composed. This study should not be neglected, for it is, I dare say, important. The rags from which we make our paper are not an economic material and every papermaker knows that this substance is becoming rare. While the consumption of paper increases every day, the production of linen remains about the same. In addition to this the foreign mills draw upon us for material. The wasp seems to teach us a means of overcoming these difficulties.'

Twenty three years later, in 1742, Réaumur's comment was :

'I am ashamed that I have not yet tried this experiment, it is more than twenty years since I first have announced it. But I had hoped that someone would have liked it to make it his occupation. It would be interesting, and possibly might be very profitable to institute experiments with wood from which these insects make their paper; for if a new material for the manufacture of paper could thus be discovered, the fortunate discoverer would be well repaid, and the country would be enriched by possessing another source of revenue.'

(quoted from : Dart Hunter, *Paper making, The History and Technique of an Ancient Craft*, Knopf, New York, 1947, p 313-315)

The fabrication of the paper you are reading now was thus inspired by wasps.

1.3. When the finder is metaphor it is an example of 'personal analogy'

Heureka!

The Greek Archimedes (287-212 BC, killed at siege of Syracuse) studied in the Egyptian city of Alexandria, at the time a centre of learning. He founded theoretical mechanics and invented the propeller, amongst other things.

The great Roman architect Vitruvius recorded this legend two centuries

after its occurrence. His description should of course be read with some reserve, however Archimedes' own writings contain some evidence in support of it. But it does give an example of pseudo-serendipity: a sought method (the measurement of volume) is found via a route not chosen for that reason (stepping into the bath). Archimedes had apparently so identified himself with the crown, that he himself had become its model. That is the 'personal analogy'. This plastically described passion for discovery represents a classic illustration of a sudden and blinding insight into a problem which the finder has long and unsuccessfully grappled with, legend or not :

'Archimedes made many and various wonderful discoveries. Of all these the one which I will explain seems to be worked out with infinite skill. Hiero was greatly exalted in the regal power at Syracuse, and after his victories he determined to set up in a certain temple a crown vowed to the immortal gods. He let out the execution as far as the craftsman's wages were concerned, and weighed the gold out to the contractor to an exact amount. At the appointed time the man presented the work finely wrought for the king's acceptance, and appeared to have furnished the weight of the crown to scale.

However, information was laid that gold had been withdrawn, and that the same amount of silver had been added in the making of the crown. Hiero was indignant that he had been made light of, and failing to find a method by which he might detect the theft, asked Archimedes to undertake the investigation. While Archimedes was considering the matter he happened to go to the bath. When he went down into the bathing pool he observed that the amount of water which flowed outside the pool was equal to the amount of his body that was immersed. Since this fact indicated the method of explaining the case, he did not linger, but moved with delight he leapt out of the pool, and going home naked, cried aloud that he had found exactly what he was seeking. For as he ran he shouted in Greek: 'Heureka, heureka!' ['I have found, I have found']

Then, following up his discovery, he is said to have taken two masses of the same weight as the crown, one of gold and the other of silver. When he had done this, he filled a large vessel to the brim with water, into which he dropped the mass of silver. The amount of this when let down into the water corresponded to the overflow of water. So he removed the metal and filled in by measure the amount by which the water was diminished, so that it was level with the brim as before. In this way he discovered what weight of silver corresponded to a given measure of water.

After this experiment he then dropped a mass of gold in like manner into the full vessel and removed it. Again he added water by measure, and discovered

that there was not so much water; and this corresponded to the lessened quantity of the same weight of gold compared with the same weight of silver. He then let down the crown itself into the vase after filling the vase with water, and found that more water flowed into the space left by the crown than into the space left by a mass of gold of the same weight. And so from the fact that there was more water in the case of the crown than in the mass of gold, he calculated and detected the mixture of the silver with the gold, and the fraud of the contractor.' (*Vitruvius on Architecture*, Volume II, Book IX, 9-12, The Loeb Classical Library, edited from the Harleian Manuscript 2767, translated by Frank Granger, Cambridge, Mass., Harvard University Press, 1970, p. 203-7)

2. One rare and surprising observation

Fracture surface

René-Just Haüy (1743-1822) remembered his observation, done in 1780, as follows :

'L'observation qui m'a donné, pour ainsi dire, la clé de la théorie, n'a point été cherchée, elle s'est présentée comme d'elle-même, en sorte qu'il suffisait de ne pas fermer les yeux sur le fait fondamental qui en était le sujet, et qui devait à son tour amener le reste. Ce fait s'est montré sur un cristal prismatique de chaux carbonatée, dont M. DeFrance, amateur très éclairé, avait eu la bonté de me faire présent, au moment où il venait de le détacher d'un groupe qui faisait partie de sa riche collection. En examinant ce cristal, lorsque je fus de retour, je m'aperçus que la fracture qui s'y était faite à l'endroit par lequel elle tenait au groupe, avait emporté une des arêtes de la base et j'observai au même endroit une face qu'il était facile de reconnaître, à la netteté de son poli et à la vivacité de son éclat, pour un des joints naturels situés entre les lames dont le prime est l'assemblage [...]' (R.J. Haüy, as cited by Jacques Nicolle in the second chapter of *La Symmetrie dans la nature et les travaux des hommes* (La Colombe), quoted by F. Lot, *Les jeux du hasard et du génie. Sur le rôle de la chance dans la découverte*, Librairie Plon, Paris, 1956, p. 18-9)

The Encyclopaedia Britannica summarises the importance of this finding :

'Haüy tried to reproduce these flat surfaces of fracture, and discovered that he got always rhombohedrons that looked like the first. This surprising observation led him his theory of the 'integrating molecule' as a 'motif, which he published in 1781. [...] In examining the fragments he discovered that they

cleve along straight planes that met at constant angles. He broke more pieces of calcite and found that, regardless of the original shape, the broken fragments were consistently rhombohedral. From subsequent experiments he derived a thoroughgoing theory of crystal structure. Fundamental to his theory were the laws of decrement and of constancy angles, whereby the cleavage forms of crystals were related geometrically to their primary forms or nuclei. Hatty subsequently applied his theory to the classification of minerals.' (*Encyclopaedia Britannica, Micropaedia*, 1989, V, p. 757)

3. Repetition of one rare surprising observation

The discovery of a 'subclinical infection' with experimental typhus

Charles-Jules-Henri Nicolle (1866-1936) received in 1928 the Nobel Prize for his discovery in 1909 that typhus is transmitted by the body louse. He discovered also the phenomenon 'subclinical infection'. For example, HIV-infected ('seropositive') people suffer, as known, from, a subclinical form of AIDS. They are 'subclinically infected'. Nicolle tells himself how he discovered a 'subclinical infection' in Tunis :

'We had just established that guinea-pigs were sensitive to exanthematic typhus. Since the typhus agent could not be cultured, we had previously been forced to hold over our research work until the seasonal return of Tunesian epidemics, when we could obtain the virus from human beings and preserve it for some weeks by passing it into monkeys, the only animal that we knew to be sensitive to it. The sensitivity of the guinea-pig, this useful animal, afforded us a convenient means for preserving the virus indefinitely by passing it into species.

Experimental typhus of the guinea-pigs is a very minimal disease. It is reduced to small changes in the temperature curve, and could not be diagnosed without a thermometer, since the animal does not seem to suffer or to have any other symptoms.

Now it happened occasionally that we discovered amongst our guinea-pigs, inoculated with the same virus, some who had no fever at all. The first time we discovered this, we thought it was do to an accident in the inoculation or to the particular resistance of the inoculated animal. These were the two hypotheses by which all bacteriologists of that era would have explained this phenomenon.

When the phenomenon kept recurring, we felt that our explanation had been too superficial, and that it must be due to another specific reason. We kept in mind the table of sensitivety to typhus of various races and species we had observed or infected. At the top of the scale was the European adult who had

immigrated to regions where typhus was endemic, and in whom the disease is most severe and often fatal. Below him appeared the aboriginal adult who is seriously infected but who, when there are no complications, generally escaped death. Then there comes the indigenous child for whom typhus, with few exceptions, is only a mild disease. Below our species there figures the chimpanzee, less sensitive still than the child, followed by the even less sensitive small monkeys, and finally the guine-pig whose infection is reduced to a thermometer curve. Could there not be below this hardly recognizable disease an even smaller degree of sensitivity, where, in the absence of fever, the only means of diagnosing typhus would be the positive results of an inoculation of blood into an animal of definitive sensitivity? That this was the case was proved by experiment. Other experiments very soon proved to us that latent typhus, exceptional as it was in the case of guinea-pigs, was the only form of typhus in some other species.

This latent typhus which we were the first to discover is a typhus of first infection. We are able to demonstrate the existence of the same sub-clinical type in other guinea-pigs, that had had primary typhus and were then reinoculated. The natural recurrence in man can also be of the sub-clinical type.

Subsequently we, as others after us, extended the notion of latent infection to a number of bacterial infections. The list increases daily.

Thus there exists a whole pathology that cannot be reached by clinical methods. If we add that it is in these unrecognizable forms that contagious and epidemic diseases are preserved, the practical importance of this new information is obvious. Now the starting point of our discovery has been the simple absence of a temperature rise in some examples of a species which commonly becomes feverish after being inoculated with a virus.' (R. Taton, *Reason and Chance in Scientific Discovery*, translated by A.J. Pomerans, The Scientific Book Guild, Beaverbrook Newspapers Limited, London, 1960, p. 81-2)

4. Several different surprising observations

Antrax

'Pasteur was curious to know how antrax persists endemically, recurring in the same fields, sometimes at intervals of several years. He was able to isolate the organisms from the soil around the graves in which sheep dead of the disease had been buried as long as 12 years before. He was puzzled as to how the organism could resist sunlight and other adverse influences so long. One day walking in the fields he noticed a patch of soil of different colour from the rest and asked the farmer the reason. He was told that the sheep

dead of antrax had been buried there the previous year.

'Pasteur, who was always examining things closely, noticed on the surface of the soil a large number of worm castings. The idea then came to him that in their repeated travelling from the depth to the surface, the worms carried to the surface the earth rich in humus around the carcase, and with it the antrax spores it contained. Pasteur never stopped at ideas but passed straight to the experiment. This justified his forecast. Earth contained in a worm, inoculated into a guinea-pig produced antrax.'

This is a fine exemple of the value of direct personal observation. Had Pasteur done his thinking in an armchair it is unlikely that he would have cleared up this interesting bit of epidemiology.' (W. I. B. Beveridge, *The Art of scientific Investigtation*, Mercury Books, London, 1961, p. 96. The quote is from E. Roux, as quoted by E. Duclaux in *Pasteur: Histotre d'un Esprit*, Sceaux, Paris, 1996)

5. Successful error

The removal of the cataractous lens

The most successful error in ophthalmology was made on 8 april 1745 in Marseille by the French Jacques Daviel (1693-1763) who did his umpteenth couching at poor Brother Felix, Hermit of Aiguille, in Provence, who already lost his right eye by a couching that failed. Couching was the conventional operation to treat cataract at that time. *Coucher* in French means 'to let lay down', in this case: the cataractous lens. Daviel pricked his sharp couching-needle in the left eye to push the cataractous lens behind the pupil downwards, but the lens broke. The anterior chamber of the eye, the space behind the cornea, filled itself with fragments of the lens and with blood. Daviel could no longer see the needle with which he had tried to press down the lens and had to stop the operation. Happily enough he remembered himself how Jean-Louis Pourfours du Petit, in 1708, had removed, with success, a lens, that had fallen in the anterior chamber, through an incision in the transparent cornea. So Daviel opened the transparant cornea with a small pair of scissors, to remove the blood and the fragments of the cataractous lens from the anterior eye chamber. He succeeded, the pupil became clear, the patient distinguished at once all objects held before him. But the eye was lost two days later by purulence. The process of infection was not known. Daviel explained the purulence with a beautifull image: 'The eye had been too tired.'

Daviel decided then not to push the cataractous lens downward behind the iris, but remove it at once via the anterior eye chamber. The first five cases went well, the first female patient recovered in two weeks. But then there

came complications. Daviel changed his mind, he regarded the method of removing the lens too deficient.

Then, on 8 April 1747, he wanted to couch the right eye of the barber and wigmaker Garion, whose left eye was already lost by couching. Daviel didn't manage to push down the lens, it broke. He removed the pieces of the lens from the space behind the iris, again through an incision in the cornea. Then Daviel started again with planned cataract extractions: 'First in some cases to get used to it gradually.'

But in the autumn of 1750 he decided to operate cataract only by removing the lens.

Already in 1752 Daviel gave two lectures for the *Académie Royale de Chirurgie* in Paris. In April he mentioned his technique of removing the lens and in November, during his prescribed second lecture, he gave his newest results: from the 206 lens extractions 182 had succeeded, 88%!

The removal of the cataractous lens has become the classic operation for cataract. Daviel's was revolutionary and took seven and a half years to develop it. In that time he perfected and compared the conventional couching with a needle and the removal of the cataractous lens. He did this, indefatigable, also by treating eyes, that were taken out of human cadavers, by alternating the two methods and by comparing the anatomical results precisely. (Serendipity in Ophthalmology, by I. Givner & A. Lachterman, *American Journal of Ophthalmology*, Vol. 63, No. 3, Part I, March, 1967, and: Daviel, J.: *American Encyclopedia of Ophthalmology*. 1917, Vol. II)

6. Successful accident

'Induced sensitization'

Charles Robert Richet (1850-1935), from 1887 onward professor of physiology in the Faculty of medicine in Paris, gave his Nobel Lecture on December 11, 1913, under the title 'Anaphylaxis':

'[...] *Phylaxis*, a word seldom used, stands in the Greek for protection. *Anaphylaxis* will thus stand for the opposite. Anaphylaxis, from its Greek etymological source, therefore means the state of an organism in which it is rendered hypersensitive, instead of being protected. [...]

These are the circumstances under which I first observed the phenomenon. You will allow me to go into some details on the origins. You will find that it is by no means the result of profound thought but a simple observation, almost a fortuitous one; so that my merit has only been in letting myself see the facts which were plain before me.

In tropical waters, Coelenterata are to be found floating on the surface, also

known as Physalia (Portuguese galleys). The basic structure of these creatures is a pocket filled with air so that they can float like a bladder. A bucco-anal cavity is subjoined to this pocket, with very long tentacles which hang in the water. These feelers sometimes run to two or three meters long and are equipped with small devices which adhere like sucking cups to objects encountered. Within each of these innumerable suction-cups is a pin-point which drives into the foreign body that is being touched. At the same time, this pin-point causes penetration of a subtle but strong poison, which is contained in the tentacles, so that contact with a feeler of the Physalia is tantamount to a multiple injection of poison. On touching a Physalia an acute sensation of pain is felt immediately, due to the penetration of this liquid venom. This is similar in relative intensity to a swimmer's mishap when he bumps into a jelly-fish in the water.

During a cruise on the yacht of Prince Albert of Monaco, the Prince advised me to study the Physalia poison, together with our friends Georges Richard and Paul Portier. We found that it is easily dissolved in glycerol and that by injecting this glycerol solution, the symptoms of Physalia poisoning are reproduced.

When I came back to France and had no more Physalia to study, I hit upon the idea of making a comparative study of the tentacles of the Actinia (*Actinia equina*, *Anemone sulcata*) which can be obtained in large quantities, for Actinia abound on all the rocky shores of Europe.

Now Actinia tentacles, treated with glycerol, give off their poison into the glycerol and the extract is toxic. I therefore set about finding how toxic it was, with Portier. This was quite difficult to do, as it is a slowly acting poison and three or four days must elapse before it can be known if the dose be fatal or not. I was using a solution of one kilo of glycerol to one kilo of tentacles. The lethal dose was of the order of 0.1 liquid per kilo live weight of subject.

But certain of the dogs survived, either because the dose was not strong enough or for some other reason. At the end of two, three or four weeks, as they seemed normal, I made use of them for a new experiment.

An unexpected phenomenon arose, which we thought extraordinary. A dog when injected previously even with the smallest dose, say 0.005 liquid per kilo, immediately showed serious symptoms: vomiting, blood diarrhoea, syncope, unconsciousness, asphyxia and death. The basic experiment was repeated at various times and by 1902 we were able to state three main factors which are the corner-stone of the history of anaphylaxis: (1) a subject that had a previous injection is far more sensitive than a new subject; (2) that the symptoms characteristic of the second injection, namely swift and total depression of the nervous system, do not in any way resemble the symptoms

characterizing the first injection; (3) a three or four week period must elapse before the anaphylactic state results. This is the period of incubation.

Once these first factors in anaphylaxis were well grounded, the field opened right up, thanks to the skilled and fruitful research of many investigators.

In 1903 Arthus, in Lausanne, showed that a first intravenous injection of serum on a rabbit causes anaphylaxis, i.e. three weeks after the first injection the rabbit is hypersensitive to the second injection. The phenomenon of anaphylaxis was becoming of general application. Instead of applying only to toxins and toxalbumins, it held good for all proteins, whether toxic at the first injection or not.

Two years later Rosenau and Anderson, two American physiologists, demonstrated in a noteworthy piece of work that the phenomenon of anaphylaxis occurs after every injection of serum, even when the injection is minute, for example of 0.00001 ml which is infinitely small amount but nevertheless sufficient to anaphylactize an animal. They quoted examples of anaphylaxis from all organic liquids: milk, serum, egg, muscle extract. They specified the reaction and clearly showed that of all subjects, the guinea-pig appeared the most sensitive in anaphylactic terms.

In 1907 I conducted an experiment which shed much light on the pathology of anaphylaxis. An anaphylactic state is produced by taking the blood of an anaphylactized animal and injected it into a normal animal subject. The anaphylactogen poison is therefore a chemical substance contained in the blood.

Such are, I think, the main stages through which our knowledge has passed. [...]’ (*Nobel Lectures, Physiology or Medicine, 1901-1921*, Elsevier, Amsterdam, p. 473-6)

7. From side-effect to main effect

‘A mixture of lime and copper sulphate was sprayed on posts supporting grape vines in Medoc with the object of frightening away pilferers. The French botanist Pierre-Marie-Alexis Millardet (1838-1902) later noticed, in October 1882, that leaves accidentally sprayed with the mixture were free from mildew. The following up of this clue led to the important discovery of the value of Bordeaux mixture in protecting fruit trees and vines from many diseases caused by fungi. It was the first fungicide to receive large-scale use the world over and can be said to have started a new era in the technology of agriculture.’ (W. I. B. Beveridge, *The Art of Scientific Investigation*, Mercury Books, London, 1961, p. 161-2. His source: G. W. Snedecor, *Statistical Methods applied to Experiments in Agriculture and Biology*.)

Collegiate Press Inc., Ames, Iowa)

8. From by-product to main product ('spin-off')

Tefal no-stick frying pan

'The Tefal no-stick frying pan was invented in December 1954. A French engineer Louis Hartmann, had been carrying out research for a number of years on PTFE (polytetrafluorethylene [Teflon, serendipitously discovered by Roy Plunkett in 1938 at du Pont]), a compound remarkable for its sliding and non-sticking properties. Hartmann discovered a process whereby one could fix a thin layer of PTFE on an aluminium surface. This was no mean feat, as to this time no metal would stick to PTFE.

A colleague, Marc Gregoire, then had the idea of using Hartmann's process for fixing PTFE on an aluminium frying pan. Gregoire founded, in 1956, the Tefal Company, in order to market this invention which has transformed the life of cooks.' (V. A. Giscard d'Estaing, *The world almanac book of inventions*, World Almanac Publications, New York, 1985, p. 154)

9. Alternative use

Oil colours and a wine-press to print books

'Drawing on established technologies Gutenberg perfected a method of casting single type from metal sufficiently uniform in size that they could be clamped together in a frame to give a surface from which it was possible to print. Water-based inks will not coat metal evenly and so Gutenberg developed from Flemish painters' oil-colours a suitable ink for printing from metal type. To press the paper uniformly against the inked types Gutenberg adapted the screw press along used for such tasks as pressing grapes.' (anonymous text, Science museum London, 1979)

10. Wrong hypothesis

Radioactivity

The French physicist Professor Antoine Henri Becquerel (1852-1908), in his Nobel Lecture, December 11, 1903 'On radioactivity, a new property of matter':

'[...] At the beginning of 1896, on the very day that news reached Paris of the experiments of Röntgen and of the extraordinary properties of the rays emitted by the phosphorescent walls of Crookes' tubes, I thought of carrying out research to see whether all phosphorescent material emitted similar rays.

The results of the experiment did not justify this idea, but in the research I encountered an unexpected phenomenon.

Of all the phosphorescent materials, uranium salts seemed particularly suitable for the investigations, because of the exceptional structure indicated by the harmonic series of the bands making up their absorption and phosphorescence spectra. [He had inherited from his father, who had done research on solar radiation and on phosphorescence, a supply of uranium salts, which phosphoresce on exposure to light.] Thus I placed sheets of double sulphate of uranium and potassium on photographic plates enveloped in black paper or protected by a sheet of aluminium and exposed them to light for several hours. On developing the plates, I found that the uranium salt had emitted rays which reproduced the silhouettes of the crystalline sheets through the black paper and various screens of metal or thin glass laid on the plates.

Under these conditions the phenomenon could be ascribed to a transformation of solar energy, like phosphorescence, but I soon recognized that the emission was independent of any familiar source of excitation, such as light, electricity or heat.

We were thus faced with a spontaneous phenomenon of a new order. Figure 1 shows the first print, which revealed the spontaneity of the radiation emitted by the uranium salt. The rays passed through both the black paper which enveloped the plate, and a thin sheet of copper in the shape of a cross. [...] (Nobel lectures, *Physics*, 1901-1921, Elsevier, Amsterdam, 1967, p. 52)

Becquerel didn't speak about the fact that when Röntgen had published that X-rays can give fluorescence (=phosphorescence) in certain compounds (in that way Röntgen discovered them, as known. Henri Poincaré, the greatest French physicist at that time, assumed that the phosphorescence of the cathode tube and the X-rays were one phenomenon. Becquerel heard that and reasoned that light rays perhaps could be transferred into deeply penetrating radiations, since he believed that cathode rays in the Crookes tubes were transferred into X-rays. To test this Becquerel started many experiments and concluded that the radiation must be coming from an element hitherto undiscovered. The phenomenon was called radioactivity. At this point he offered Madame Curie as her doctor's thesis problem the search for the element that possessed this peculiar property.

11. No hypothesis

'I therefore made experiments to see things about which I had absolutely

no preconceived idea.'

The French physiologist Claude Bernard (1813-1878) experimented on 'curare' and wrote :

'In 1845, Monsieur Pelouze gave me a toxic substance, called *curare*, which had been brought to him from America. We then knew nothing about the physiological action of this substance. From old observations and from the interesting accounts of Alex. von Humboldt and of Roulin and Boussingault, we knew only that the preparation of this substance was complex and difficult, and that it very speedily kills an animal if introduced under the skin. But from the earlier observations I could get no idea of the mechanism of death by curare; to get such an idea I had to make fresh observations as to the organic disturbances to which this poison might lead. I therefore made experiments *to see* things about which I had absolutely no preconceived idea. First, I put curare under the skin of a frog: it died after a few minutes; I opened it at once, and in this physiological autopsy I studied in succession what had become of the known physiological properties of its various tissues. [...] Now in my frog poisoned with curare, the heart maintained its movements, the blood was apparently no more changed in physiological properties than the muscles, which kept their normal contractility. But while the nervous system had kept its normal anatomical appearance, the properties of the nerves had nevertheless completely disappeared. There were no movements, either voluntary or reflex, and when the motor nerves were stimulated directly, they no longer caused any contraction in the muscles. To learn whether there was anything accidental or mistaken in this first observation, I repeated it several times and verified it in various ways; for when we wish to reason experimentally, the first thing necessary is to be a good observer and to make quite certain that the starting point of our reasoning is not a mistake in observation. In mammals and in birds, I found the same phenomenon as in frogs, and disappearance of the physiological properties of the motor nervous system became my constant fact. Starting from this well established fact, I could then carry analysis of the phenomena further and the mechanism of death from curare. I still proceeded by reasonings analogous to those quoted in the above example, and, from idea to idea and experiment to experiment, I progressed to more and more definite facts. I finally reached this general proposition, that *curare causes death by destroying all the motor nerves, without affecting the sensory nerves.*

In cases where we make an experiment in which both preconceived idea and reasoning seem completely lacking, we yet necessarily reason by syllogism without knowing it. In the case of curare, I instinctively reasoned in the following way: no phenomenon is without a cause, and consequently no

poisoning without a physiological lesion peculiar or proper to the poison used; now, thought I, curare must cause death by an activity special to itself and by acting on certain definite organic parts. So by poisoning an animal with curare and by examining the properties of various tissues immediately after death, I can perhaps find and study the lesions peculiar to it.

The mind, then, is still active here, and an experiment in order to see is included, nevertheless, in our general definition an experiment. In every enterprise, in fact, the mind is always reasoning, and, even when we seem to act without a motive, an instinctive logic still directs the mind. Only we are not aware of it, because we begin by reasoning before we know or say that we are reasoning, just as we begin by speaking before we observe that we are speaking, and just as we begin by seeing and hearing before we know what we see or what we hear.' (Cl. Bernard, *An introduction to the study of experimental medicine*, translated by H. C. Greene, Dover Publications, Inc, New York, 1957, p. 157-59)

12. Inversion

The result was therefore precisely the reverse!

Claude Bernard, in the same book, written during his nervous breakdown :

'About the year 1852, my studies led me to make experiments on the influence of the nervous system on the phenomena of nutrition and temperature regulation. It had been observed in many cases that complex paralysees with their seat in the mixed nerves are followed, now by a rise and again by a fall of temperature in the paralyzed parts. Now this is how I reasoned, in order to explain this fact, basing myself on known observations and then on prevailing theories of the phenomena of nutrition and temperature regulation. Paralysis of the nerves, said I, should lead to cooling of the parts by slowing down the phenomena of combustion in the blood, since these phenomena are considered as the cause of animal heat. On the other hand, anatomists long ago noticed that the sympathetic nerves especially follow the arteries. So, thought I inductively, in a lesion of a mixed trunk of nerves, it must be the sympathetic nerves that produce the slowing down of chemical phenomena in capillary vessels, and their paralysis that then leads to cooling the parts. If my hypothesis is true, I went on, it can be verified by severing only the sympathetic, vascular nerves leading to a special part, and sparing the others. I should then find the part cooled by paralysis of the vascular nerves, without loss of either motion or sensation, since the ordinary motor and sensory nerves would be still intact. To carry out my experiment, I therefore sought a suitable experimental method that would

allow me to sever only the vascular nerves and spare the others. Here the choice of animals was important in solving the problem; for in certain animals, such as rabbits and horses, I found that the anatomical arrangement isolating the cervical sympathetic nerve made this solution possible.

Accordingly, I severed the cervical sympathetic nerve in the neck of a rabbit, to control my hypothesis and see what would happen in the way of change of temperature on the side of the head where this nerve branches out. On the basis of a prevailing theory and of earlier observation, I had been led, as we have just seen, to make the hypothesis that the temperature should be reduced. Now what happened was exactly the reverse. After severing the cervical sympathetic nerve about the middle of the neck, I immediately saw in the whole of the corresponding side of the rabbit's head a striking hyperactivity in the circulation, accompanied by increase of warmth. The result was therefore precisely the reverse of what my hypothesis, deduced from theory, had led me to expect; thereupon I did as I always do, that is to say, I at once abandoned theories and hypothesis, to observe and study the fact itself, so as to define the experimental conditions as precisely as possible. Today my experiments on the vascular and thermo-regulatory nerves have opened a new path for investigation and are the subject of numerous studies which, I hope, may some day yield really important results in physiology and pathology. This example, like the preceding ones, proves that in experiments we may meet with results different from what theories and hypothesis lead us to expect. But I wish to call more special attention to this third example, because it gives us an important lesson, to wit: without the original guiding hypothesis, the experimental fact which contradicted it would never have been perceived. Indeed, I was not the first experimenter to cut this part of the cervical sympathetic nerve in living animals. Pourfour du Petit performed the experiment at the beginning of the last century and discovered the nerve's action on the pupil, by starting from an anatomical hypothesis according to which this nerve was supposed to carry animal spirits to the eye. Many physiologists have since repeated the same operation, with the purpose of verifying or explaining the changes in the eye which Pourfour du Petit first described. But none of them noticed the local temperature phenomenon, of which I speak, or connected it with the severing of the sympathetic nerve, though this phenomenon must have occurred under the very eyes of all who, before me, had cut this part of the sympathetic nerve. The hypothesis, as we see, had prepared my mind for seeing things in a certain direction, given by the hypothesis itself; and this is proved by the fact like the other experimenters, I myself had often divided the cervical sympathetic nerve to repeat Pourfour du Petit's experiment, without perceiving the fact of heat production which I later discovered when an

hypothesis led me in this direction. Here, therefore the influence of the hypothesis could hardly be more evident; we had the fact under our eyes and did not see it because it conveyed nothing to our mind. However it could hardly be simpler to perceive, and since I described it, every physiologist without exception has noted it and verified it with the greatest ease.

To sum up, even mistaken hypotheses and theories are of use in leading to discoveries. This remark is true in all sciences. The alchemists founded chemistry by pursuing chimerical problems and theories which are false. In physical science, which is more advanced than biology, we might still cite men of science who make great discoveries by relying on false theories. It seems, indeed, a necessary weakness of our mind to be able to reach truth only across a multitude of errors and obstacles.

What general conclusions shall physiologists draw from the above examples? They should conclude that in the present state of biological science accepted ideas and theories embody only limited and risky truths which are destined to perish. They should consequently have very little confidence in the ultimate value of theories, but should still make use of them as intellectual tools necessary to the evolution of science and as suitable for the discovery of new facts. (Cl. Bernard, *An introduction to the study of experimental medicine*, translated by H.C. Greene, Dover Publications, Inc, New York, 1957, p. 168-70)

13. Testing popular 'belief'

A vague opinion prevailed that Cow Pox was a preventive of the Small Pox

In the history of smallpox immunisation the years c. 1750-1798 may be viewed as the age of inoculation, when the medical innovation of causing mild illness to prevent a major attack had overcome the considerable opposition of its early years to become a widely accepted practice. [...] The first scientific experiment with cowpox was carried out in 1796 by Edward Jenner (1749-1823), a successful Gloucestershire doctor and experienced inoculator, who had been interested in the subject since about 1775. Writing in 1801 he recalled how :

'My inquiry into the nature of Cow Pox commenced upwards of twenty five years ago. My attention to this singular disease was first excited by observing, that among those whom in the country I was frequently called upon to inoculate, many resisted every effort to give them the Small Pox. These patients I found had undergone a disease they called the Cow Pox, contracted by milking cows affected with a peculiar eruption on their teats.

On inquiry, it appeared that it had been known among the dairies time immemorial, and that a vague opinion prevailed that it was a preventive of the Small Pox.

This opinion I found was, comparatively, new among them: for all the older farmers declared that they had no such idea in their early days - a circumstance that seemed easily to be accounted for, from my knowing that the common people were very rarely inoculated for the Small Pox, till that practice was rendered general by the improved method introduced by the Suttons. So that the working people in the dairies were seldom put to the test of the preventive powers of the Cow Pox.'

On 14 May 1796 Jenner vaccinated an eight-year-old boy, James Phipps, with material from a cowpox sore on the hands of Sarah Nelmes, the daughter of a local farmer, and on 1 July inoculated him with matter from a smallpox pustule, several incisions being made on both arms. No smallpox symptoms resulted and Jenner then submitted a paper to the Royal Society. Further experiments and observations followed; in 1798 he vaccinated John Baker, aged five years, with material from an infected cow and he recorded 17 other cases of people who appeared to have obtained immunity from smallpox by contracting cowpox directly from cows. Later in 1798 Jenner published his *Inquiry into [...] Variolae Vaccinae [...] known by the name of cowpox*. The idea of calling cowpox smallpox of the cow and giving it a Latin name was Jenner's. His argument was based on the propositions that those who had suffered from naturally acquired cowpox were immune from smallpox, that cowpox could be carried from arm to arm and that all who had been vaccinated would be protected from smallpox. The advantages of cowpox over smallpox lymph were that it was safer to use, generally caused a much less severe illness and the treatment was cheaper to carry out. Moreover, unlike smallpox, cowpox was not transmitted aurally and the use of cowpox virus therefore did not entail the risk of spreading disease among unprotected members of the population. Jenner also believed (incorrectly) that cowpox gave lifelong immunity from smallpox. [...]

In 1967 the World Health Organisation embarked on an intensified global programme of smallpox eradication which it was hoped be achieved within ten years. This, the first ever world-wide eradication scheme, was brought to a successful conclusion in 1977.

Smallpox was the first disease for which a preventive measure was discovered. It caused an average of 6,000 deaths a year throughout the century and could scar or blind those who survived. Up to the mid-century (19th) variolation or inoculation ('buying the smallpox') was practiced. Material from smallpox postule was transferred to the arm of a healthy

person to give a mild form of the disease. Variolation was carried out by both medical practitioners and 'amateur' inoculators, and general inoculations were sometimes paid for by subscribers or from the poor rate.' (*The speckled monster*, J. R. Smith, 1987, Essex Record Office and Science museum text, anonymous, 1997, London)

14. Child, student, outsider or animal

Rosa Parks, a black woman who refused to go to the back of a bus

The American philosopher Richard Rorty asked in an interview, citing Havel

:
'History, what can you say about it? You know that story about Rosa Parks, a black woman who refused to go to the back of the bus because she was tired? Somehow the story appealed, the media managed to convince the whites in America to think about what it was for an old black woman to stand in a bus... the pain she felt at that moment. From then on it was very difficult to maintain the segregation of races. Black had felt that pain for years and years. But nobody saw it. And now all of a sudden we all saw it ...

You can't plan such things. It happens once upon a time that a person or happening becomes paradigmatic and then something changes. But even what changes appears unpredictable. Together with the civil rights movement a black underclass has been formed. And the gap between the black middle class and the black underclass has become larger and larger.' (De Volkskrant, the Dutch 'Gardian', a newspaper)

15. Chance encounter of two professionals

To share an umbrella and a table with a French student

The Indian author Bhagwan S. Gidwani wrote in 1976 :

'It was some 13 years ago that I began collecting material on Tipu Sultan. A chance conversation led to my interest in him. It so happened that whilst in London, a French student and I were leaving the British Museum at the same time. We nodded to each other as strangers would. I offered him the protection of my umbrella against the slight drizzle which had just begun. We were both headed for the nearby restaurant and there we shared a table. It was then that I learned that he was not there merely for sight-seeing as I was but had interest of a more serious kind. He had come to the museum in search of material for a thesis which he was writing on kings and emperors who had died in battle. There were so few of them, according to him, and it seemed to be his complaint that in the face of defeat, kings often chose either to surrender or to flee in order to live or to fight another day. I continued to

listen somewhat uninterested until he caught my interest with his concluding remark : 'But your Tipu Sultan was the one who died on a battlefield - what a great man he was.'

With a courteous smile, I acknowledged his compliment to my countryman though silently I remembered that the history books imposed on me during my school and college days had not rated Tipu Sultan very high.

The observation of the French student remained in my mind. [...] my gratitude is to the French student who initially inspired this research. I am sorry that I had not asked him his name, nor had he asked for mine.' (From the Author's Note of *The Sword of Tipu Sultan (A historic novel about the life and legend of Tipu Sultan of India)* by Bhagwan S. Gidwani, Allied Publishers Limited, 1976, p. XIII -XVII)

16. Disturbance

Anti-oxidants

'The first interesting products preventing the unwanted and fast oxidation of food, caoutchouc (natural rubber), etc., were found by Moureu and Dufraisse, investigating the modification of acrolein, to be used as war gaz during the first World war. A crude acrolein (containing certain impurities) conserved itself better than the purified product.' (Ch. Moureu et Ch. Dufraisse, *Comptes rendus de l'académie des sciences*, 1922, 174, p. 258; *Chemical reviews*, 1926, 3, p. 119, as mentioned by Jean Jacques, *L'imprévu ou la science des objets trouvés*, Editions Odile Jacob, Paris, 1990, p. 177)

17. Scarcity, emergency, 'kairos'

'which I judged unfit'

About a battle in 1536 in Italy during the third war between Francois I and Charles V of Spain and Austria the French military surgeon Ambroise Paré (1510-1590) wrote:

'The first Discourse wherein wounds made by gunshot are freed from being burnt or cauterized according to *Vigoes methode*'

'In the yeare of our Lord 1536. *Francis* the French King [...] sent a puissant Army beyond the Alpes [...]. I was in the Kings Army the Chirurgion of Monsieur of *Montejan* Generall of the foote. The Imperialists had taken the straits of *Suze*, the Castle of *Villane*, and all the other pasages; so that the Kings army was not able to drive them from their fortifications but by fight. In this conflict there were many wounded on both sides with all sorts of weapons, but cheefely with bullets. I will tell the truth, I was not very expert

at that time in matters of Chirurgery; neither was I used to dresse wounds made by Gunshot. Now I had read in *John de Vigo* (*Lib. I de Vulner. Cap. 8.*) that wounds made by Gunshot were venenate or poisoned, and that by that reason of the Gunpowder; Wherefore for their cure, it was expedient to burne or cauterize them with oyle of Elders scalding hot, with a little Treacle (Treacle, i.e. theriacs, or antidotes to poison) mixed therewith. But for that I gave no great credite neither to the author, nor remedy, because I knew that caustickes could not be powdered into wounds, without excessive pain; I, before I would run a hazard, determined to see whether the Chirurgions, who went with me in the army, used any other manner of dressing to these wounds. I observed and saw that all of them used that Method of dressing which *Vigo* prescribes; and that they filled as full as they could, the wounds made by Gunshot with Tents and pledgets dipped in this scalding Oyle, at the first dressings; which encouraged me to doe the like to those, who came to be dressed of me. It chanced on a time, that by reason of the multitude that were hurt, I wanted this Oyle. Now because there were some left to be dressed, I was forced, that I might seeme to want nothing, and that I might not leave them undrest, to apply a digestive made of the yolke of an egge, Oyle of Roses, and Turpentine. I could not sleepe all that night, for I was troubled in minde, and the dressing of the precedent day, (which I judged unfit) troubled my thoughts; and I feared that the next day I should finde them dead, or at the point of death by the poyson of the wound, whom I had not dressed with the scalding oyle. Therefore I rose early in the morning, I visited my patients, and beyond expectation, I found such as I had dressed with a digestive onely, free from vehemencie of paine to have had good rest, and that their wounds were not inflamed, nor tumified; but on the contrary the others that were burnt with the scalding oyle were feaverish, tormented with much paine, and the parts about their wounds were swolne. When I had many times tryed this in divers others, I thought thus much, that neither I nor any other should ever cauterize any wounded with Gun-shot.' (*The apologie and treatise of Ambroise Paré containing the voyages made into divers places with many of his writings upon surgery.* Edited and with an introduction by Geoffry Keynes, 1951, Falcon Educational Books, London, p. ix, x, xi, 137-8)

18. Interruption of work

'Jump cut'

The value of pioniers is their example, not their number. Georges Méliès (1861-1938) was a film pionier who founded 'the cinematografic spectacle'. Between 1896 and 1912 he made 503 little movies full of tricks and fantasies. As a former illusionist he designed also his own decors and little film studio,

the first in the world. He worked there with ten people. He invented what is now called 'special effects'. About his first 'trouvaille', done in 1896, he wrote in 1907 :

'One wants to know how the first idea came to me to apply a trick to the cinematography? Very simple, I believe. A blocking-up of the apparatus that I used in the beginning (a rudimentary apparatus, in which the film often broke, hooked up and refused to move) produced an unexpected effect, a day I photographed prozaically the *Place de l'Opéra*: a minute was necessary to unblock the film and get the apparatus running again. During that minute the people that passed, the omnibusses, the vehicels, had changed from place, of course. Projecting the film, rewelded where the rupture was produced, I saw suddenly an omnibus *Madeleine-Bastille* changed in a hearse, and men changed in women. The trick by substitution, called stop trick [*truc à arrêt*], was found, and two days later I executed the first metamorphoses of men in woman and the first sudden disappearances, that had, in the beginning, such a success. It is thanks to this extremely simple trick that I executed my first spectacular plays: *Le Manoir du Diable*, *Le Diable au couvent*, *Cendrillon*, etc., etc.' ('Les vues cinématographiques', Causerie par Georges Méliès, Directeur du théâtre Robert-Houdin, *Annuaire Général et International de la Photographie*, Librairie Plon, Paris, 1907, pp. 363-92 + p. 385)

19. Playing

'That dull day on boring Ibiza appeared a revolution in building technique.'

'Around 1960 the French engineer Henri Vidal annoyed himself on the beach of this holiday island. As a right-minded engineer he could not stop to knock up, even during the holiday, so he played with sand and pine needles. As known, sand cannot tolerate load. It streams away like water. On that summer day Vidal asked himself whether he could stabilize sand. For example by putting a layer of pine needles in it, on regular distances. That appeared indeed possible. Where first the sand sank away, now, through the needles the hill of sand stayed together. That dull day on boring Ibiza appeared a revolution in building technique. Five years later Vidal presented his theory on '*terre armée*', a technique is now known worldwide. More than twelve thousand works are now made with *terre armée* like viaducts, land-abutments, quai walls and storage basins. The idea of Henry Vidal was based on the introduction of friction, not between the grains of sand, but between the grains of sand and the metal strips, horizontally placed in the mountain of sand. If you stand on such a mountain, then the grains want to slide, but they

are arrested by the metal strips. The friction between the sand and the strips of galvanized steel prevents the shoving away.' (*NRC Handelsblad*, the Dutch '*Le Monde*')

20. Joke and fun (initiating a hypothesis)

'Give 'em an ice cream and cake diet!'

'Six baboons are stars of a serendipity incident in San Antonio, Texas. Last spring [in 1957], Dr. N.T. Werthessen of the Southwest Foundation for Research and Education began studying these creatures, normally vegetarian, because of recent clues that they are the only known animals that can develop atherosclerosis similar to the disease in man. As he left the laboratory to take a trip, Dr. Werthessen told the animal keeper: 'Give 'em an ice cream and cake diet.' Today Dr. Werthessen says: 'It was just a figure of speech - I just meant for him to feed them well. But when I returned I found that the baboons really were given ice cream and cake - and, to our surprise, they ate it. As any French pastry cook knows, ice cream will cover up the taste of most foods. Now we expect to feed them just about anything we please, to study effects of diet on heart disease.' (*J.A.M.A.*, Milton Golin: *Medicine at work, Serendipity - Big word in medical progress, Does pure luck deserve all the credit?* Dec., 1957, p. 2086)

21. Dream or 'forgetting-hypothesis'

'This sudden illumination, I know of it'

The Frenchman Charles-Jules-Henri Nicolle (1866-1936) gave the account of his discovery of the transmission of typhus by fleas, a classical example of a sudden flash of creative illumination (he got the Nobel prize for Physiology or Medicine in 1928 for his work on typhus).

'This shock, this sudden illumination, this instantaneous selfcertainty of a new fact - I know of it, I have experienced it in my own life. It is in this way that the mode of transmission of exanthematic typhus was revealed to me [in 1909]. Like all those who for many years frequented the Moslem hospital of Tunis, I could daily observe typhus patients bedded next to patients suffering from the most diverse complaints. Like those before me, I was the daily and unhappy witness of the strange fact, that this lack of segregation, although inexcusable in the case of so contagious a disease, was nevertheless not followed by infection. Those next to the bed of a typhus patient did not contract the disease, while, almost daily, during epidemic outbreaks, I would diagnose contagion in the douars (the Arabic quarters of the town), and

amongst hospital staff dealing with the *reception* of patients. Doctors and nurses became contaminated in the country, in Tunis, but never in the hospital wards.

One day, just like any other, immersed no doubt in the puzzle of the process of contagion in typhus, in any case not thinking of it consciously (of this I am quite sure), I entered the doors of the hospital, when a body at the bottom of the passage arrested my attention.

It was a customary spectacle to see poor natives, suffering from typhus, delirious and fibrile as they were, gain the landing and collapse on the last steps. As always I strode over the prostrate body. It was at this very moment that the light struck me. When, a moment later, I entered the hospital, I had solved the problem. I knew beyond all possible doubt that this was it. This prostrate body, and the door in front of which he had fallen, had suddenly shown me the barrier by which typhus had been arrested. For it to have been arrested, and, contagious as it was in entire regions of the country and in Tunis, for it to have remained harmless once the patient had passed the Reception Office, the agent of infection must have been arrested at this point. Now, what passed through this point? The patient had already been stripped of his clothing and of his underwear; he had been shaved and washed. It was therefore something outside himself, something that he carried on himself, in his underwear, or on his skin, which caused the infection. This could be nothing but a louse. Indeed, it was a louse. The fact that I had ignored this point, that all those who had been observing typhus from the beginnings of history (for it belongs to the most ancient ages of humanity) had failed to notice the incontrovertible and immediately fruitful solution of the method of transmission, had suddenly been revealed to me. I feel somewhat embarrassed about thus putting myself into the picture. If I do so, nevertheless it is because I believe what happened to me it is a very edifying and clear example, such as I have failed to find in the case of others. I developed my observation with less timidity. At the time it still had many shortcomings. These, too, appear instructive to me.

If this solution had come to me with an intuition so sharp that it was almost foreign to me, or at least to my mind, my reason nevertheless told me that it required an experimental demonstration.

Typhus is too serious a disease for experiments on human subjects. Fortunately, however, I knew of the sensitivity of monkeys. Experiments were therefore possible. Had this not been the case I should have published my discoveries without delay, since it was of such immediate benefit to everybody. However, because I could support the discovery with a demonstration, I guarded my secret for some weeks even from those close to me, and made the necessary attempts to verify it. This work neither excited

nor surprised me, and was brought to its conclusion within some months.

In the course of this very brief period I experienced what many other discoverers must undoubtedly have experienced also, viz. strange sentiments of the pointlessness of any demonstration, of complete detachment of the mind and of wearisome boredom. The evidence was so strong, that it was impossible for me to take any interest in the experiments. Had it been of no concern to anybody but myself, I well believe that I should not have pursued this course. It was because of vanity and self-love that I continued. Other thoughts occupied me as well. I confess this failing. It did not arrest my research work. The latter, as I have recounted, led easily and without a single day's delay to the confirmation of the truth, which I had known ever since that revealing event, of which I have spoken.' (Charles Nicolle, *Biologie de l'invention*, Alcan, Paris, 1932, p. 56-60 translated by A.J. Pomerans in *Reason and Chance in Scientific Discovery* by R. Taton, The Scientific Book Guild, Beaverbrook Newspapers Limited, London, 1960, 76-8)

So far my model for 'accidental findings', an 'order' of 21 serendipity patterns in what was regarded as 'chaos'. The first 17 serendipity patterns I found were published in 1994 by the *British Journal for the Philosophy of Science* (45, 631-648). The physicist J. M. Campaniaro, from Madrid, interested in measuring the percentage of authors who mention of the role of serendipity in scientific articles, used my model and confirmed five (Nr. 2, 3, 4, 5 & 14) of these 17 patterns by reading 205 *Citation Classics* and later comments on them by the authors themselves (*Scientometrics*, 1996 Vol. 37, 1, 3-24). Now I am finishing a book '*Gardens of Serendip*', on request of the Editor Life Sciences of the Cambridge University Press.

MY RELATION WITH MY MODEL : QUESTIONS AND ANSWERS

Why did I create this model?

As microsurgeon and inventor, I worked twenty years in experimental ophthalmology, developing an artificial cornea (a French idea of Pellier de Quensey) for cornea blind people in economically poor countries that is now implanted daily in Amritsar in India. Just for fun, I began to collect cases of serendipity, and gradually my hobby became my main interest, because patterns of serendipity emerged from my collection of 'serendipities' that 'happened' mostly in the experimental sciences. I didn't look for any order in my alphabetical serendipity files, but the patterns of serendipity offered themselves spontaneously, as unsought 'heapings' of similar, comparable cases.

How I test my model?

The ways in which unsought findings are made are so evident and pertinent that they invite you, so to say, to call them patterns.

How I let the model develop itself?

By collecting, reading and studying authentic cases of serendipity, preferably written by the serendipitists themselves, in their own language (Dutch, English, French, German I can read myself). I judge on my own to which pattern they belong most, and in doing so I sometimes find a new pattern or a better formulation for an already described pattern.

What are the intrinsic limites of the model?

Well each case of serendipity is by definition unique, because it is a (serendipitous) original discovery, invention or creation. But in spite of that they have in common that the way they presented themselves to the serendipitist as a 'chance' observation is 'recognizable'. The limit of the model is that these ways, or patterns, can sometimes coexist, overlap and/or cooperate. So my list of 'patterns' is just a list and not a classification as in a flora, for example. But the fact that some serendipitous findings fit in more patterns at the same time does underlines the value of these patterns in a funny and ironic way.

How do I trust my model?

Well, I found, developed, changed, applied it myself. And it gets richer and richer in explaining and illustrating serendipity in science, technique and art.

To what extent do I believe it the model?

I don't think in terms of belief. The model helps me to analyse, describe, understand, explain, illustrate the large subject of serendipity in a broad field.

How do I change it?

By adding, redefining, splitting, deleting and by (re)combining patterns.

Does the model help to exclude other practices?

Loaded question! No, because, as far as I know there are no other practices or models that are excluded by the model.

How I use it?

It has a methodological function. 'Method' in Greek means 'μετα οδοσ' that is: 'road along which'. Serendipity is one road along which a discovery, invention or creation can be 'found' or 'done'. Patterns of serendipity enable

a searcher to recognize, understand and describe better and clearer the (serendipitous) road that was taken. It has also a didactical function when you publish or lecture about serendipity.

Is it an instrument of power?

A very French question, I think. No, what started as a academical interest, is now a model with a clear analytical and practical value, which gives the possibility to describe the anatomy of serendipity. I am interested in the instrumental value of the model of these patterns, not in the power of the idea, not in the idea of power. Personally I don't believe in the idea of power.

In what does my belief in this model transform?

Paradoxically: my belief in the model gives me the desire, the energy to develop and fine-tune the model.

In which respect do these practices of elaboration distinguish themselves from non-modern ones?

There is no pre-modern or pre-postmodern practice of elaboration of serendipity patterns. A serendipity case was seen, as by myself in the beginning, as a 'chance', 'random', 'fortuitous', 'accidental' event. But 'accidental' means in this case, as explained above, psychologically that the surprising observation 'falls' towards you, so to speak. It has nothing to do with the mathematical connotation of 'accident', 'random', 'fortuitous' or 'chance'.

QUOTES

I glued together a 'quote' of a Nobel prize winning physicist, the American Irving Langmuir about his laboratory policy regarding serendipity :

'That work wasn't planned. It was allowed to grow by curiosity, by having fun. It was partly because it was not planned that it was so successful. You can't plan to make discoveries, but you can plan work that will lead to discoveries. You can organize a laboratory so as to increase the probabilities that useful things will happen there. And in so doing, keep the flexibility, keep the freedom. We know from our own experience that in true freedom we can do things that could never be done through planning. Serendipity is the art of profiting from the unexpected. Freedom of opportunity as developed by democracy is the best human reaction to divergent phenomena. We may in fact, define freedom as the opportunity to profit from the unexpected.' (I. Langmuir (1888-1957): Freedom, the Opportunity to Profit from the Unexpected, *G.E. Res. Lab. Bull.* Schenectady (N.Y.), US :

General Electric Corporate Research & Development Publications, 1956, glued together by myself)

And at last, one British quote of N.W. Pirie about the relevance of the philosophy of a scientist :

‘Some people think that the philosophy a scientist accepts is not of very much importance; his job is to observe the phenomena. This is a gross oversimplification and involves the subsidiary hypothesis that all scientists are fully equipped with serendipity. A sensible philosophy controlled by a relevant set of concepts saves so much research time that it can nearly act as a substitute for genius; it may be that this is what we mean by genius. Thus equipped we avoid the pseudo-problems or, more correctly, the problems that are real in adjacent fields but pseudo in our own. A scientist can have no ore valuable skill than the ability to see whether the problem he is investigating exists and whether the concepts he is using are applicable. But the lure of a concept that clarifies another field is strong and many of us follow it, like the music of the Pied Piper, to our scientific detriment.’ (N.W. Pirie, Concepts out of context: the Pied Pipers of science, *British Journal for the Philosophy of Science*, 1952, II, 8, p. 280)

In my talk I will discuss, in French, about serendipity, its model and the possible role of both in discoveries, inventions and creations. To be honest I think it is time to introduce the word ‘serendipité’ in the French sciences. I did the same in the Dutch scientific world.

31-2-1997, Feerwerd, Groningen, The Netherlands, t.t. Pek van Aniel.

FIN

Psychopathologie, modèles et pouvoir

Martine TIMSIT, Christophe PARISSÉ & Evelyne ANDREEWSKY
<timsit.berthier@infoboard.be>, <parisse@ext.jussieu.fr>,
<andreews@ext.jussieu.fr>

1. Avant-propos : modèles et pouvoir - quelques repères

A propos de repères, Poirot-Delpech, dans *le Monde* d'il y a quelques mois, en proposait (des plutôt réjouissants), sur le mode "on reconnaît"¹, mode permettant de baliser rapidement ces repères.

On reconnaît le pouvoir à l'autorité qu'il confère - et celui qui le détient, à ce qu'il est le plus fort.

La raison du plus fort est toujours la meilleure.

On reconnaît un savant (de même qu'un pape) *à son infaillibilité* :

- William Thomson, futur Lord Kelvin, déclarait en 1895 que les machines volantes plus lourdes que l'air ne peuvent exister. Propos infaillibles, comme on sait ; depuis Aristote et sa théorie des graves, les corps lourds *doivent* en effet rejoindre leur lieu naturel (cf. J.-B. Grize).

- Et pourtant, elle tourne ! déclarait (in petto) Galilée, en dépit de sa confrontation avec l'infaillibilité pontificale ...

Depuis Galilée, excommunication et autres exclusions ne cessent de menacer ceux qui osent remettre en question les modèles-dogme des pouvoirs, scientifiques, religieux, ou autres.

On reconnaît un "modèle-dogme" - ou une *pensée fermée* (cf. E. Morin) - "à son arrogance et à sa suffisance scientiste".

C'est ainsi que pour certains, la pensée n'est ni plus ni moins que l'activité (électro-bio-chimique) du cerveau - ce qui les amène à proclamer (avec arrogance et suffisance) que le psychologique (et le social) ne peuvent et ne doivent être conçus qu'en termes de neurones ; d'où, entre

¹ex: *on reconnaît* un chirurgien à ce qu'il porte un noeud papillon", *on reconnaît* "les porteurs de noeud papillon à ce qu'ils s'imaginent à l'abri du ridicule".

autres exemples, un excellent modèle de la lecture et des dyslexies : l'activité des neurotransmetteurs ... (cf. Changeux²). Pourquoi pas ! La génétique voit bien dans le génome l'identité des individus, à l'instar d'ailleurs de la phrénologie, qui la voyait, elle, dans la morphologie de leur cerveau ...

On sait que la conception des relations entre cerveau et esprit n'a jamais été neutre. Elle est en effet en prise directe avec questionnements et certitudes de base sur nous même. On ne s'étonne donc pas de la trouver en filigrane non seulement dans les approches cognitives actuelles, mais dans toute l'histoire de la psychiatrie (cf. Postel & Quétel, 1983). Dans ces deux domaines, les *modèles scientifiques* - constructions *objectives* par définition - sont cependant marqués du sceau de ces interrogations. C'est ce qui explique sans doute que ces modèles, loin d'être neutres, sont susceptibles d'engendrer des outils redoutables, de *très subjectifs instruments de pouvoir*. Nous présentons, à partir d'exemples empruntés à la psycholinguistique et à la psychopathologie, quelques réflexions sur ces problèmes.

2. Si les faits contredisent la théorie, eh bien tant pis pour les faits ! (Hegel)

Pour illustrer le titre du présent paragraphe dans les domaines de la psycholinguistique et de la psychopathologie, citons des cas assez prototypiques, où des faits peu susceptibles d'être expliqués avec les modèles en vigueur, ont été éliminés (tant pis pour ces faits, ils sentaient par trop le soufre !).

2. 1. Phénomènes subliminaires.

En psycholinguistique, les modèles usuels considèrent la (ou les) significations des mots comme "figurant" dans un lexique interne, le "lexique mental". Si l'on s'en tient à cette hypothèse, un mot présenté à un sujet donné ne peut être compris, ou identifié par ce sujet, que dans la mesure où cette présentation a réussi à "activer" (ou "accéder à") la représentation interne de ce mot.

²Colloque "Rodin Remediation Acad.", Paris, 1993.

Dans un tel cadre théorique, les phénomènes *subliminares*³ sont difficiles à accommoder. En effet un mot affiché en subliminaire peut très bien, sans pour autant avoir été identifié par les sujets, avoir des effets liés à sa signification (voir note 3., pour un exemple). Les modèles de compréhension lexicale en termes d'accès en "tout ou rien" aux représentations des significations (telle une porte, qui ne peut être que ouverte ou fermée), ont à l'évidence beaucoup de mal à expliquer de tels phénomènes. Dans ces conditions, les tenants de ces modèles ne peuvent que dénier l'existence même des phénomènes subliminaires (ce qui a été fait à satiété), sauf à leur trouver une explication compatible avec le cadre théorique traditionnel. Ont ainsi été proposées des hypothèses parfaitement *ad hoc*, que l'on peut résumer comme suit : *le mot présenté en subliminaire serait dans un premier temps (très court) identifié ; sa signification peut alors paramétrer les mécanismes du lexique mental. Elle est ensuite immédiatement inhibée, ce qui fait que les sujets n'ont pas conscience de cette identification, cf. Marcel, 1983.*

Dans certains cas, l'hypothèse *ad hoc* précédente ne fonctionne pas. En effet, dans la mesure où l'inhibition de la signification des mots présentés en subliminaire est considérée comme due au trop court laps de temps de cette présentation, comment alors expliquer que des phénomènes très analogues peuvent avoir lieu avec des items présentés tout à loisir ? Il s'agit de phénomènes se produisant chez des individus devenus (à la suite d'une lésion hémisphérique) *totale*ment incapables de lire et comprendre l'écrit, autrement dit, totalement alexiques, mais qui sont néanmoins capables de trier les mots qui leur sont présentés (à loisir, sur papier), sur des critères sémantiques implicites.

A défaut de pouvoir expliquer ce nouveau phénomène, le mieux est sans doute de l'occulter ! C'est en tout cas ce qui s'est passé pour la première expérience ayant mis en évidence ce phénomène. Il s'agissait d'un patient présentant une "alexie totale", très soigneusement attestée à l'aide de nombreux tests psycholinguistiques de lecture, au V. A. Hospital, Boston, USA. Ce patient devenu parfaitement incapable de comprendre

³où on observe qu'un mot donné, affiché trop rapidement pour que les sujets soient en mesure de l'identifier (ou même de s'apercevoir que quelque chose se passait sur l'écran), entraîne cependant des conséquences liées à sa signification (ce qu'on peut objectiver en demandant aux sujets de lire à voix haute les items affichés sur l'écran. On observe, par exemple, que l'item "fils" est énoncé en général /fis/, à la suite d'un mot comme "père" affiché au préalable en subliminaire, mais /fil/, si le mot préalable était par exemple "aiguilles".

correspondre point par point le nom des propriétés thérapeutiques des nouvelles drogues (antipsychotiques, antidépresseurs et anxiolytiques), et la classification des maladies mentales (troubles psychotiques, troubles de l'humeur et troubles anxieux - cf. DSM III et IV).

Cette modélisation pharmacologique de la clinique a un corollaire logique : l'utilisation exclusive des médicaments dans le traitement des difficultés psychologiques. Ainsi, quand un patient présente une symptomatologie importante, associant par exemple dépression et anxiété, l'on parle de "co-morbidité" et on prescrit deux drogues au lieu d'une. Et lorsqu'on découvre qu'une même drogue est active dans des tableaux cliniques différents, "on ne laisse plus l'objet clinique s'imposer à l'observateur, mais au contraire, on modèle la saisie clinique par rapport à une finalité" (Partiot et al., 1991), et on pratique un découpage "transnosologique" à l'aide d'échelles comportementales et de psychopathologie quantitative, fabriquées à la lumière des concepts pharmacologiques. Actuellement, plus de 100 nouvelles molécules antidépressives et anxiolytiques sont en essai thérapeutique dans les hôpitaux belges ...

Une telle modélisation qui postule l'isomorphisme des différents niveaux d'observation (de la molécule à la neuro-anatomie, au comportement animal et au psychisme) ne permet que prescription de médicaments en réponse à des difficultés psychologiques, et mise en évidence de résultats convergents entre données psychologiques, neurobiologiques et pharmacologiques. Ces données témoignent toujours de la maladie en termes "d'hypofonctionnement", de "déficit", que l'on doit, naturellement, corriger avec des drogues. Cette modélisation se renforce chaque jour par les données, toujours dans le sens déficitaire, apportées par les nouvelles techniques d'imagerie (PET SCAN, IRM fonctionnelle) et par les résultats des essais thérapeutiques évalués positivement grâce à l'utilisation d'échelles construites "sur mesure".

Dans un tel cadre théorique, les résultats qui font apparaître des divergences entre niveaux d'observation différents, et qui montrent par exemple un fonctionnement cérébral normal ou "hyperactif" chez des patients déprimés et ralentis, sont difficiles à interpréter (cf. résultats apportés par la Neurophysiologie). De la même façon, l'importance de l'effet "placebo" dans la dépression (37,4% des cas, Lapierre, 1995) est systématiquement occulté. Et à défaut de pouvoir expliquer ces données qui remettent en question le modèle, l'attitude habituelle consiste à les exclure délibérément, à les taire, ou à les disqualifier.

Le bilan d'un tel modèle n'est pas réjouissant : appauvrissement de la clinique psychiatrique, absence de perspective de recherche, déshumanisation de la pratique ...

Pourtant, c'est le savoir issu de ce modèle qui est actuellement diffusé largement dans les chaires universitaires de Psychiatrie, dans les Congrès internationaux et dans la grande presse. Faut-il le souligner, tous ces organismes de diffusion du savoir sont largement financés par les firmes pharmaceutiques, et "les discours académiques ne se différencient guère des messages promotionnels de l'industrie" (Zarifian, 1997).

3. Pouvoir des modèles et modèle des modèles

On peut se demander comment des modèles élaborés dans des laboratoires scientifiques, qui devraient être source de savoir, peuvent ainsi se transformer en "modèles-dogmes" qui entraînent par leur immobilisme (et leur torpeur), une situation qui n'est pas favorable à l'évolution de la pensée et de la pratique.

La réponse ne se trouve pas dans la nature des modèles eux-mêmes mais dans le rôle que certains leur font jouer pour justifier et légitimer leur pouvoir. Si l'on en croit Sloterdijk, 1987, "Il n'y a plus de savoir dont on pourrait être l'ami. Il ne nous vient pas à l'esprit d'aimer le savoir. Nous nous demandons au contraire comment faire pour vivre avec lui sans nous pétrifier. Ce que nous présentons ici est une méditation sur la proposition : *savoir, c'est pouvoir*, proposition qui au XX^e siècle est devenue le fossoyeur de la philosophie ... avec elle prend fin la tradition d'un savoir qui comme l'indique son nom, était une théorie érotique - amour de la vérité et vérité de l'amour. Du cadavre de cette philosophie ont surgi au XIX^e siècle les sciences modernes et les théories du pouvoir".

Les relations de pouvoir entre la science et l'argent (public ou privé) ne fonctionnent que rarement sur un mode épistémologique - quel qu'il soit - mais elles s'établissent plutôt à partir des croyances et surtout des intérêts du décideur, et traduisent sa culture et souvent son attachement au concret, au réel et à la garantie de résultats. Tout ceci reflète les effets de l'appât du gain et de la peur de l'inconnu et n'est possible que parce que l'on peut en généralisant de façon abusive un modèle scientifique faire admettre des croyances culturelles.

Dans le cadre d'une telle utilisation n'importe quel modèle risque de devenir un "prêt-à-penser" susceptible de répondre à toutes les questions, non seulement à celles posées par une pratique scientifique

donnée, toujours limitée, mais aussi à celles soulevées par les autres pratiques, sociales et collectives. Et comme le souligne I. Stengers (1997b: 85), les modèles scientifiques courants utilisent souvent la simplicité comme argument d'efficacité, et non comme aveu d'ignorance ... Il devient nécessaire de les stériliser et de les scléroser afin d'y asseoir des certitudes. L'hérésie scientifique, avec excommunication (cf. par exemple l'affaire Benveniste) est sans doute une manifestation des pathologies psychosociologiques qui traduit le pouvoir des modèles utilisés dans une telle perspective. Il s'agit en effet de ne surtout pas remettre en question ces modèles-dogmes qui conditionnent la cohérence de nos interprétations du monde. Il y va peut-être de notre identité, mais surtout, de notre confort intellectuel ! *Ne touche pas à mon modèle !* confère à ce type de modèle un pouvoir exorbitant (il appartient en effet à une "méta-théorie sécurisante", qui mène droit à "la pensée unique" (Badiou, 1969).

Nous pouvons au contraire accepter la remise en question de nos modèles, et risquer de les confronter à d'autres modèles, élaborés dans d'autres disciplines (ce qui fait l'intérêt des approches pluridisciplinaires), en évitant "les exclusions actives des *ou bien ... ou bien*", Stengers, 1997). Il ne s'agit pas seulement de vouloir ce changement, il faut le provoquer avec une dynamique qui permet d'induire des constructions et des pratiques nouvelles, satisfaisant ainsi notre besoin compulsif de compréhension du monde et de nous-même. Les modèles conservent alors le pouvoir de nous rassurer, mais perdent celui d'inhiber notre pensée.

On peut essayer, à titre d'illustration, de faire interagir les deux modèles-dogmes de la psychopathologie : d'une part, le *modèle médical*, où il est admis que les perturbations du fonctionnement cérébral sont "causes premières" des perturbations du vécu et du comportement du sujet, indépendamment de son intentionnalité et de sa volonté, et de l'autre, le *modèle psychanalytique*, dans lequel le sujet est son propre objet d'étude, de réflexion et de savoir, et son propre thérapeute.

En neuropsychologie comme en psychiatrie, le patient, en se soumettant aux examens qu'on lui demande, confirme la légitimité du modèle médical. Ce type de modèle suppose une mise à distance et une soumission de celui qui est pris comme objet d'étude, mais qui est exclu du savoir dont il est l'objet (cf. I. Stengers, 1997b : 74). En quelque sorte, le patient "subit" à la fois le pouvoir des neuropsychologues, des médecins, et de son propre cerveau. Cependant, il faut bien reconnaître qu'il est aussi producteur de connaissance sur ce cerveau dont il ne semble avoir aucune maîtrise. Mais les résultats du savoir ainsi produit ne lui sont jamais

retournés et s'adressent exclusivement aux spécialistes compétents des Neurosciences Cognitives et de la Psychiatrie (a-t-on jamais vu les résultats d'un "essai thérapeutique", impliquant un nouveau médicament, discutés avec les patients ayant participé à cet essai ?). En revanche, le sujet qui entre en analyse ignore qu'il a un cerveau et ne veut pas le savoir.

Il est possible d'introduire un certain dynamisme interactif dans ces modèles en envisageant les relations entre cerveau et esprit. En effet, si l'état psychologique peut être affecté par des substances chimiques, l'inverse est non moins vrai, et une idée à caractère émotionnel, un souvenir, un chagrin, affecte également l'activité cérébrale. Dans cette perspective, il ne s'agit plus de considérer le patient comme le reflet d'un cerveau déficient, objet d'examen de plus en plus sophistiqués, et cible de substances de plus en plus spécifiques, mais comme un sujet, un alter-ego, susceptible de participer à l'élaboration du savoir qui concerne son cerveau. Cette modélisation doit prendre en compte à la fois les représentations du patient et celles utilisées par son interlocuteur. Elle peut déterminer non seulement un renouvellement du savoir, mais un certain effet thérapeutique.

Certes, il n'existe pas de solution "simple" au problème "complexe" du pouvoir des modèles. En effet, la capacité de ce pouvoir à masquer les difficultés marque toute tentative de modélisation. Et pour arriver à rompre l'engrenage des relations entre savoir et pouvoir, et supprimer les abus des modèles, il faudrait peut-être ressourcer notre pratique épistémologique dans une *culture* de la science et de ses mécanismes, et retrouver "un savoir dont on pourrait être l'ami" (Sloterdijk, 1987).

Références

- BADIOU A. (1969), *Le concept de modèle*, Maspero, Paris.
- DELOCHE G. & ANDREEWSKY E. (1982), "From Neuropsychological Data to Reading Mechanisms", *Int. J. of Psychology*, 17, 259-279.
- FOERSTER H., von (1988), "La Construction d'une réalité", in Watzlawick, P. (Ed.), *L'invention de la réalité : contributions au constructivisme*, Le Seuil, Paris, 45-69.
- HENDERSON L. (1987), Word Recognition, A Tutorial Review, in COLTHEART M. (ed.), *Attention and Performance XII, The Psychology of Reading*, Lawrence Erlbaum, London, 171-200.

- JOUVENT R. (1991), "Des hypothèses et des modèles en Psychopharmacologie des comportements". *Psychologie française*, 36-3, 263 p.
- LE MOIGNE J.-L. (1994), *Le Constructivisme, tome 1 : des fondements*, Paris, ESF.
- LAPIERRE Y. D. (1995), "Placebo : A potent but misunderstood Psychotrope", *J. Psychiatry Neurosci.*, 20, 3, 173-174.
- MARCEL T. (1983), "Conscious and Unconscious Perception : Experiments on Visual Masking and Word Recognition", *Cognitive Psychology*, Vol. 15, 197-237.
- MORIN E. (1987), *la Méthode*, Tome III - *La connaissance de la connaissance*, Vol. 1, Paris, Seuil.
- PARTIOT A., JOUVENT R., PIERSON A., EL MASSIOUI F. & WILDÖCHER D. (1991), "Potentiels évoqués tardifs et psychopathologie : des indices cognitifs de l'action de psychotropes", *Psychologie Française*, 36-3, 241-248.
- PEIRCE Ch. S. (1995), "Lectures of 1898", *Le raisonnement et la logique des choses*, Paris, Cerf.
- SLOTERDIJK P. (1987), *Critique de la raison cynique*, Christian Bourgeois éditeur, Paris.
- STENGERS I. (1997), *Cosmopolitiques, tome 7. "Pour en finir avec la tolérance"*, Editions La Découverte.
- STENGERS I. (1997b), *Sciences et pouvoirs*, Paris: Editions La Découverte.
- SWAIN G. (1994), *Dialogue avec l'Insensé*, Nrf.
- WINOGRAD T. & FLORES F. (1986), *Understanding Computers and Cognition*, Norwood, N. J., Ablex.
- ZARIFIAN E. (1996), *Le prix du Bien-Etre. Psychotropes et société*. Odile Jacob. Paris.

L'autonomie comme approche de nos interactions avec nos modèles : la clôture selon F. Varela ou selon C. Castoriadis ?

Guillaume Deffuant

Laboratoire d'Ingénierie des Systèmes Complexes

Cemagref

Parc de Tourvoie

92 185 Antony Cedex

Résumé : Nous proposons d'aborder nos relations avec nos modèles à l'aide d'une conception générale de la cognition, fondée sur l'autonomie. Pour cela, nous tentons d'articuler deux visions de l'autonomie : celle de F. Varela et celle de C. Castoriadis. Varela part de la cellule biologique comme exemple typique et associe l'autonomie à la création d'une clôture opérationnelle. Castoriadis se focalise sur l'homme en société en prenant comme exemple typique l'autonomie de la polis Grecque, associée à une rupture de la clôture des significations imaginaires sociales imposées de l'extérieur. L'articulation proposée repose tout d'abord sur une distinction entre clôture physique des opérations et clôture des significations (clôture du pour-soi). La clôture du pour-soi suppose l'autonomie, car elle est l'affirmation d'un monde autonome, pour-soi. Nous proposons une interprétation des étapes de l'évolution cognitive par l'instauration d'une nouvelle clôture des opérations (vision de Varela) mais aussi comme rupture d'une clôture du pour-soi, qui est remplacée par une clôture plus large (vision de Castoriadis). La rupture de la clôture des significations imaginaires sociales dans la polis Grecque est une rupture de la clôture du pour-soi des sociétés hétéronomes, et correspond donc à la création d'une clôture des opérations (les interactions sociales) d'un type nouveau (celui de la compétition entre égaux). Ces différents types de clôtures identifiés nous permettent de proposer une compréhension renouvelée de l'époque moderne. Notre thèse est donc que Castoriadis et Varela parlent bien de la même chose mais sous deux angles différents. Ce même cadre fournit un certain nombre de propositions pour comprendre le fonctionnement cognitif sous-jacent aux fétiches, faitiches (modèles scientifiques) et aux dieux.

Introduction

Ce texte aborde la question de nos relations en la confrontant à deux approches générales de l'autonomie et de la cognition : celle de F. Varela et celle de C. Castoriadis. Ce choix comporte des difficultés car des contradictions importantes entre ces deux auteurs semblent apparaître au premier abord. L'effort principal est donc porté sur la construction d'un point de vue permettant de lever ces contradictions apparentes. Cet effort nous permet au passage de progresser dans la compréhension de nos relations avec nos modèles.

L'autonomie selon Varela se focalise principalement sur la biologie, et part de la cellule biologique comme exemple typique. Castoriadis voit l'autonomie comme centrée sur l'homme en tant qu'être social, et utilise comme référence principale la démocratie Grecque antique. Au delà de l'éloignement de ces références, le terme « clôture » commun aux deux auteurs, semble avoir des sens très différents, voire incompatibles, dans les deux approches. En effet, selon F. Varela, l'autonomie est la *création* d'une clôture (opérationnelle), alors que chez C. Castoriadis, l'autonomie est *rupture* d'une clôture (des significations imaginaires sociales imposées dans les sociétés hétéronomes). De telles différences ou contradictions apparentes brouillent la vision globale que l'on pourrait avoir de l'autonomie.

De plus, au sein même de l'œuvre de chacun des deux grands auteurs, il nous semble que des ambiguïtés gênantes subsistent sur l'utilisation du terme clôture. Cela est surtout vrai pour Varela, dont le concept de « clôture opérationnelle » semble faire l'objet de nombreuses interprétations erronées. Ainsi, Castoriadis l'assimile-t-il dans l'un de ses textes à une clôture informationnelle, c'est à dire une absence d'interactions avec l'extérieur (Castoriadis 86b), ce qui est clairement contradictoire avec les textes de Varela. Dans le même texte, Castoriadis considère la vision de l'autonomie de Varela et la sienne comme opposées. Bien que Castoriadis semble avoir évolué ultérieurement sur ce point (Castoriadis 97), cette discussion révèle bien une difficulté. Nous tentons tout d'abord d'explicitier les ambiguïtés internes aux définitions de Varela. Pour ce faire, nous proposons une distinction, souvent implicite chez Varela, entre la clôture des opérations d'un processus, et la clôture cognitive d'un monde « pour-soi », d'une monade. Ces deux clôtures, qui peuvent paraître équivalentes dans certains textes de Varela doivent selon nous être clairement distinguées, car elles correspondent à des positions épistémologiques très différentes par rapport au système considéré.

La clôture dans son sens cognitif, c'est-à-dire clôture d'un monde pour soi, est celle qui nous permet d'établir une articulation entre l'autonomie chez Varela et chez Castoriadis. La thèse générale défendue est que les contradictions apparentes (création ou rupture de clôture) peuvent se résoudre dans une vision générale de la cognition faisant apparaître ces deux types de clôture et leurs transformations au cours des grandes étapes de l'évolution. Il s'agit donc d'une tentative très différente de celle de Paul Dumouchel, qui proposait un niveau intermédiaire d'autonomie entre les deux visions (Dumouchel 83). Notre thèse est que Varela et Castoriadis parlent d'une même chose, mais vue sous des angles différents.

Nous proposons donc de retracer les grandes étapes, les grandes ruptures décelables dans cette clôture du pour-soi, et les créations de clôtures opérationnelles correspondantes. La première est celle de l'émergence du vivant, avec la création du premier « monde pour soi », dont part Varela. Il s'agit donc de l'autonomie par création d'une clôture (opérationnelle et d'un monde pour soi). Nous proposons d'interpréter les autres étapes comme créations d'une nouvelle clôture opérationnelle, engendrant une nouvelle clôture du pour-soi qui englobe la (ou les) précédente(s), sans totalement la (les) détruire. Il s'agit donc de mondes pour-soi nouveaux, plus larges, mais qui peuvent coexister avec les anciens. Nous laisserons de côté la question de la création elle-même, essentielle pour Castoriadis. Le terme employé par Castoriadis de rupture de la clôture des significations imaginaires sociales opérée par la polis grecque serait donc selon nous similaire à d'autres ruptures repérables à d'autres étapes de l'évolution cognitive. La différence essentielle est que nous habitons cette dernière clôture, et que nous ne sommes donc pas capables d'en concevoir les limites (il faudrait pour cela pouvoir en sortir).

Le texte propose tout d'abord de préciser la distinction entre clôture des opérations et clôture du pour-soi, puis examine successivement les étapes de transformations de ces deux types de clôture dans les grandes étapes de l'évolution :

- passage d'une clôture du pour-soi de la monade (cellule biologique), à une clôture sensori-motrice faisant émerger des invariants sensori-moteurs stables.
- passage de la clôture sensori-motrice à une clôture sociale de significations imposées, faisant émerger une conscience d'appartenance à un groupe social,

- passage d'une clôture des significations imaginaires sociales imposées, à la construction d'un moi social individuel.
- passage d'une clôture sensori-motrice liée au corps à une clôture sensori-motrice technologique, typique de la modernité.

Au cours de ce chemin, nous obtenons des indications précieuses sur les questions des fétiches, des faitiches et des divinités, c'est à dire sur nos modèles.

Une double vision de la clôture

Le concept de « clôture opérationnelle » mérite une grande attention. Rappelons en effet que la thèse de Varela est qu'un système est autonome si et seulement si il est opérationnellement clos (Varela 89). L'objectif est de définir l'autonomie par une caractérisation de ses dynamiques fondamentales. La définition de la clôture opérationnelle peut en effet se résumer de la manière suivante : *un système est opérationnellement clos s'il se définit par des dynamiques de création et destruction de processus bouclés les uns sur les autres, et orientées vers le maintien d'une unité.* Cette vision a souvent été assimilée à une clôture physique, impliquant une absence d'interactions avec l'extérieur, comme nous l'avons signalé dans l'introduction. Il est clair qu'il s'agit là d'une interprétation erronée, car tous les textes de Varela évoquent longuement les interactions du système avec l'extérieur. Nous proposons ici une explication possible de cette incompréhension : l'idée de clôture opérationnelle (de même que l'autopoïèse) contient deux visions très différentes de l'idée de clôture sous le même concept. La première est la clôture des opérations, et la seconde la clôture du monde « pour-soi ».

La clôture des opérations

La première condition de la clôture opérationnelle est la clôture des opérations du système. Il s'agit là d'une description à caractère biologique, ou même à un certain niveau d'abstraction, mathématique (certaines analogies ont été proposées avec les corps algébriquement clos (Castoriadis 97)). L'idée centrale est celle d'un réseau de processus bouclés les uns sur les autres, dont l'activité structure le système. Certains processus sont reliés à l'extérieur du système, ce qui assure les interactions avec l'extérieur.

Une telle description est entièrement compatible avec une approche scientifique traditionnelle, et de telles propriétés, moyennant sans doute un

travail pour les rendre un peu plus précises, peuvent aisément faire l'objet de consensus parmi une communauté d'observateurs scientifiques.

Dans cette première approche de la clôture opérationnelle en tant que clôture des opérations, nous ne voyons pas de réelle difficulté conceptuelle. Il nous semble que la difficulté ressentie par de nombreux observateurs vient d'une deuxième interprétation, toujours affleurante dans les textes de Varela, mais peut-être pas suffisamment explicite.

La clôture du « pour-soi »

En effet, à partir de cette clôture des opérations, un saut conceptuel est effectué en amenant à considérer un tel système d'une façon différente et un peu inhabituelle pour une approche scientifique. Il s'agit de passer du point de vue de la commande au point de vue de la clôture (Varela 89). L'utilisation du terme « point de vue » est pour nous caractéristique. On demande un changement de point de vue, de position épistémologique par rapport au système. Nous défendons ici la thèse suivante : dans cette autre versant de la clôture, Varela nous invite implicitement à nous mettre à la place du système, à entrer par la pensée dans son monde « pour-soi ».

En effet, Varela évoque notamment l'impossibilité pour le système de distinguer les perturbations d'origine externe des perturbations d'origine interne (ou compensation de perturbation externe). C'est ainsi que s'introduit l'autre idée de la clôture, beaucoup plus forte selon nous. En effet, comment un système tel qu'une amibe pourrait-il avoir l'équivalent d'un monde extérieur ? Il ne se produit que des modifications de son fonctionnement interne, dues pour certaines à les interactions avec le monde externe, mais il lui est impossible de les distinguer de celles qui viennent de perturbations internes. Il s'agit de la clôture de la monade initiale, qui est « sans porte ni fenêtres » selon l'expression de Leibniz.

Cette clôture peut être assimilée à un isolement, mais c'est un isolement « perçu ». Plus exactement, il est impossible à ce type de structure d'avoir une quelconque « sensation » ou conception d'un extérieur. Mais cela n'est pas du tout incompatible avec des interactions effectives, repérables dans la vision scientifique. Cette dualité explique selon nous les ambiguïtés d'interprétations que l'on a pu trouver.

L'accès à la clôture du monde pour-soi requiert donc cet acte extrêmement difficile : s'identifier au système, imaginer son monde. Comment se figurer le monde « pour-soi » d'une cellule biologique ? Il nous faut nous projeter dans un être sans conscience, sans désir, sans anticipations, sans monde

extérieur, sans mémoire au sens où nous l'entendons. Un tel acte est pourtant indispensable selon nous pour saisir l'autre versant de la clôture opérationnelle : la clôture du pour-soi.

La clôture du « pour-soi », qui n'est pas toujours explicite dans les textes de Varela, fournit le lien le plus manifeste avec le sens usuel de l'autonomie. Cette idée est en revanche extrêmement claire dans tous les textes de Castoriadis sur ces questions. En effet, dès que l'on peut attribuer au système une clôture du monde pour-soi, il y a création d'un monde nouveau, avec son organisation et ses règles propres. Cette organisation et ces règles sont liées au maintien de l'unité du système, et donc acquièrent une indépendance par rapport aux lois de la physique.

Prenons un exemple pour illustrer ce point : soit une amibe créant des pseudopodes lors de l'ingestion d'un protozoaire. Le développement des pseudopodes est causé par la concentration de substances spécifiques à ce protozoaire dans l'environnement immédiat de l'amibe. Comment imaginer le monde « pour-soi » de l'amibe dans ce contexte ? C'est évidemment presque impossible pour nous. Nous savons simplement que le protozoaire n'existe pas dans ce monde, car aucun organe ne permet à l'amibe d'avoir accès au protozoaire en tant que protozoaire, elle n'a accès qu'à un changement de ses dynamiques qui produisent les pseudopodes. Ce monde contient-il une trace des modifications des dynamiques (développement des pseudopodes puis ingestion) ? de quelle manière ? Le fait qu'il nous soit quasiment impossible d'en parler consacre l'autonomie radicale de ce monde par rapport au notre et à celui de la physique.

La question de l'équivalence des deux clôtures et lien avec l'autonomie

On peut déceler une équivalence implicite entre les deux types de clôtures dans certains textes de Varela (Varela 89). Cette équivalence ne nous semble pas aller de soi. En effet, comme nous l'avons signalé, on observe un changement d'attitude extrêmement important par rapport au système : le passage de l'observation scientifique à une identification au système ou une tentative d'établir indirectement quelques propriétés de son monde pour-soi, s'il est impossible d'y avoir accès, même par imagination.

Si l'on considère que l'autonomie se manifeste lorsqu'une clôture du « pour-soi » s'instaure, la question de l'autonomie devient : quelle type d'organisation minimale engendre une clôture du « pour-soi » ? Ou encore, quand un processus chimique ou physique devient-il une entité créatrice d'un monde autonome ?

La réponse implicite de Varela, telle que nous l'interprétons, est qu'une clôture du pour-soi est automatiquement créée par une clôture des opérations qui organise le maintien d'une unité. Cette deuxième condition (maintien d'une unité) est sans doute très importante pour donner une possibilité d'imaginer un monde pour-soi du système. Elle indique en effet une possibilité d'accorder une forme de valeur aux situations en fonction de leur caractère menaçant pour l'unité ou non.

Illustrons concrètement ce dernier argument. Il est possible de décrire une clôture des opérations dans des systèmes dissipatifs simples comme les tourbillons de Bénard par exemple, ou bien encore des systèmes chimiques clos comme une étoile. En revanche, il est plus difficile de soutenir que de telles structures visent à un maintien de leur unité, en favorisant certaines interactions par rapport à d'autres, comme la déformation de la membrane cellulaire de notre amibe le faisait. Il serait donc plus facilement acceptable de dire que l'amibe produit un monde « pour-soi », que ne le font un tourbillon de Bénard ou une étoile, dont la structure ne traduit pas de visée particulière vers des interactions plus propices à maintenir leur unité.

Mais il s'agit là d'une question qui nous paraît extrêmement délicate, et nous ne prétendons nullement y apporter une réponse définitive. Notre propos vise à identifier les ambiguïtés et à souligner la nécessité d'un acte d'identification au système et donc d'une vision de son monde de l'intérieur, pour comprendre le sens de la clôture d'un monde pour-soi. Il nous semble important de souligner cette vision, qui justifie de manière plus profonde le choix du terme clôture. Nous laissons ouverte la discussion de notre interprétation de la thèse de Varela, selon laquelle cette clôture du pour-soi apparaît dès qu'une clôture opérationnelle se met en place, ou la proposition de procédures alternatives, scientifiques, permettant de déterminer si une clôture du « pour-soi » s'est formée ou non. Une telle discussion dépasserait largement le cadre de ce texte.

Armé de cette double vision de la clôture, nous poursuivons maintenant notre étude sur la création ou rupture de clôture (des opérations ou du pour-soi) en considérant d'autres formes de mondes « pour-soi » que celui de la cellule biologique.

Passage de la clôture de l'instant à celle du « présent remémoré »

Le monde pour-soi de la cellule biologique n'a pas de temps (ni d'espace), il est prisonnier de l'instant. Le passage à des interactions sensori-motrices est

décrit par Varela comme l'instauration d'une nouvelle clôture des opérations, celle des actions / sensations qui intègrent le monde extérieur dans la boucle (Maturana et Varela 94). Comme telles, ces interactions ne changent pas fondamentalement le monde pour-soi des systèmes tant qu'elles ne sont pas accompagnées d'une capacité de mémoire et d'anticipation.

La création d'un monde extérieur par des opérations d'anticipations sur les stimulations proximales

En effet, cette capacité permet de lier les instants entre eux dans une durée correspondant à une anticipation ou une attente. Ainsi, le monde « pour-soi » obtenu se libère de certaines limites précédentes : il dépasse le cadre de l'instant, pour s'établir dans une durée. Cette durée est pour nous analogue à celle du présent remémoré de Edelman dans la conscience primaire (Edelman 92). Il exprime la création d'un temps autonome, limité à une anticipation ou une attente. Dans cette fenêtre de temps, l'avant et l'après peuvent avoir un sens. Il s'agit donc d'une rupture de la clôture de l'instant.

De plus, la capacité de mémoire et d'anticipation permet aux organismes de construire un monde de perceptions présentant une certaine stabilité par rapport à leurs mouvements, dans un espace différent de leur espace interne. Ce monde stable dépend évidemment des capacités perceptives de ces organismes, ainsi que de leurs capacités motrices. Il y a donc à ce niveau, création d'un monde projeté, construit à partir des stimulations proximales. En effet, en intégrant la perception actuelle dans une histoire des perceptions passées, permettant des anticipations, les différentes stimulations proximales fusionnent et se combinent pour constituer des « percepts » stables par rapport aux anticipations. Ainsi, anticipons nous les différents aspects qu'un objet prendra en fonction de l'angle selon lequel il est perçu et des mouvements en cours (Petitot 92). Cette fusion des stimulations proximales est un acte créateur qui crée un monde de situations (plus ou moins) stables et prévisibles. J. Proust (Proust 97) aborde cette question dans un cadre théorique très différent, mais propose des directions de travail qui rejoignent en grande partie notre propos. Chez l'animal pourvu d'un cortex primaire (et donc selon Edelman d'une conscience primaire), nous pouvons prévoir que la proie existe dans son monde lorsqu'il la guette ou la chasse (alors que le protozoaire ne pouvait pas exister pour l'amibe) ainsi qu'une forme de l'espace qui l'en sépare, même si le mode d'existence de cette proie et de cet espace nous resteront probablement toujours largement inaccessible.

Le fétiche et la nécessité d'introduire la distinction saillance / prégnance

Cette description déjà très rapide serait selon nous très incomplète si l'on y ajoute pas la distinction entre saillance et prégnance (Thom). Rappelons que les saillances correspondent à un ensemble de singularités (bords, coins, trous, changements d'intensité etc...) décelable de manière sensori-motrice (l'aspect moteur est crucial pour juger de la stabilité). La notion de prégnance est beaucoup plus diffuse. Une prégnance est généralement associée à une saillance, et peut se diffuser d'une saillance à une autre, comme par contagion. Ainsi, certaines saillances sont-elles génétiquement prégnantes sexuellement par exemple, et d'autres saillances neutres pourront-elles être investies d'une prégnance sexuelle si elles sont associées (dans le temps ou l'espace) aux saillances initialement sexuelles. Ainsi la cognition sensori-motrice est-elle fondée sur l'anticipation (à l'aide de la mémoire) des saillances, mais aussi des prégnances associées. La diffusion des prégnances à des saillances similaires ou liée à celle de la perception initiale, renforce les capacités d'anticipation.

Notre description est très proche de celle du présent remémoré de Edelman, la prégnance étant l'équivalent de la valeur des catégories perçues pour Edelman. Nous insistons ici plus que Edelman sur la propriété de diffusion de la prégnance (valeur) d'une situation à l'autre par similarité. D'autre part, l'émergence de formes stables par anticipation des transformations des saillances est également étranger à Edelman. Notre hypothèse est que les catégories de Edelman correspondent à ces formes stables.

Ce fonctionnement cognitif est typique d'un grand nombre d'animaux (ceux possédant un cortex primaire selon Edelman). Il nous semble également que l'homme garde en partie ce fonctionnement (auquel est superposé un autre, que nous décrirons au paragraphe suivant). Ainsi, un grand nombre de comportements cognitifs humains peut-il s'expliquer dans le cadre de ce schéma cognitif. Nous prendrons comme exemple la création de fétiches. Un fétiche est un objet particulier (une saillance) auquel est associée une prégnance forte (positive ou négative), qui diffuse sa prégnance autour de cet objet. Pratiquer de telles attributions (positives ou négatives) est une tendance forte de la cognition (nous la plaçons au niveau sensori-moteur), indépendante des cultures. Il suffit de donner l'exemple de Bjorn Borg, champion de tennis de la fin des années 70, qui portait la même tenue pour tous ses matches de Wimbledon. Nous ne considérons ici que des objets fétiches privés, attribués par une seule personne. Le mécanisme de désignation des objets fétiches publics est différent, nous le verrons dans la suite.

Cette cognition fondée sur l'anticipation de saillances et de prégnances, fait émerger un espace (celui dans lequel se stabilisent les anticipations) et un temps (celui de l'anticipation). Le monde pour-soi associé est clos, car il est impossible de sortir de la clôture de la fenêtre de temps et de l'espace engendré dans cette période. Cette nouvelle clôture rompt la précédente en en dépassant les limites, et en l'englobant (sans la détruire totalement, puisqu'à certains niveaux la précédente demeure). Nous voyons donc que la distinction entre clôture des opérations et clôture du monde pour-soi permet de faire apparaître à la fois une rupture et une création de clôture, ce qui n'apparaissait pas sans cette distinction.

Passage de la clôture sensori-motrice à la clôture des significations sociales imposées

Nous proposons maintenant de nous intéresser à une nouvelle rupture, celle occasionnée par l'émergence des sociétés humaines. Nous entrons ici dans un domaine où l'identification, et l'imagination du monde « pour-soi » deviennent beaucoup plus naturelle, même si elles conservent de sérieuses limites.

Empathie : modèle de l'autre ou l'autre comme modèle ?

Selon Varela et Maturana, la création d'un nouveau type de clôture (des opérations et du monde pour-soi) est due à un nouveau type d'interactions : la communication par le langage (Maturana et Varela 94). Nous retrouvons un raisonnement similaire à celui de la clôture des opérations sensori-motrices, qui fait émerger un monde nouveau, à partir d'interactions nouvelles. De même, Edelman associe la création de la conscience d'ordre supérieur, spécifiquement humaine, au langage (Edelman 92).

Sans remettre totalement en cause ce rôle spécifique du langage, nous souhaitons le compléter. Plus fondamentalement encore que le langage, il nous semble en effet que la capacité l'empathie est à la base des interactions sociales humaines. Nous nous inspirons en cela les travaux de J.P. Dupuy sur ceux de Adam Smith ou de R. Girard (Dupuy 92) comme nous le proposons dans (Deffuant 96). L'empathie est en effet la capacité à s'identifier à quelqu'un d'autre, même partiellement, c'est à dire imaginer son monde, ses désirs, ses craintes, etc...L'empathie ne peut apparaître que chez les animaux dont la conscience primaire s'est développée, et qui présentent une vie collective structurée (sociale). Pour de tels animaux, les interactions entre les membres du groupe prennent une importance de plus

en plus considérable par rapport aux autres interactions. Les anticipations les plus utiles dans un tel groupe sont donc les anticipations du comportement des congénères. On imagine aisément l'avantage que procure la capacité d'empathie pour anticiper ces réactions. En effet, se mettre à la place de l'autre, et imaginer ses désirs et ses craintes, permet beaucoup plus justement de prévoir ses actions.

Nous rencontrons ici pour la première fois le modèle, dans son sens le plus riche. En effet, comme le fait remarquer J.P. Dupuy (Dupuy 92), le terme modèle a deux sens qui peuvent vus comme contradictoires. Prendre une autre personne comme modèle, c'est en général vouloir l'imiter, alors que les modèles scientifiques cherchent à imiter la nature. Ici nous nous concentrons sur le premier sens de modèles. Nous considérons que prendre un modèle humain ne peut se faire sans une part d'empathie. En effet, vouloir être l'autre ne se fait (chez l'homme) sans imaginer le monde « pour-soi » de l'autre, c'est à dire en imaginant la vie en étant à sa place. Prendre quelqu'un comme modèle impose donc de s'être fait un modèle (dans le sens scientifique) de son monde. Les deux sens du mot se retrouvent ici. Le mimétisme Girardien lui même exige à la fois de prendre l'autre comme modèle et de s'en faire un modèle. En effet, ce mimétisme porte sur le désir imaginé chez l'autre. Il suppose donc un modèle des désirs de l'autre, donc un modèle de son monde « pour-soi ».

Empathie et langage

Il semble que parmi les animaux, en dehors des hommes, seules certaines espèces de singes soient capables d'empathie. Mais on peut douter de l'importance et de la fréquence de l'utilisation de cette faculté dans le type d'organisation sociale de ces animaux. Il ne semble pas qu'elle soit fondamentale.

En revanche, l'empathie ne devient véritablement fondamentale lorsque l'évolution permet le langage, qui lui est intimement lié. En effet, le langage n'est possible que si la capacité d'empathie est présente, car il est impossible de parler à quelqu'un sans s'imaginer ce qu'il est capable de comprendre. Il s'agit d'une évidence telle que personne ne semble y faire attention. Elle explique cependant la manière dont on s'adresse aux très jeunes enfants et aux animaux (très différente de celle dont on s'adresse aux adultes de même niveau culturel). La difficulté des dialogues avec des locuteurs d'une langue inconnue est une autre illustration. Ces exemples

témoignent d'une anticipation constante de l'influence des mots sur le point de vue de l'autre. Un homme qui parle à une borne kilométrique ou à un arbre parle en fait à un être imaginaire dont il simule l'empathie et pour lequel il simule une empathie. L'utilisation d'un mot ou d'une expression suppose l'anticipation de son effet sur divers interlocuteurs potentiels. Cette anticipation suppose une forme d'empathie.

A l'inverse, la capacité d'empathie est multipliée par l'utilisation du langage. En effet, le langage permet de donner des indications beaucoup plus complètes et précises sur le point de vue de chacun, et donc de s'imaginer par la pensée regardant le monde d'un point de vue différent du sien propre. Il donne la possibilité d'agir directement sur le point de vue de l'autre, d'une manière beaucoup plus large et subtile. En résumé, l'utilisation du langage fait tendre l'essentiel des relations sociales vers des relations empathiques (à des degrés divers).

Nous voyons bien l'avantage de considérer l'empathie dans le cadre d'une exploration du monde « pour-soi ». En effet, la clôture sensori-motrice rend les organismes prisonniers de leur point de vue dans une fenêtre d'espace et de temps, ils sont dans l'impossibilité d'en sortir. L'empathie donne au système la capacité d'accéder (fictivement) à un point de vue différent du sien. Capacité certes limitée, et sujette à bien des erreurs et approximations.

Mais il est difficile d'aborder les répercussions de cette faculté d'empathie sur le monde pour-soi, sans aborder ses conséquences dans l'organisation globale de la société. En effet, il est clair que l'organisation de cette société et le monde « pour-soi » de ses membres sont intimement liés.

La société hétéronome comme attracteur stable de la clôture par empathie

Nous avons développé ailleurs les éléments nouveaux sur la compréhension du social, et notamment de l'organisation religieuse des sociétés archaïques ou traditionnelles, liés à l'empathie vue comme base des interactions sociales (Deffuant 98). Nous nous inspirons à nouveau des travaux de Dupuy (Dupuy 92) et de sa vision de ceux de R. Girard (Girard 81, 82). Nous tentons d'en résumer grossièrement les arguments principaux. Ils nous permettent d'aborder la définition des sociétés hétéronomes selon Castoriadis.

Nous proposons donc maintenant considérer deux clôtures qui interagissent : la clôture des opérations empathiques au niveau de la société entière, qui peut produire un monde « pour-soi » au niveau global, et la clôture des interactions empathiques d'un individu, qui influence le monde « pour-soi »

de cet individu. La clôture des interactions empathiques produit en effet un nouveau système dont le monde du pour-soi, en tant que système global, est difficilement accessible. On peut supposer qu'il s'apparente à son échelle, dans ses grandes propriétés, à celui de la cellule élémentaire. Mais nous ne chercherons pas trop à entrer dans cette question très difficile. Toujours est-il que l'on peut considérer que ces interactions empathiques collectives satisfont à la clôture opérationnelle et donc produisent une unité autonome au sens de Varela. Ces dynamiques ont une cohérence, une créativité dans leurs comportements propres qui échappent aux individus. Aucun individu n'est véritablement responsable de la tradition qui s'est instituée, elle se constitue au niveau de cette cohérence des dynamiques collectives. S'il est difficile d'accéder au monde pour-soi de cette clôture des opérations, nous pouvons en caractériser les formes générales des dynamiques.

En effet les interactions empathiques généralisées à tous individus deux à deux et dans toutes les combinaisons produisent une société fondamentalement instable, qui change constamment d'idoles, développe la violence et se vit comme une situation de crise, comme l'a identifié R. Girard (Girard 81, 82). Cette situation qu'il appelle « crise sacrificielle » est selon nous beaucoup plus compréhensible à l'aide d'interactions empathiques, que par les interactions mimétiques que considère Girard. Nous avons amorcé une analyse des différences entre ces conceptions dans (Deffuant 96). Nous nous bornerons à suggérer que l'organisation des flux empathiques est stabilisée par l'introduction de personnages externes à la société, hors de portée, qui polarisent ces flux. En effet, on peut montrer que les interactions empathiques entre deux personnes sont intrinsèquement instables. L'un des deux doit dominer et absorber l'autre, mais la situation peut se retourner à tout moment. Le seul moyen de stabiliser cette situation est d'introduire un intercesseur (médiateur) qui permet aux deux parties de se voir en interaction. Les divinités jouent ce rôle d'intercesseur en polarisant l'ensemble des flux empathiques. On voit immédiatement quelle garantie de stabilité ces personnages extérieurs amènent : la société étant conçue de l'extérieur, aucun de ses membres ne peut la changer. Cet état attracteur est donc extrêmement stable.

On peut se demander si l'utilisation du langage ne requiert pas un flux d'empathie vers une sorte de point fixe, garanti ici par les divinités, qui permettent l'existence des mots. La crise sacrificielle en effet, telle qu'elle est décrite par Girard, pousse la détérioration de la société jusqu'à une détérioration des mots et du langage qui deviennent inutilisables. L'empathie, ainsi que ses relations avec le langage donnent des pistes

intéressantes à explorer à ce propos. Cette analyse mérite d'être poursuivie, mais elle dépasse les limites de ce texte.

Tentons maintenant de pénétrer le monde pour-soi des membres d'une telle société. L'adoption du point de vue divin (ou d'un point de vue quelconque, auquel il est possible de s'identifier, tout en étant extérieur à la société, comme typiquement, le bouc émissaire) permet à chaque membre de se voir exister, comme membre du groupe, dans le regard des divinités (ou de ce personnage externe à la société auquel on s'identifie). En effet, la divinité voit la société comme un ensemble, sans bien distinguer ses membres, puisqu'elle voit cette société de l'extérieur. Le monde pour-soi est donc vécu comme appartenance (voire identification) à un groupe. Comme nous l'avons signalé, il y a là rupture de la clôture sensori-motrice auto-centrée, car d'autres mondes pour-soi deviennent accessibles (fictivement). C'est ainsi qu'il devient possible de se concevoir comme membre d'un groupe, dans un temps indépendant, le temps de la divinité. On comprend ainsi pourquoi de nombreuses cultures rendent les dieux responsables de la création du temps.

Les comportements deviennent orientés par ce nouveau point de vue externe, qui peut facilement aller à l'encontre de celui de la clôture sensori-motrice (comportements de jeûne, sacrifice, automutilation, tabous sexuels...). Cette création d'un imaginaire collectif partagé par empathie permet donc aux membres de la communauté de ne plus être soumis uniquement à la clôture de la conscience primaire. En cela, ils sortent de cette clôture, pour construire une conscience d'ordre supérieur correspondant à leur appartenance à une communauté. Cette conscience d'ordre supérieur est à notre avis un peu différente de celle qu'Edelman décrit (Edelman 92), car cette dernière correspond à celle des individus modernes. Nous sommes en accord avec Louis Dumont, pour lequel l'individu social tel que nous l'entendons est une invention moderne. Les sociétés exhibant le type de dynamiques d'interactions sociales que nous décrivons sont les sociétés traditionnelles (de type indo européen, ou les empires mésopotamien et égyptien par exemple). Nous avons indiqué ailleurs (Deffuant 98) les différences (importantes) que présentent les sociétés judéo-chrétiennes ou bien les sociétés sans état telles que les décrit Pierre Clastres (Clastres 74). Ces différences ne changent pas fondamentalement la suite de l'argumentation.

Dans notre perspective, la rupture liée à une capacité d'empathie généralisée est primordiale. Le fait que les divinités sont intentionnelles, douées d'une capacité d'empathie, est crucial puisqu'il est un élément essentiel du monde

pour-soi de ce type de culture. En effet, l'empathie polarisée sur un être externe au groupe fait émerger un monde pour soi dominé constamment par l'image du groupe dans sa totalité, reflet du regard de l'être externe. Dans ce cadre, l'assimilation du Dieu au fétiche, que B. Latour semble pratiquer (Latour), est un contresens énorme. En effet, même si la prégnance du fétiche est associée par combinaison d'empathie (fétiche public en opposition à un fétiche privé), il reste un objet, et ne peut avoir le même pouvoir structurant sur la société que la divinité.

Ces sociétés, dont la stabilité est acquise au par un attracteur d'empathie perçu comme externe, sont considérées comme hétéronomes par Castoriadis car leurs valeurs et leur organisation sont vécues comme imposées de l'extérieur, et impossible à remettre en question. Il est impossible de les contester, ou même de penser en dehors d'elles. L'individu est prisonnier d'une clôture de significations qu'il vit comme imposée de l'extérieur, sur laquelle il n'a aucune prise. En effet, l'existence sociale dans ces sociétés est conditionnée par le passage obligé des divinités et de la tradition qui est commune à tous. Des êtres qui ne partagent pas cette tradition sont inaccessibles à l'empathie (ils ne sont pas considérés comme de véritables humains). L'organisation des flux empathiques des sociétés archaïques et traditionnelles permet donc de créer de nouveaux mondes collectifs, mais également de nouvelles clôtures dans lesquelles les membres de ces communautés sont enfermés. C'est cette clôture, qui est construite lors d'un processus collectif lent, que Castoriadis considère comme une hétéronomie pour les membres de la communauté. Cependant, cette clôture est aussi une rupture de la clôture sensori-motrice qu'elle englobe dans un monde plus large.

Cette hétéronomie est donc relative à la clôture suivante que nous allons aborder maintenant, mais il s'agit d'une autonomie par rapport à la clôture précédente (sensori-motrice).

Passage de la clôture par empathie aux dieux à la clôture par empathie publique : le miracle grec

L'évolution menant au projet de l'autonomie sociale a lieu en Grèce, au du 7^{ème} au 5^{ème} siècle avant J.C.. On ne pourra évidemment le considérer ici que de manière très superficielle. Nous espérons cependant pouvoir indiquer comment les contradictions apparentes entre Varela et Castoriadis peuvent s'aplanir et leurs deux visions prendre place dans une plus large cohérence.

L'œuvre de Castoriadis est imprégnée en profondeur par le projet de la polis Grecque. Le miracle Grec, qui construit une société entre citoyens égaux, soumis à la même loi, constitue une rupture par rapport à l'ordre archaïque ou traditionnel. En effet, la construction de la société est vue comme le résultat de la confrontation des points de vue, qui sont a priori d'une égale valeur (à Athènes après la réforme Clithénienne). La loi qui en résulte s'applique à tous, et n'est pas vue comme l'expression de la volonté d'un personnage particulier, s'imposant aux autres. Cette loi est considérée explicitement comme une construction, à laquelle chacun est invité à participer. La différence par rapport aux autres sociétés est que l'organisation est conçue comme résultant de l'action de ses membres et non comme donnée de l'extérieur.

Cette transformation s'accompagne d'une plus grande capacité d'ouverture aux autres sociétés, qui tout en étant déclarées barbares, deviennent accessibles par l'empathie. On trouve ce souci d'impartialité déjà chez Homère dans l'Iliade, comme le fait remarquer Castoriadis, en citant Hannah Arendt (Castoriadis 86a). Il s'agit d'une attitude totalement nouvelle, qui est impossible dans le cadre des clôtures de significations sociales imposées décrites précédemment.

Comment interpréter ces faits en termes de clôture (des opérations ou du « pour-soi ») ? Considérons tout d'abord la clôture des interactions entre les citoyens, au niveau collectif. On constate un changement dans la dynamique des interactions empathiques, qui s'organisent sans polarisation stable (le rôle des divinités dans la vie publique, tout en étant important, n'est plus le seul élément organisateur des flux empathiques). Les conditions spécifiques du monde Grec, ont en effet permis d'élaborer un évitement de la « crise sacrificielle » différent de celui des autres sociétés.

Considérons la clôture du « pour-soi » des membres de la communauté, qui nous est a priori plus accessible. Le premier point très important est que les grecs ont réussi les premiers à réaliser une empathie pour d'autres sociétés que la leur, c'est à dire à prendre en compte le point de vue d'autres significations imaginaires sociales que les leurs, au lieu de les refuser et de les considérer uniquement de l'extérieur. Ils ont donc les premiers réussi à effectuer au niveau collectif, l'opération correspondant à la faculté d'empathie au niveau individuel. Leur organisation n'est donc plus la seule possible, seule envisageable. Elle fait partie d'un ensemble de possibles, plus ou moins souhaitables. En ce sens, il s'agit d'une rupture de la clôture des significations sociales imposées, et donc un élargissement sans

précédent du monde pour-soi des citoyens. C'est principalement sur cette rupture que Castoriadis insiste.

Un tel exploit vient de l'activité de débat contradictoire et de l'organisation de la compétitions publiques entre égaux qui généralisent la situation où chacun est observé par tous les autres, au moins potentiellement. Au lieu de prendre appui sur un personnage extérieur imaginaire regardant la communauté, il y a eu création d'un point de vue public global, directement accessible à chacun par l'expérience des actions publiques. Cette organisation peut donc rester stable car l'empathie pour l'ensemble de la communauté donne une sorte d'attracteur, qui est la communauté tout entière (le public). Cet attracteur est interne et non externe, mais il acquiert une sorte de transcendance, en allant au delà d'une volonté particulière. Il évite la situation de violence généralisée par empathie.

Dans le monde du pour-soi de chaque individu, il s'agit d'une transformation radicale. En effet, cette nouvelle dynamique donne la possibilité à chacun d'exister socialement aux yeux de la communauté, en tant qu'individu, alors que les sociétés traditionnelles ne lui donnent la possibilité de se voir que comme membre d'un groupe. Chacun est invité à jouer un rôle dans les compétitions ou les débats, et peut donc s'identifier à celui qui est au centre de la scène. De même, chacun se voit également comme membre du groupe (en étant dans le public), ou en ayant une empathie pour le public s'il est sur la scène. C'est ainsi qu'apparaît le premier individu social, qui se crée un espace d'existence et d'action au sein du groupe. Un nouveau monde pour soi est donc instauré, dont l'espace est créé par l'empathie pour le regard du groupe sur soi. Ce n'est donc pas seulement l'autonomie individuelle sociale, mais bien l'existence même de l'individu social qui est instaurée.

Le projet d'autonomie sociale de Castoriadis, à la fois individuelle et collective, s'interprète donc dans notre perspective par un double mouvement :

- L'individu se trouve dans une société qui organise l'existence d'une empathie publique sur chacun, lui permettant de construire son être social comme reflet de cette empathie. Son monde pour-soi change donc radicalement par rapport aux sociétés précédentes, il acquiert une composante supplémentaire : celle de l'individu social.
- L'individu social acquiert une certaine capacité de se concevoir socialement en dehors de son groupe. Il peut ainsi parvenir à une empathie pour les membres d'organisations sociales autres que la sienne

(notamment hypothétiques) et à en rompre la clôture pour en construire une nouvelle, issue du débat contradictoire sur les organisations possibles.

Comme nous l'avons vu jusqu'à présent, toute rupture de la clôture du pour-soi se fait par la constitution d'une autre clôture du pour-soi, créant un monde plus large, dans lequel la clôture précédente apparaît comme une possibilité parmi d'autres, intégrées dans une cohérence plus large. C'est encore vrai partiellement pour la dernière rupture de clôture que nous venons d'indiquer. En effet, la mise en débat de l'organisation sociale, et la capacité d'en imaginer d'autres représentent cette cohérence plus large. Cependant, ce changement est lié à une transformation des flux empathiques qui changent l'expérience de la vie en communauté. Cette nouvelle clôture du pour-soi fondant l'autonomie de l'individu social est-elle d'une nature différente de celles que nous avons considérées précédemment ? Peut-on la considérer comme une rupture indéfinie comme Castoriadis semble le faire ? S'agissant à nouveau du monde pour-soi, cette question est très délicate. En effet, nous habitons un monde de ce type (et même sans doute un peu plus pauvre), nous sommes donc à l'intérieur des limites d'une telle clôture. Ses limites ne pourraient apparaître que si nous réussissions à la briser, pour en constituer une nouvelle qui l'engloberait. Par définition, la clôture du pour-soi dans laquelle on vit n'apparaît pas comme une clôture.

Nous pouvons maintenant confronter ce projet d'autonomie proposé par Castoriadis à la situation de l'individu moderne, à propos de laquelle ce dernier ne ménage pas ses critiques.

La modernité ou le passage de la clôture sensori-motrice à la clôture des sciences expérimentales

Castoriadis, bien que reconnaissant la filiation entre la société moderne et le projet d'autonomie Grec, ne cache pas un certain malaise, voire un certain mépris pour cette société. Nous proposons d'interpréter cette attitude dans le cadre de notre questionnement entre rupture et création de clôture. Au passage, nous glanerons des indications sur la question des modèles.

La principale innovation de la modernité par rapport à la civilisation Grecque est celle des sciences expérimentales à la renaissance. Le développement formidable de ces sciences et leurs succès immenses ont transformé les sociétés humaines et les modes de vie comme jamais il ne fut même imaginable auparavant. Ces succès ont conduit l'idéologie moderne à être largement dominée par les relations aux choses, aux objets, à l'économie, alors que la cité Grecque vivait pour l'interaction entre égaux,

l'action publique, politique. L'individu moderne est expert en relations avec les choses (surtout dans leur fabrication), alors que le citoyen grec était expert en relations avec ses semblables. Ce sont là des lieux communs.

Ces lieux communs trouvent selon nous un intérêt renouvelé dans notre perspective. Notre proposition est la suivante : les sciences expérimentales ont permis la rupture de la clôture sensori-motrice, la remplaçant par une nouvelle clôture utilisant des prolongements technologiques du corps. L'exemple typique est donné par Galilée qui le premier introduit une boucle de ce type pour l'utilisation de la lunette astronomique. A l'aide de ces outils techniques, qui permettent de nouvelles boucles de type « perception-action », les modernes se sont ouvert un champ cognitif nouveau, immense. La raison, inventée par les Grecs dans le cadre du débat politique, a fuit la cité et trouve sa pleine efficacité dans la conception de ces boucles sensori-motrices technologiques, et leur stabilisation sous forme de modèles mathématiques. Elle mène au délire moderne tant de fois dénoncé par Castoriadis de pseudo-maîtrise pseudo-rationnelle de la nature. A nouveau, nous sommes tentés de dire qu'il ne s'agit pas d'une nouvelle clôture, car nous ne voyons que la rupture de la clôture sensori-motrice du corps qui ouvre sur une exploration infinie toujours renouvelée par les progrès techniques. Cependant, la prudence est à nouveau de mise, car il s'agit de la clôture que nous habitons, et donc dont les limites ne nous sont pas perceptibles tant que nous ne l'avons pas dépassée par une autre.

B. Latour nomme fétiches les inventions des sciences modernes, s'inspirant du terme fétiche (Latour 96). Il justifie ce choix par l'acte de croyance toujours nécessaire pour donner une existence aux objets scientifiques qui dépasse le contexte de leur création. Dans notre perspective, cet acte de croyance est à assimiler à celui de la perception d'une saillance. En effet la création des saillances de notre espace perceptif usuel fonctionne selon les mêmes principes : il s'agit de repérer des stabilisations de boucles « perception-action ». L'assimilation au fétiche nous paraît injustifiée car le fétiche se distingue par un acte de croyance qui porte sur la prégnance de l'objet, et non sur sa saillance, qui elle va de soi. Or ce qui caractérise les objets découverts par les sciences est leur prégnance neutre (un photon porte-t-il bonheur ? un neutrino a-t-il des pouvoirs bénéfiques ou maléfiques ?). Un acte de foi sur des prégnances, analogue à celui qu'on constate par rapport au fétiches est exclu de la démarche scientifique, tout simplement car les prégnances « bon » ou « mauvais » sont exclues du champ scientifique.

Un des effets de cette nouvelle clôture technologique est qu'elle produit un monde qui échappe en partie à celui des interactions sensori-motrices usuelles, du corps. Ce monde techno-scientifique (celui de la physique et de la biologie, de l'infiniment petit et de l'infiniment grand) se présente cependant avec une légitimité beaucoup plus forte que le monde sensori-moteur biologique, pour-soi, auquel les sciences ont bien du mal à avoir accès. On aboutit à ce paradoxe insensé que le monde de la vie de tous les jours, des « sachverhalte », qui est la base sur laquelle tout se fonde, devient étranger à l'homme moderne. Varela (Varela et al. 91) et Petitot (Petitot 92) portent le projet de réintégrer ce monde dans l'idéologie moderne, c'est à dire de l'exprimer par une démarche scientifique renouvelée.

L'autre versant, plus positif, est l'intégration dans la vie de tous les jours d'une instrumentation qui rompt la clôture physique du pour-soi en l'élargissant considérablement. En effet, si certaines explorations des sciences expérimentales mènent à des mondes étrangers à celui du corps, d'autres le transforment en lui offrant des prolongements technologiques. Les nouveaux moyens de locomotion, de transmission changent radicalement le monde sensori-moteur du pour-soi en l'élargissant. Ils produisent un homme nouveau, habitant un monde sensori-moteur technologique nouveau. Les dernières technologies des « réalités virtuelles » en sont l'expression la plus aboutie. Cette élargissement de la clôture correspond à un véritable élargissement de son autonomie.

Cependant, le monde social « pour-soi » de l'individu moderne reste en dehors de ses discours dominants. En effet, la modernité, en se dotant d'une idéologie dans laquelle l'approche scientifique réifiante est dominatrice, se prive de la possibilité d'intégrer véritablement la capacité d'empathie, qui est pourtant au centre des comportements sociaux humains. Cette capacité, prise dans toute sa complexité, permettant des fusions, des influences, des réflexions croisées indéfiniment entre les individus échappe donc à l'image que l'homme moderne a de lui-même, dans les discours dominants. En effet, le raisonnement fondé sur l'empathie proposé par Weber par exemple, même s'il constitue un premier pas dans la voie que nous proposons, reste extrêmement partiel, en ne s'autorisant qu'une empathie pour la partie rationnelle de l'individu. Castoriadis a montré les limites de cette approche, qui n'admet dans l'empathie que ce qui est spécifique de la clôture moderne du pour-soi (Castoriadis 90). Il est donc extrêmement difficile à l'individu moderne de comprendre la clôture empathique qui forme malgré lui son individualité sociale, par là même en grande partie atrophiée.

Encore plus que la société Grecque, la société moderne vit constamment au bord de la crise sacrificielle, dans chaque homme est en compétition avec chaque autre, et qui bascule dans une violence généralisée des relations de concurrence deux à deux. Comment cette société fait-elle pour tenir, dans une situation qui est synonyme de chaos et de destruction généralisée pour toutes les autres ? Il est clair que cette question outrepassé les limites de ce texte. Suggérons cependant que cette atténuation de son individualité sociale est liée à la capacité de la modernité à côtoyer constamment la crise sacrificielle. En effet, l'individualisme moderne, qui est le contraire de l'individu social Grec, est curieusement une manière de potion anesthésiante permettant de détourner le regard de la compétition sauvage, la rivalité généralisée qui font rage dans la société moderne. L'individualisme moderne permet de faire semblant de ne pas être en compétition avec les autres, puisque la seule soi-disant référence est soi-même. On cherche à réussir pour se réaliser, s'accomplir soi-même. Le fait qu'une fable aussi grossière puisse faire l'objet de consensus aussi unanime est le signe d'un aveuglement bénéfique, de ceux qui permettent à la société de tenir. Gardons nous donc de le dénoncer trop hâtivement.

Notre perspective fait donc apparaître un curieux retournement : selon J.P. Vernant, les sciences sont en Grèce fille de la cité, c'est à dire que la conception des phénomènes naturels est issue de la révolution de l'organisation sociale. Les sciences ont été profondément transformées par l'approche expérimentale moderne qui les a focalisées sur la nature et sa maîtrise (pseudo rationnelle). Elles reviennent ensuite vers la cité en étant totalement détachées. Notre analyse indique qu'elles ont perdu dans ce mouvement l'essentiel de leur adéquation aux questions de la cité, et que leur influence sur elle a de grande chance d'être désastreuse, comme l'histoire du 20^{ème} siècle l'a amplement démontré.

Conclusion : Les dieux ne sont pas des fétiches, ni des faitiches

La distinction entre clôture des opérations et clôture du pour-soi permet d'articuler l'autonomie selon Varela et selon Castoriadis. A la lumière de notre parcours, l'autonomie apparaît comme un mouvement graduel, contribuant à la rupture partielle d'une clôture du pour-soi afin d'en mettre une autre plus large en place. Ce discours se fonde donc sur une capacité à accéder au moins indirectement à cette clôture du monde pour soi du système considéré, car les clôtures des opérations ne fournissent qu'une

vision partielle. Les contradictions entre rupture ou création de la clôture viennent selon nous du fait que Varela examine le processus plutôt du point de vue de la clôture des opérations (bien que la clôture du pour-soi soit bien souvent implicite chez lui), alors que Castoriadis prend en considération la clôture du « pour-soi », qui fait apparaître clairement les ruptures. Enfin, Castoriadis s'intéresse en premier lieu aux dernières ruptures du monde pour-soi qui ont eu lieu chez les humains (celle de l'empathie publique et celle des sciences expérimentales). Nous habitons les nouvelles clôtures mises en place, et par définition, nous ne pouvons en concevoir les limites.

Dans notre cadre, l'équivalence implicite de B. Latour entre Dieux et faitiches, ou fétiches est donc extrêmement critiquable. En effet, nous avons insisté dans notre présentation sur la rupture très importante, et bien identifiée par I. Stengers, advenant dès que la cognition prend pour objet des sujets. En effet, le monde n'est plus un, mais devient multiple, car il faut prendre en compte le monde des sujets. Si ces sujets disposent de la capacité d'empathie, ils prennent eux-mêmes en compte le monde des autres. Nous avons tenté de montrer qu'il y a là rupture de la clôture sensori-motrice et changement radical de monde « pour-soi ».

Or les fétiches et les faitiches sont des objets (ils ne créent pas de monde « pour-soi »), alors que les Dieux sont des sujets intentionnels. Négliger cette différence c'est tomber sous le coup de la malédiction de la caricature de tolérance proposée par I. Stengers. En effet, l'effort proposé par I. Stengers dans les cosmopolitiques est fondé sur cette difficulté de la multiplicité des mondes, et de la reconnaissance du droit de chacun à vivre pleinement. Or assimiler les Dieux à des fétiches, donc à des objets, revient à nier un monde, celui du Dieu. Car un fétiche ne crée pas de monde, alors que tout Dieu en crée un.

Dans notre perspective, Dieu et fétiches apparaissent comme issus de fonctionnements cognitifs extrêmement différents et leur assimilation nous semble source de contresens très regrettables. Les divinités engendrent selon nous un monde « pour-soi » qui est la source du fonctionnement de la société et de sa perception d'elle même. C'est ce monde qui permet à chaque membre de la communauté de se vivre comme partie d'un tout. C'est d'ailleurs ce qui explique pourquoi une maladie frappant un membre du groupe s'interprète naturellement comme un message envoyé à l'ensemble du groupe.

Partager la conviction que les électrons appartiennent au monde réel ne nous semble pas pouvoir engendrer des conséquences du même genre. En fait, nous avons argumenté sur le caractère atrophié de la cognition sociale

(fondée sur l'empathie) du monde moderne, qui a tenté de bannir cette capacité de tous ses discours dominants (sauf dans l'art du roman, mais ce n'est pas un discours dominant).

L'enjeu de l'après modernité que fait apparaître ce texte est celui de la conciliation entre la clôture cognitive engendrée par les sciences expérimentales et la réhabilitation d'une clôture empathique explicite dans un monde social permettant de concevoir véritablement les individus sociaux. Castoriadis a ironisé à juste titre sur le terme « post-modernité », qui sonne comme un aveu de vacuité. En effet ce terme indique une incapacité à s'affirmer de manière autonome, puisqu'il se définit comme l'après quelque chose.

Nous proposons de donner du contenu à ce terme en l'appelant « deuxième renaissance ». En effet, la première renaissance nous a apporté les sciences expérimentales, donc donné un essor formidable au monde, primaire, sensori-moteur. Le monde empathique, social, avec toutes les ressources de création d'êtres reflets d'empathies croisées et multipliées a été au contraire ignoré, refoulé (sauf en partie dans le roman). C'est cette partie avortée de la renaissance qu'il importe de mettre maintenant en œuvre, après les errements désastreux du 20^{ième} siècle dans ce domaine.

Bibliographie

Castoriadis C. « La polis Grecque et la création de la démocratie ». *Domaines de l'homme. Les carrefours du labyrinthe II*. Seuil 1986a

Castoriadis C. « La logique des Magmas et la question de l'autonomie ». *Domaines de l'homme. Les carrefours du labyrinthe II*. Seuil 1986b

Castoriadis C. « Individu, société, rationalité, histoire ». *Le monde morcelé. Les carrefour du labyrinthe III*. Seuil 1990.

Castoriadis C. « De la monade à l'autonomie ». *Fait et à faire*. Le seuil, 1997.

Clastres P. *La société contre l'état*. Editions de Minuit. 1974.

Deffuant G. « Mécanismes empathiques et complexité sociale : autour de J.P. Dupuy ». *Actes des journées de Rochebrune* 1996.

Deffuant G. « Les modèles cognitifs à l'épreuve du phénomène religieux : Proposition de directions de recherche nouvelles fondées sur l'empathie ». *A paraître dans Intellectica*.

Dumont L. *Essais sur l'individualisme*. Le Seuil. 1983.

- Dumouchel P. in *l'Auto-organisation : du physique au politique*. Seuil 1983.
- Dupuy J.P. *Introduction aux sciences sociales*. Ellipses. 1992.
- Edelman G. *Biologie de la conscience*. Odile Jacob. 1992.
- Girard. R. *La violence et le sacré*. Grasset. 1981.
- Girard R. *Le bouc émissaire*. Grasset. 1982.
- Latour B. *Petite réflexion sur le culte moderne des dieux Faitiches. Les empêcheurs de penser en rond*. Synthélabo. 1996
- Maturana H. et Varela F. *L'arbre de la connaissance*. Addison-Wesley. 1994.
- Petitot J. *Physique du sens*. Editions du CNRS. 1992.
- Proust J. *Comment l'esprit vient aux bêtes. Essai sur la représentation*. Gallimard. 1997.
- Stengers I. *Cosmopolitiques*. Tome 1. La découverte / les empêcheurs de penser en rond. 1997
- Thom R. *Esquisse d'une sémiophysique*.
- Varela F. *Autonomie et connaissance*. Seuil. 1989.
- Varela F. , Thompson E., Rosh E. *The embodied mind*. MIT Press. 1991.

Dépot légal : 1^{er} trimestre 1998
imprimé par l'Ecole Nationale Supérieure des Télécommunications à Paris
ISSN : 1242 - 5125

