

École des Hautes Études en Sciences Sociales

Thèse présentée pour obtenir le grade de
DOCTEUR DE L'ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
en
SCIENCES COGNITIVES

NÉCESSITÉ DES MODÈLES EN SCIENCES COGNITIVES :
DE LA MODÉLISATION À LA MISE EN PARANGON

Pierre SAUREL

Thèse soutenue le 16 janvier 1998 devant le jury composé de :

MM.

Michel	IMBERT	Président
Claude	LAMOUREUX	Rapporteur
John	STEWART	Rapporteur
Paul	BOURGINE	Directeur
Wolf	EBERWEIN	Examinateur

Sommaire

Une table des matières détaillée figure à la fin de la thèse

Introduction	9
Chapitre 1 L'évolution de la notion d'axiomatisation et ses relations avec l'étude du cerveau 1903-1996	29
1 Quarante ans d'évolution de la notion d'axiomatisation à travers le parcours de von Neumann (1903-1943)	31
2 Les années quarante et l'unification de la conception du formalisme (1943-1956)	49
3 L'intelligence artificielle et les sciences cognitives, quarante ans de confrontation à une ontologie formaliste (1956-1996)	75
Chapitre 2 Comment les sciences cognitives conçoivent l'étude des processus	87
Introduction du chapitre 2	89
1 Diversité et unité des sciences cognitives	91
2 Le connexionnisme, un cadre général pour modéliser les processus "émergents"	107
3 En guise de conclusion sur la façon d'étudier les processus cognitifs	131
Chapitre 3 La complexité : des théories de la quantité ?	141
1 Les modèles de la complexité/quantité, cas particuliers du cadre de Farmer	143
2 Questions et outils pour aborder la complexité/quantité	179
3 Trois méthodologies de la complexité/quantité	197
Chapitre 4 La pratique de la modélisation : la mobilisation de masse en RDA en 1989	249
1 Introduction à la pratique de la modélisation	251
2 La modélisation mise en pratique : la mobilisation de masse en RDA en 1989	265
3 Pour aller plus loin dans la pratique de la modélisation	301
Chapitre 5 La complexité associée à la cognition nécessite la matérialité : une thèse pour une complexité au-delà de la complexité/quantité	313
Introduction	315
1 La complexité au-delà de la quantité	319
2 Que nous impose cette nouvelle complexité ?	347
Table des matières	365

En hommage à Jean,

exemple de courage et d'abnégation, parangon de vertu,
qui a cherché à vivre en harmonie avec ses conceptions.
Il est mort pour faire vivre les idées qui nous animent.

Remerciements

Il est difficile d'exprimer toute la gratitude que j'ai envers Paul BOURGINE, qui m'a appris tout ce qui est présenté dans cette thèse, qui m'a initié à la recherche, et qui, malgré ses multiples et débordantes activités, a souvent été d'une disponibilité remarquable.

Je remercie Monsieur le professeur IMBERT qui me fait l'honneur de présider le jury. Mes remerciements vont également à Messieurs les professeurs LAMOUREUX et STEWART qui ont accepté la tâche ingrate d'être rapporteur de ce travail.

Je sais gré à Herr Professor Doktor Wolf-Dieter EBERWEIN d'avoir bien voulu participer à ce jury. Je le remercie également de m'avoir permis de travailler à ses côtés.

Je remercie les chercheurs du CRÉA pour leur collaboration à ce travail et plus particulièrement Jean PETITOT dont les discussions enrichissantes m'ont permis de mieux comprendre les Sciences Cognitives et m'ont appris que le projet de la modélisation en Sciences Humaines et Sociales était possible.

Je remercie Hugues BERSINI de m'avoir formé à l'apprentissage par renforcement et d'avoir été à l'initiative de ce travail.

Je remercie Vincent DOUZAL qui est à l'origine d'un nombre important des analyses développées dans ce travail.

Je remercie Calixte TAYORO de m'avoir ouvert les yeux et de m'avoir amené à une lecture plus attentive des travaux de Turing et de Wolfram.

Je remercie Jürgen SCHMIDHUBER qui m'a montré la force des positions symboliques.

Je remercie mes collègues successifs de la Division ÉLAN du CÉMAGREF d'Antony, de la Société CORTOSYS, du Laboratoire CRÉA de l'École Polytechnique, du lycée MICHELIS à Amiens et de l'IUT d'Évry de m'avoir accordé un

soutien chaleureux pendant la durée de ce travail.

Je remercie la Direction de la Prospective de la Société DASSAULT-AVIATION pour avoir financé une partie de ce travail.

Je remercie Jean-Pierre DUPUY de m'avoir accepté au sein du pôle complexité du CRÉA, à Palaiseau, et de m'avoir ainsi permis de terminer ce travail dans des conditions exceptionnelles.

Je remercie Marie-France HANSELER, Chantal LEMAIRE et Cyril SUQUET qui m'ont été d'une aide précieuse et efficace en toutes occasions.

Je remercie les membres du Collège de Polytechnique et de la Direction du Troisième Cycle, colocataires sympathiques.

Je n'aurais pas pu accéder aux sources bibliographiques sans la disponibilité du personnel de la Bibliothèque Centrale de l'École Polytechnique. Qu'ils en soient tous remerciés et plus particulièrement Michel MULTAN pour sa patience infinie.

Je remercie tous mes proches, famille et amis, qui m'ont manifesté leur soutien moral pendant la réalisation de ce travail.

Je remercie Patrice PREZ pour ses discussions passionnées, pour ses critiques acérées, pour l'acuité de son regard, pour ses coups de colère salvateurs. Je le remercie pour sa volonté inébranlable et sa détermination à aller de l'avant, à repousser les limites, quoi qu'il advienne et quelles qu'en soient les conséquences. Il m'a montré que chercher n'était pas seulement une profession mais aussi un engagement.

Je remercie André SAUREL, d'avoir lu les différentes versions de ce travail avec la minutie et la patience qui le caractérisent.

Je remercie également Ghislaine BARBERY, qui m'a accompagné, épaulé et soutenu sans faille pendant toute la durée de ce travail.

Je remercie enfin tous ceux qui ont hanté mes nuits et m'ont offert ainsi la possibilité de dialoguer avec le passé.

INTRODUCTION

Le renvoi au technique est une échappatoire.

Jean CAVAILLÈS

Sur la logique et la théorie de la science, 1947¹

*Le véritable théoricien épargne les formules, autant qu'il le peut ;
il exprime avec des mots ce qui se laisse dire en mots,
tandis que dans les livres des praticiens
les formules ne figurent trop souvent qu'à titre de simple ornement.*

Ludwig BOLTZMANN

Über die Bedeutung von Theorien, 1890²

¹Références précises [Cavaillès, 1976] [p. 71].

²Cité dans [Dugas, 1959][p. 122].

Par goût pour la pluridisciplinarité j'ai choisi de faire un DEA, puis une thèse, en sciences cognitives. La pluridisciplinarité est souvent conçue comme une simple juxtaposition de disciplines. On peut au contraire la considérer comme l'occasion de construire des connaissances par une imbrication étroite de savoirs issus de sources très différentes. La pluridisciplinarité serait alors l'intégration de connaissances provenant de diverses disciplines et leur dépassement dans une problématique nouvelle. Dans cette optique, les sciences cognitives étaient un espoir formidable. Elles formaient une science en constitution qui concernait de nombreuses disciplines académiques et avait comme objectif important une mise en commun et une confrontation étroite de résultats déjà obtenus par ailleurs. On pouvait envisager comme réaliste l'établissement d'une science à part entière qui briserait les carcans disciplinaires. Je voulais participer à une telle aventure.

Mon objectif personnel était de contribuer à la construction de modèles qui relèveraient de ces sciences. Mon souhait était d'appliquer des mathématiques à des exemples précis. J'étais guidé à la fois par le souci de rester très proche des problèmes concrets et par la volonté de contribuer à leur résolution en mettant en œuvre les outils abstraits issus des mathématiques.

Grâce à l'initiative d'Hugues Bersini, enseignant-chercheur à l'IRIDIA, laboratoire de l'université libre de Bruxelles (ULB), j'ai travaillé, pendant le DEA sur des problèmes d'apprentissage par renforcement ; il s'agissait d'étudier et de démontrer la convergence d'un algorithme élaboré par Watkins, le *Q-learning* [Watkins, 1989]. Les différents algorithmes d'apprentissage par renforcement avaient été conçus à l'origine à partir de considérations et d'expériences de psychologie proposées par Rescorla et Wagner [Rescorla et Wagner, 1972]. Ces travaux avaient fourni un support conceptuel à Barto et Sutton [Barto *et al.*, 1983], qui avaient proposé des algorithmes dont les mécanismes semblaient correspondre à ceux observés expérimentalement.

J'obtins, pendant ce DEA, des résultats de convergence du *Q-learning* avec un procédé constructif qui ne faisait pas appel aux probabilités, contrairement aux résultats que l'on connaissait à l'époque. J'étais heureux d'avoir obtenu un résultat technique pour un algorithme dont la plausibilité expérimentale ne semblait faire aucun doute. Pour approfondir ce travail, je montrai les liens étroits entre le mécanisme qui permettait le bon fonctionnement du *Q-learning* et d'autres, fréquents dans certains modèles de physique. J'étudiai alors différents modèles de physique, comme **celui du tas de sable**, et **les modèles d'agrégation limitée par diffusion** (modèles DLA pour Diffusion Limited Aggregation). Ils semblaient permettre une analyse plus fine du comportement des différents algorithmes d'apprentissage par renforcement³. En ce qui concerne le Q-learning, les difficultés mathématiques et formelles semblaient maintenant bien déblayées et en partie évacuées. Je pus

³On ne trouvera que peu de traces de ce travail dans cette thèse.

désormais me consacrer complètement au projet pluridisciplinaire initial, l'étude de l'apprentissage par renforcement, c'est-à-dire des relations entre ces algorithmes et différentes expériences de psychologie [Blancheteau, 1978]. Je souhaitais approfondir cette analyse au cours du travail de thèse.

Mais des difficultés systématiques de confrontation avec des données réelles et effectives se manifestaient. J'avais déjà pressenti et présenté certaines de ces difficultés dans mon travail de DEA [Saurel, 1992]. De nouveau sous l'impulsion d'Hugues Bersini, je fis un séjour aux États-Unis, au cours duquel je rencontrais les meilleurs spécialistes internationaux de l'apprentissage par renforcement (Richard Sutton, Andrew Barto, Andrew Moore, Leslie Pak Kaelbling, Christopher Watkins, Peter Dayan, Satinder Singh). Malgré cela mes objectifs, qui consistaient à mettre en place, dans le cadre de l'apprentissage par renforcement, une véritable modélisation, semblaient presque irréalistes. Je ne rejétais pas la spécialisation dans son principe – elle est nécessaire et j'avais tout à fait joué le jeu pendant le travail de DEA. Je constatais cependant que **la modélisation, c'est-à-dire la confrontation étroite entre des travaux formels et des expériences**, en l'occurrence de psychologie, n'intéressait pas les spécialistes de l'apprentissage par renforcement. *La spécialisation extrême était un frein à la pluridisciplinarité telle que je la concevais, et elle me semblait stérile pour une discipline neuve comme les sciences cognitives.*

“L’homme de science engagé dans une recherche, dans le domaine de la physique par exemple, peut attaquer son problème immédiatement. Il peut aller tout de suite au cœur du sujet, c'est-à-dire au centre d'une structure organisée. C'est qu'il existe déjà un corps structuré de doctrines scientifiques et avec lui un état du problème généralement accepté. Il peut donc laisser à d'autres le soin d'ajuster sa contribution à l'édifice en construction de la connaissance scientifique” [Popper, 1978][p. 11].

La situation en sciences cognitives ne permettait pas l'approche spécialisée des problèmes dont parle Popper pour la physique. Une clarification préalable des objectifs, des méthodologies et des concepts en jeu autour de l'apprentissage par renforcement et au sein des sciences cognitives, m'apparaissait nécessaire. À y réfléchir, ceci n'était pas surprenant, puisque les sciences cognitives sont somme toute d'origine récente, en comparaison avec la longue histoire de la physique.

Je rencontrais donc des obstacles à ma volonté de modélisation de l'apprentissage par renforcement. Cette modélisation ne pouvait pas se pratiquer seul et la spécialisation observée semblait une difficulté insurmontable. Pour aller plus loin, je pouvais me consacrer à un travail isolé sur les algorithmes formels d'apprentissage par renforcement. Revenir à ma discipline d'origine ne m'aurait sans doute pas gêné s'il s'était agi de la psychologie cognitive, des neurosciences ou de la philosophie de l'esprit, où l'on peut envisager travailler sur des faits concrets et obtenir ainsi sa propre justification expérimentale. J'aurais simplement pratiqué la discipline dont j'étais issu en affirmant que j'en étudiais les aspects cognitifs. Cela n'engage à rien.

Mais je connaissais suffisamment les Mathématiques pour ne pas oser me présenter en sciences cognitives comme mathématicien. J'étais un modélisateur, pas un mathématicien. Comment le modélisateur pourrait-il revenir à sa discipline d'origine ? Que seraient des modèles cognitifs si on enlevait le terme cognitif ? Des systèmes formels, mais sans sémantique qui les éclaire et permette de les interpréter. Le terme de modèles cognitifs se réduit alors à la dénomination d'une classe de modèles. Mais le terme est trompeur, car ces modèles n'ont aucun lien particulier avec la cognition. On peut conserver le terme de modèle, mais on ne peut plus évoquer la notion de modélisation. **Celle-ci nécessite l'établissement d'une sémantique appropriée et non arbitraire, que l'on peut associer à un système formel sur la base de concepts cohérents étudiés sous d'autres aspects scientifiques.** Le risque était donc grand d'un enfermement sur des algorithmes et des modèles qui ne pouvaient pas être étroitement rattachés à d'autres disciplines relevant des sciences cognitives. On peut penser que ces difficultés sont liées au fait que les systèmes formels ne peuvent rien apporter aux sciences cognitives. Mais cela serait méconnaître l'origine et le fondement de ces sciences dans la notion de machine et de symbole. J'aurais aussi pu considérer comme certains que les sciences cognitives ne sont qu'une "bulle spéculative". Cela me semblait à tout le moins insuffisant, et une attitude plus constructive me paraissait nécessaire. Choisir un repli sur ma discipline aurait été en complète contradiction avec les motivations initiales qui m'avaient amené aux sciences cognitives.

Je me plaçais ainsi nettement en position de modélisateur. Pour le modélisateur en sciences cognitives, plus sans doute que pour les autres participants à ces sciences neuves, une réflexion s'imposait, qui nécessitait comme *a priori* absolu de considérer ces sciences comme un tout, un ensemble qui possédait une cohérence profonde.

Deux nécessités s'imposaient alors à moi. Je dégageai à partir d'elles deux objectifs de thèse. La première nécessité était de mieux connaître ma nouvelle discipline, les sciences cognitives, ses enjeux et son histoire, pour comprendre quelle place j'y prenais en tant que modélisateur. Ceci devenait mon premier objectif de thèse. La deuxième était de faire mes preuves comme modélisateur en sciences cognitives. Mes résultats théoriques obtenus en DEA m'avaient permis de montrer mes capacités comme manipulateur de systèmes formels. Pour être considéré comme un modélisateur, il faut cependant bien plus que cela. Mon expérience de l'apprentissage par renforcement m'avait montré que la modélisation nécessite d'autres exigences et en particulier **la volonté de rapprocher étroitement, dans un cadre d'interprétation commun, des systèmes formels et des faits d'expérience.** Je souhaitais réaliser effectivement la modélisation d'un phénomène précis, en relation étroite avec un spécialiste du phénomène. Ce deuxième objectif était nécessaire afin de mieux comprendre quelles formes d'interactions le modélisateur pouvait avoir avec le spécialiste du phénomène modélisé.

L'occasion me fut donnée par mon directeur de thèse, Paul Bourgine, de rencon-

trer Wolf-Dieter Eberwein, spécialiste de sciences politiques du Wissenschaftszentrum Berlin (WZB), qui s'intéressait en particulier au mouvement révolutionnaire en RDA en 1989. Ce phénomène, quoiqu'en apparence très différent des préoccupations relevant des sciences cognitives, nécessitait cependant pour sa compréhension des modèles très proches de ceux proposés actuellement pour étudier les systèmes nerveux : les modèles connexionnistes de réseaux neuronaux biologiques. *Pour un modélisateur, les systèmes formels étaient du même type.* J'aurais pu travailler sur ces modèles avec un biologiste. Les circonstances firent que je travaillais sur les mêmes modèles mais avec un politologue. Il s'agissait là, à partir d'un type de formalisme déterminé, de réaliser une véritable modélisation. Peu importait que ce fût en sciences politiques ou en biologie. Le principal était de savoir comment on attribue une sémantique à un modèle et comment on modifie le modèle, ces deux actions étant étroitement dépendantes l'une de l'autre. L'essentiel était de réaliser ainsi une modélisation pluridisciplinaire, en s'attachant à ne pas se contenter d'une simple juxtaposition de deux discours. Je remplissais ainsi mon deuxième objectif de thèse.

Pour un modélisateur la différence n'existant donc pas, au stade actuel de la pratique de la modélisation en sciences cognitives, entre un modèle du mouvement révolutionnaire en RDA et un modèle de systèmes nerveux. De plus il faut noter que les données sur le mouvement révolutionnaire en RDA, quoique partielles, étaient d'un accès relativement aisé. On possède également de très nombreuses données en biologie. Mais en comparaison les données sur les systèmes nerveux biologiques semblent très délicates à obtenir et à interpréter. De plus l'occasion ne s'était tout simplement pas présentée de travailler en relation étroite avec un spécialiste expérimentateur et théoricien en neurobiologie.

La conception et l'étude de cette modélisation m'a permis de comprendre à quel point la modélisation elle-même est une activité dynamique, qui nécessite des échanges incessants entre le "modélisateur" et le spécialiste du phénomène. **En ce sens la modélisation permet également une véritable pluridisciplinarité, conçue comme une interaction étroite entre des spécialistes de disciplines différentes.** Ce qui m'est apparu également avec netteté, c'est le rôle essentiel dans la modélisation de plusieurs "domaines" scientifiques "élémentaires", à savoir la théorie, les modèles, les simulations, l'analyse mathématique des modèles, l'analyse des simulations, la mesure sur les simulations, la mesure sur les phénomènes et peut-être d'autres encore.

L'activité scientifique consiste bien souvent à travailler de manière spécialisée sur un seul de ces domaines, ce qui demande déjà une très grande technicité. C'était d'ailleurs ainsi que je concevais initialement le rôle du modélisateur cherchant à analyser des modèles formels bien définis.

Ce qui m'a marqué, dans ce travail, a été le rôle essentiel que joue l'interaction entre ces différents domaines. Cette interaction est la trame qui permet de justifier

scientifiquement la modélisation proposée. Néanmoins chacun de ces domaines doit être distingué des autres. Il a sa propre nécessité et ne peut être réduit aux autres éléments. Le retrait d'un seul de ces domaines entamerait fortement la crédibilité scientifique de la modélisation réalisée.

Cette conception dynamique de **la modélisation comme une activité, assise sur une volonté de tenir ensemble la théorie, le modèle, la simulation et les autres éléments associés**, ne correspondait pas, semble-t-il, à la pratique effective en sciences cognitives. Cette constatation allait au-delà d'une simple analyse sociologique et n'était pas liée aux qualités scientifiques irréprochables des spécialistes de sciences cognitives. Cette pratique avait apparemment un enracinement historique. L'analyse de cette question me permit d'atteindre mon premier objectif de thèse : mieux connaître ma discipline, son histoire et donc les contraintes qu'elle s'impose.

Je savais que, depuis le développement de l'intelligence artificielle, et en particulier les travaux fondateurs de Simon et Newell, les systèmes symboliques sont souvent conçus pour eux-mêmes, et l'interaction étroite et forte avec un autre spécialiste du phénomène à modéliser est exceptionnelle. En sciences cognitives on parle encore de modèles à propos de ces systèmes formels. Il serait cependant abusif de dire que ces modèles sont issus d'une quelconque modélisation dans le sens que nous venons de discuter. D'un point de vue historique on pouvait penser trouver là un parfait bouc émissaire qui aurait été à l'origine des errements méthodologiques, observés aujourd'hui encore chez ceux qui prétendent concevoir des modèles mais ne réalisent jamais de modélisation dans le sens que nous avons précisé.

Si effectivement Simon et l'intelligence artificielle avaient été à l'origine de telles dérives méthodologiques, il importait de savoir quelles positions avaient été défendues auparavant. J'avais déjà entendu parler des conférences Macy, qui ont eu lieu aux États-Unis dans la deuxième moitié des années quarante. Elles sont souvent considérées comme l'acte fondateur des sciences cognitives. Je m'interrogeai alors naturellement sur les positions méthodologiques qui avaient été défendues à l'époque et sur la conception de la modélisation qui avait été préconisée. Ma surprise fut grande. Je découvrai une multitude de points de vue très contrastés. Les modélisateurs de l'époque, comme Wiener ou von Neumann, étaient également des mathématiciens. Même s'ils y prenaient peu d'attention, ils défendaient clairement des positions où une collaboration étroite avec le spécialiste du phénomène était nécessaire [von Neumann, 1951]. Pour eux le système formel ne pouvait se concevoir indépendamment d'une référence systématique à des faits d'expérience. *Les similarités constatées n'étaient jamais prises au pied de la lettre, naïvement, comme si elles constituaient des vérités ontologiques. En fait seule une position métaphysique statuant sur la nature ontologique des objets manipulés pouvait réduire cette multiplicité de positions à l'étude isolée des systèmes formels.*

Je m'attachai plus particulièrement à la personnalité de von Neumann. J'étais

surpris qu'il réagisse si peu aux positions ontologiques défendues autour de lui et en particulier à l'amalgame, souvent réalisé par les différents participants aux conférences Macy, entre les modèles ou les systèmes formels d'une part, et les machines symboliques matérielles ou les ordinateurs d'autre part. On identifiait également machine symbolique et machine vivante. Cette double identification des modèles aux machines symboliques, et simultanément des machines vivantes aux machines symboliques, expliquait que l'on en conclut, par une simple relation de transitivité, que l'on pouvait identifier les modèles et les machines vivantes. Les positions de Simon en découlaient simplement. *Si on avait identifié ces trois termes, on pouvait se contenter de travailler sur un seul d'entre eux, par exemple les systèmes formels, tout en affirmant que l'on travaillait bien en même temps sur les deux autres.*

Même si une position métaphysique était clairement nécessaire pour en arriver à ce point, il m'apparaissait également qu'on ne trouvait pas de telle position ontologique chez von Neumann et qu'il défendait seulement une position méthodologique. Comment était-il alors possible qu'il acceptât cette relation étroite entre modèles et machines symboliques matérielles ? Seule une analyse plus fine me permit de répondre à cette question. Il s'agissait en fait d'une compréhension profonde, de la part de von Neumann, des liens qui existent entre les systèmes formels et les machines matérielles. Von Neumann avait participé à l'évolution de la conception de la notion d'axiomatisation. La notion d'axiomatisation est évidemment très proche de la notion de modèle car elles posent toutes deux des fondements formels à partir desquels on essaye d'obtenir des conséquences de manière déductive.

La conception initiale et “naïve” de l'axiomatique, préconisée par Cantor et Hilbert [Hilbert, 1990] à la fin du siècle dernier, a profondément évolué à la suite des travaux sur la théorie des ensembles, auxquels von Neumann participa activement. La théorie des ensembles avait du mal à trouver ses propres fondements. Von Neumann montra même que les paradoxes ne pouvaient sans doute pas être supprimés de cette théorie [von Neumann, 1928]. Le modèle, considéré comme une axiomatique particulière, pouvait alors difficilement trouver ses propres fondements. Les travaux de Gödel en logique [Gödel, 1931], comme ceux de von Neumann sur la théorie des ensembles, montraient également l'extrême complication que pouvait constituer un système axiomatique. Les logiciens des modèles se trouvaient là face à un véritable nœud gordien. Il appartenait à Turing de le trancher [Turing, 1936]. Turing donna un nouveau sens à la notion de système formel. Il proposa d'associer cette notion à *une machine matérielle, quoiqu'abstraite*.

Von Neumann avait parfaitement compris cette modification de la notion d'axiomatisation. Ceci avait évidemment des conséquences sur sa conception des modèles, qui pouvaient ainsi être à la fois rattachés aux mathématiques et à *des systèmes matériels comme les ordinateurs*, sans que cela ne fût en aucune manière lié à une quelconque position ontologique. Ses positions soutenues lors des conférences Macy

en résultèrent. De plus les écrits que von Neumann nous a légués et qu'il a rédigés sur son lit de mort [von Neumann, 1992] prennent un sens très différent si on ne leur associe pas une position ontologique, comme les interprétations ultérieures ont pu le faire. L'étude de ces travaux a eu une profonde influence sur ma conception de la place des modèles et du rôle de la modélisation en sciences cognitives. *La matérialité y apparaît sous un jour nouveau, qui va bien au-delà de la machine de Turing et des machines symboliques.* Pour aller plus loin j'étais d'abord amené à approfondir ma connaissance des machines symboliques. Mais cette conception de la matérialité ne pouvait s'y retrouver que de manière "asymptotique". Il m'apparut alors avec clarté qu'il devenait nécessaire d'introduire une nouvelle notion, celle de **machine processuelle**, en opposition aux machines symboliques, et qui peut seule rendre compte immédiatement de cette matérialité.

Ainsi j'avais atteint les deux objectifs que je m'étais fixés. Ma thèse portait initialement sur l'apprentissage par renforcement. C'est pourtant la modélisation et sa place en sciences cognitives qui étaient devenues mon fil conducteur. Elles étaient systématiquement présentes, au moins en toile de fond, dans tous les aspects de mon travail depuis le DEA. Selon la conception qui me guidait, l'analyse de systèmes formels, pour elle-même, est essentielle, en particulier dans une discipline comme les mathématiques appliquées. Se contenter d'une telle analyse et prétendre faire de la modélisation relève cependant de l'escroquerie, nous semble-t-il, dans une discipline qui doit mettre en relation les modèles et les phénomènes, comme essayent de le faire les sciences cognitives.

Cette thèse a été rédigée en cinq chapitres à partir de ce parcours. Elle n'en constitue cependant pas une narration linéaire. **Dans tout ce travail nous réservons l'emploi du terme de modèle au sens de modèle formel qu'on lui attribue en physique.** La modélisation, et plus précisément la place des modèles en sciences cognitives, est évidemment le fil conducteur qui mène d'un chapitre au suivant. On peut cependant en trouver d'autres. **En particulier la notion de complexité est centrale dans notre travail.** On la retrouve sous diverses formes. Chaque chapitre la traite sous un angle différent. Le terme de complexité est celui généralement employé actuellement, mais il a eu de nombreux avatars. On a parlé dans les années quarante de cybernétique (chapitre 1), puis de systémique, dont le connexionnisme (chapitre 2) constitue une des formes les plus évoluées. **Le terme de complexité a ensuite été employé par les physiciens des transitions de phase (chapitre 3).** Le terme d'auto-organisation a également été utilisé dans ce vaste contexte. Nous évoquons en particulier les états critiques auto-organisés dans le chapitre 3. Dans le cinquième chapitre, nous sommes resté fidèle à cet état d'esprit même si, selon nous, nous abordons ici une complexité qui n'est pas étudiée actuellement et qui nécessiterait peut-être un nouveau terme.

L'objectif du chapitre 1 est de préciser, à partir du parcours de von Neumann, les fondements historiques dont sont issues les conceptions méthodologiques pratiquées en sciences cognitives.

Le chapitre 1, *L'évolution de la notion d'axiomatisation et ses relations avec l'étude du cerveau*, rassemble les éléments historiques concernant la place de la modélisation dans l'étude du cerveau. Il se divise en trois sections qui se succèdent dans un ordre chronologique de 1903 à 1996. Comme nous l'avons déjà expliqué, les positions de von Neumann sont profondément ancrées dans le formalisme de Hilbert et dans son évolution suite aux travaux de Gödel puis de Turing. La première section de ce chapitre, *Quarante ans d'évolution de la notion d'axiomatisation à travers le parcours de von Neumann (1903-1943)*, se concentre sur cet aspect. La deuxième section, *Les années quarante et l'unification de la conception du formalisme*, a pour thème principal l'évolution de la conception de la modélisation au cours des conférences Macy. Nous insistons sur la diversité de points de vue observée au cours des premières conférences. Mais les analogies qui ont permis ce premier dialogue pluridisciplinaire se transforment progressivement en identifications. On assiste alors au développement d'une conception formaliste de l'intelligence, que l'on observe également dans les travaux de Turing de l'après-guerre. C'est plus généralement une ontologie formaliste qui se développe et tend à unifier les positions des différents participants aux conférences Macy. **L'intelligence artificielle, dont les sciences cognitives sont les héritières, s'est fondée sur cette position métaphysique.** Chez Simon, ce n'est pas seulement l'intelligence mais le monde dans son ensemble qui est de nature symbolique. Ainsi le socle sur lequel s'appuient les sciences cognitives, comme l'intelligence artificielle, est formé d'une position ontologique symbolique issue d'une triple identification entre les modèles, les phénomènes et leur reconstruction matérielle. Cette troisième période, mieux connue, est rapportée brièvement dans notre troisième section intitulée *L'Intelligence Artificielle et les sciences cognitives, quarante de confrontation à une ontologie formaliste (1956-1996)*.

L'objectif du chapitre 2 est d'analyser ce qui peut constituer l'unité des sciences cognitives. Dans ce contexte, la notion de processus cognitif est centrale. Ces processus constituent l'objet d'étude principal des sciences cognitives. De plus la conception qu'en ont les sciences cognitives permet de préciser un certain nombre de choix méthodologiques, aussi bien pour les théories développées que pour les expériences mises en œuvre.

Le chapitre 2, *Comment les sciences cognitives conçoivent l'étude des processus*, est centré sur l'analyse de la notion de processus, considéré comme concept unificateur des sciences cognitives. Il se divise en trois sections. La première aborde

la notion de processus d'un point de vue général. La seconde et la troisième étudient les modes de traitement théorique et expérimental des processus. La première section, intitulée *Diversité et unité des sciences cognitives*, a pour objectif de montrer qu'au-delà de leur diversité apparente les sciences cognitives sont unifiées dans leur conception des processus. Dans un premier temps nous définissons dans notre travail une notion d'objet et de processus. Nous constatons ensuite que les sciences cognitives se chargent d'étudier les processus cognitifs, sans spécifier ce que l'adjectif cognitif apporte de plus à la notion de processus. Puis nous montrons que les sciences cognitives, d'un point de vue général, réduisent les processus à une modification d'objets sous-jacents. Ainsi, en sciences cognitives, les processus ne sont jamais considérés comme premiers. Dans une deuxième section, intitulée *Le connexionnisme un cadre général pour modéliser les processus "émergents"*, nous nous intéressons à une théorie particulière, le connexionnisme – tel qu'il est défini par Farmer. Nous établissons que dans ce cadre théorique également les processus sont réduits à une modification d'objets. Nous rappelons, comme le fait déjà Farmer, que, d'un point de vue purement mathématique, le cadre connexionniste ne nécessite pas d'être distingué des systèmes de règles, si ce n'est pour les représentations particulières qu'il privilégie. Nous discutons enfin les limites du connexionnisme liées au choix d'une causalité particulière, celle de la mécanique des chocs. Dans une troisième section, intitulée *En guise de conclusion sur la façon d'étudier les processus cognitifs*, nous discutons rapidement la façon dont on considère le concept de processus dans les études expérimentales. Là encore les processus sont réduits à des modifications d'objets qui sont seuls à être effectivement accessibles à la mesure. Nous montrons également que dans ce contexte le choix de paramètres, ou de plages de certains paramètres considérés comme cognitifs, ne va pas sans problème.

Ainsi, même si l'unité des sciences cognitives se retrouve dans le concept de processus, ces sciences réduisent systématiquement ces processus à des objets. La détermination de processus particuliers, que l'on pourrait considérer comme cognitifs, devrait alors se faire par la détermination d'objets particuliers considérés comme cognitifs ou par la spécification de fonctionnements particuliers de ces ensembles d'objets, ces fonctionnements seuls pouvant être considérés comme cognitifs. À l'heure actuelle nous ne connaissons pas de critère permettant de dénommer des objets ou des paramètres comme cognitifs et, ainsi que nous le montrons dans la troisième section, les critères envisageables introduiraient d'importantes difficultés. De même les fonctionnements particuliers, qui devraient permettre de spécifier quels systèmes sont cognitifs, sont étroitement liés à l'élaboration d'une causalité particulière et d'une logique spécifique associée, qui pourraient être considérées comme proprement cognitives. Comme nous l'établissons pour le connexionnisme, la causalité actuellement utilisée est celle de la mécanique des chocs, qui ne permet pas de distinguer parmi les systèmes physiques ceux qui sont cognitifs de ceux qui ne le sont pas. S'il s'avérait impossible d'établir un tel critère de démarcation, il serait alors inutile de rechercher une quelconque unité aux sciences cognitives. Certes les

systèmes cognitifs ne sont sans doute que des systèmes physiques. Mais ils seraient alors des systèmes physiques identiques à ceux actuellement étudiés par la physique. On pourrait dans ce cas s'interroger sur les éléments qui nous permettent tout de même, en tant qu'observateurs humains, de distinguer les systèmes cognitifs des autres systèmes physiques.

On peut donc penser que l'ensemble de ces difficultés vient du fait que les sciences cognitives réduisent les processus à des modifications d'objets et se trouvent alors dans l'impossibilité de préciser quel est l'aspect cognitif de leur réduction. Pour éviter ces difficultés nous suggérons qu'il faut considérer les processus comme premiers et éviter cette réduction. C'est en tant que processus que les processus particuliers effectivement étudiés en sciences cognitives sont cognitifs. C'est alors directement dans la notion immédiate de processus cognitif que doit être cherchée l'unité des sciences cognitives. Nous développons ce point de vue dans le chapitre 5.

L'objectif du chapitre 3 est de montrer comment on étudie aujourd'hui certains modèles précis qui entrent bien dans le cadre connexioniste de Farmer. Ces modèles font tous partie de ce que l'on appelle aujourd'hui la complexité. Mais la complexité n'est ici qu'une dénomination pour désigner le nombre important, voire infini, d'éléments et d'interactions entre ces éléments. On trouve dans la manipulation de cette infinité le principal point commun de ces différentes théories de la complexité. Nous parlons alors de complexité/quantité. Cette unité de la complexité recouvre une unité des outils employés mais cache une diversité des méthodologies mises en œuvre.

Le chapitre 3, *La complexité : des théories de la quantité ?*, essaye de cerner ce qui fait l'unité des théories actuelles de la complexité. Nous nous concentrons sur certains aspects de deux de ces théories, celle des transitions de phase et celle de la complexité algorithmique. Pour la première nous insistons essentiellement sur les modèles de type Ising et pour la deuxième nous nous attachons uniquement aux automates cellulaires, et plus particulièrement à leur étude par Stephen Wolfram. Dans une première section, *Les modèles de la complexité/quantité, cas particulier du cadre de Farmer*, nous mettons en évidence que les modèles d'Ising, de même que celui de Heisenberg ou que les automates cellulaires, peuvent être décrits dans le cadre très général défini par Farmer. Il est donc naturel de s'intéresser à ces modèles comme exemples précis de ce cadre. Le premier intérêt de ces modèles pour les sciences cognitives est donc de fournir des exemples précis de modèles "connexionnistes"⁴ largement étudiés en physique et en *Computer Sciences*. Ces modèles présentent encore un autre intérêt pour les sciences cognitives. En effet ils constituent les éléments techniques auxquels on renvoie implicitement lorsque

⁴Le terme doit être évidemment entendu dans le sens de Farmer pour désigner une classe de modèles. Ce terme ne renvoie pas spécifiquement à des modèles en relation avec les réseaux de neurones biologiques.

l'on veut justifier le bien fondé des conceptions “émergentielles” de la cognition en sciences cognitives. On ne peut donc pas éluder l'étude détaillée de ces modèles, sauf à laisser le champ libre à des discours généraux sur le concept de cognition, qui prétendraient s'appuyer sur des résultats techniques soi-disant incontestables obtenus par la physique ou par les *Computer Sciences*.

On constate en fait que l'unité de ces modèles se retrouve principalement dans la nécessité d'un nombre infini d'éléments pour obtenir des résultats théoriques. La quantité est alors introduite pour des raisons purement techniques. Parallèlement, on observe un grand nombre de neurones dans le cerveau. Cette observation est le meilleur argument pour justifier l'emploi de ces modèles en sciences cognitives. Paradoxalement, un facteur introduit pour des raisons techniques devient prétendument un argument de validation expérimentale. La quantité est le cœur de ces conceptions de la complexité. Les modèles de la physique des transitions de phase renvoient alors préférentiellement à des modèles dont les variables sont continues, alors que les automates cellulaires constituent des modèles dont les variables sont discrètes. La différence entre les deux approches est donc essentiellement de point de vue. Dans cette section nous mettons également en évidence la multiplicité des comportements qualitatifs des différents modèles de transitions de phase. Nous insistons sur le nombre restreint de résultats théoriques exacts et sur leur technicité. Nous mettons également en avant le fait que le comportement d'un modèle obtenu avec un calcul exact peut être très différent de celui obtenu avec un calcul approché, par exemple avec une approximation de champ moyen. La différence peut même être radicale puisque les comportements peuvent être qualitativement différents.

Dans une seconde section, *Questions et outils pour aborder la complexité/quantité*, nous sommes amené à envisager quels outils peuvent être utilisés pour étudier ces modèles. En effet cette question est importante dans la mesure où Farmer, dans son article, propose un cadre général mais ne précise pas quels sont les outils disponibles. Nous présentons ces outils dans le cadre discret des automates cellulaires étudiés par Wolfram. On y retrouve des outils pour l'étude locale et l'étude globale des automates, comme les notions de loi puissance, d'exposant critique, d'approximation en champ moyen, d'équation maîtresse, d'exposant de Lyapounov, et celle d'entropie. Ces outils proviennent donc à la fois de la mécanique statistique et de la théorie de l'information ou de la théorie de la complexité algorithmique. Ainsi nous avons déjà dans les mains des outils d'étude théorique permettant d'analyser l'ensemble des modèles qui entrent dans le cadre de Farmer. Mais ces modèles portent également avec eux des questions spécifiques que l'on est susceptible de se poser, de manière plus générale, pour l'ensemble des modèles qui relèvent du cadre de Farmer.

Les théories de la complexité présentent donc aujourd’hui une certaine unité par le biais de leur objet d'étude, qui nécessite la quantité. Il importe cependant de noter que cette unité sur la spécificité de l'objet d'étude, que l'on retrouve également dans certains outils utilisés, cache en fait de profondes divergences dans la conception de la modélisation et des méthodologies d'étude des modèles. Dans la troisième

section *Trois méthodologies de la complexité/quantité*, nous présentons trois modèles différents, d'abord les modèles de gaz sur réseau, puis le modèle du tas de sable, et enfin des modèles d'évolution. À chacun de ces modèles on peut associer une méthodologie différente de l'étude des modèles. Nous discutons et comparons ces trois méthodologies.

Ce chapitre est suivi de trois annexes. La première, *Quelques dynamiques des modèles d'Ising*, présente une liste de dynamiques des modèles d'Ising. La deuxième, *Les vingt problèmes de Wolfram*, est une présentation et une discussion personnelle de vingt questions posées par Wolfram. Elle permet d'illustrer par des exemples les questions que l'on se pose habituellement lorsque l'on étudie des modèles qui relèvent de la complexité/quantité. Dans la troisième, *La classification de Gutowitz*, nous expliquons brièvement un mode de classification des comportements des automates à partir d'une approximation avec des chaînes de Markov.

L'objectif du chapitre 4 est de mettre en pratique sur un exemple précis les outils et les méthodologies de la modélisation mis en évidence dans les théories de la complexité/quantité.

Le chapitre 4, *La pratique de la modélisation : la mobilisation de masse en RDA en 1989*, est centré autour de la modélisation d'un phénomène de sciences politiques. Il se divise en trois sections. La première section, *Introduction à la pratique de la modélisation*, trace à grands traits et de manière rudimentaire notre conception de la pratique actuelle de la modélisation, résumée par une figure. Les méthodologies mises en évidence pour les théories de la complexité et étudiées dans la dernière section du chapitre 3 s'intègrent toutes trois dans ce schéma général de la modélisation. L'objectif de cette section est double. D'une part nous souhaitons montrer que la modélisation est une activité extrêmement sophistiquée, qui est loin de se réduire à l'étude de modèles. Au contraire elle met en jeu un nombre important de domaines comme la théorie, la simulation, les mesures sur ces simulations, les expériences, etc. D'autre part nous voulons insister sur le caractère dynamique de la modélisation, celle-ci étant une activité, contrairement au modèle qui est par lui-même statique. Dans la deuxième section, *La modélisation mise en pratique : la mobilisation de masse en RDA en 1989*, nous présentons la modélisation réalisée, dans son état actuel d'avancement. Nous décrivons successivement le phénomène étudié et son interprétation théorique puis les éléments qui ont été dégagés, et nous donnons une interprétation en langue naturelle du modèle, avant d'en présenter une expression quasi-formelle. Nous décrivons ensuite la simulation effectuée dans l'environnement de programmation Smalltalk et nous analysons les résultats obtenus. Dans une troisième section, *Pour aller plus loin dans la pratique de la modélisation*, nous présentons les directions à suivre pour poursuivre cette modélisation. Après avoir mis en évidence les difficultés rencontrées, nous établissons les liens étroits entre cette étude de sciences politiques et la pratique de la modélisation en sciences cognitives.

Cette proximité est attribuable en partie au fait que le modèle défini dans cette étude est extrêmement proche, dans son expression, des modèles connexionnistes. Nous montrons également que certaines questions et certains résultats obtenus dans ce cadre peuvent et doivent être posés plus généralement en sciences cognitives. Nous mettons notamment en relief l'intérêt d'étudier l'évolution du comportement qualitatif des modèles de réseaux neuronaux en fonction du nombre de neurones du modèle. Ce type de travail n'est pas réalisé, à notre connaissance, en sciences cognitives, pour les simulations de réseaux neuronaux. Ceci amène à s'interroger sur la sémantique que l'on peut alors attribuer aux "neurones" de ces simulations et sur les moyens qui permettent de comparer les observations sur ces simulations et celles réalisées sur des neurones réels.

Le cinquième et dernier chapitre fait office de conclusion et d'ouverture. Ce chapitre a deux objectifs. Le premier est de montrer l'extrême complication que peut constituer une tentative de modélisation ou même l'élaboration d'un cadre de modélisation du cerveau/esprit dans son ensemble. Le deuxième est de proposer une voie certes étroite, mais qui permet tout de même d'étudier la cognition d'un point de vue unifié et global tout en respectant les contraintes de la modélisation. Pour cela nous proposons et défendons la thèse suivante : la complexité associée à la cognition nécessite la matérialité.

Le chapitre 5 s'intitule *La complexité associée à la cognition nécessite la matérialité : une thèse pour une complexité au-delà de la complexité/quantité*. Dans une première section, *La complexité au-delà de la quantité*, nous montrons que les caractéristiques du cerveau qui nous amènent à parler de complexité à son propos ne se résument pas au nombre astronomique de ses composants et des connexions qu'ils établissent. Il faut également tenir compte de la diversité des éléments cérébraux, ce qui entraîne de nombreuses contraintes sur les modèles envisageables. Lorsque l'on effectue des études locales, cette diversité peut être réduite grâce à la thèse que nous nommons thèse de l'identité logique. Cette thèse consiste simplement à postuler que chaque élément pris isolément est isomorphe (ou a un fonctionnement isomorphe) à une unique structure logique de référence. Chaque élément peut alors être rapporté *isolément* à une structure de référence. Cependant, pour réduire la diversité à une unité plus profonde quand on recherche un cadre d'explication global, il faut que l'on soit capable de ramener, *en même temps*, l'étude des différents éléments à celle d'une unique structure logique. Pour cela il faut préciser un isomorphisme global dont chaque restriction correspondrait à l'isomorphisme postulé pour chaque élément par la thèse de l'identité logique. Nous appelons alors isolation normative l'attitude qui consiste à étudier une unique structure logique de référence sans expliciter le moyen de ramener dans chaque cas la structure particulière à cette structure de référence. Cette attitude consiste donc à éviter de préciser les isomorphismes globaux et leur

restriction. Nous établissons les difficultés logiques que présente cette isolation normative lorsque l'on cherche à établir un cadre d'explication global. Le rejet de l'étude d'une unique structure logique nous amène à considérer la nécessité de proposer des modèles comportant une mixité de structures logiques. De plus les observations biologiques concordent sur le fait que le cerveau est structuré. Pour proposer un cadre global de modélisation pour le cerveau, il faut donc proposer d'étudier une hiérarchie de structures logiques mixtes, ce que nous appelons hétéarchie.

Les difficultés s'accumulent donc pour celui qui veut définir un cadre global pour la modélisation du cerveau. Il devient nécessaire de proposer un autre angle d'approche. C'est ce que nous faisons dans la deuxième section, *Que nous impose cette nouvelle complexité ?* Pour cela nous énonçons et défendons la thèse selon laquelle la complexité associée à la cognition nécessite la matérialité. La matérialité est considérée ici comme essentiellement processuelle. Comme nous l'expliquons cette thèse est un stimulant, un repoussoir, qui nous oblige à concevoir de nouvelles machines symboliques, toujours plus complexes. Elle est l'occasion également de proposer une nouvelle définition de la cognition. Cette définition permet d'écartier l'étude des systèmes cognitifs sous la forme de leur réduction quantique, ce qui serait nécessaire si on considère que les systèmes cognitifs sont des systèmes physiques sans autre spécificité. Cette thèse permet ainsi de considérer les processus comme premiers et d'éviter ainsi une conception réifiante de nos abstractions sous forme de processus. Nous définissons également une notion d'infini en acte comme transcendance particulière dont nous montrons le caractère opérationnel. Dans ce contexte le rôle et le statut de l'observateur sont extrêmement importants, puisque la caractéristique principale de la thèse défendue est de proposer une conception pour laquelle certaines émergences observées sont également des émergences intrinsèques. Cette thèse n'est pas pour autant ontologique. Il convient alors de comparer la thèse défendue à celle, symbolique, de Simon. On est ainsi amené à concevoir de nouvelles machines, les machines processuelles, qu'il faut opposer aux machines symboliques et qui en constituent un prolongement dans un sens très spécifique. On aboutit alors à une nouvelle simplification en sciences cognitives, très différente de celle qui consiste à étudier un nombre important d'éléments simples et tous identiques. Nous proposons d'étudier des machines processuelles simples, ce qui permet de préserver aux systèmes cognitifs une complexité autre que la complexité/quantité. Pour cela la méthodologie de la modélisation, telle qu'elle est présentée dans le chapitre 4, ne suffit plus. Un nouveau domaine intervient, la reconstruction du phénomène sur le même substrat. Nous ne parlons plus de modélisation mais de mise en parangon.

RÉFÉRENCES BIBLIOGRAPHIQUES DE L'INTRODUCTION

- [Barto *et al.*, 1983] BARTO A. G., SUTTON R. S. ET ANDERSON C. W. Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Transactions SMC*. 1983, vol. 13, no 5, p. 834–846.
- [Blancheteau, 1978] BLANCHETEAU M. *Motivations et guides des déplacements chez l'animal (renforcements positifs)*. Thèse d'État en Psychologie, Université René Descartes, Paris, 1978.
- [Cavaillès, 1976] CAVAILLÈS J. *Sur la logique et la théorie de la science*. Troisième édition. Librairie philosophique J. Vrin, Paris, 1976. 94 p. Collection Problèmes et controverses. Première édition Presses Universitaires de France, 1947.
- [Dugas, 1959] DUGAS R. *La théorie physique au sens de Boltzmann*. Éditions du Griffon, Neuchâtel, 1959. 314 p. Volume 33 de la collection Bibliothèque scientifique.
- [Gödel, 1931] GöDEL K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*. 1931, vol. 38, p. 173–198. Reproduit dans les Collected works de Kurt Gödel. S. Feferman, J. Dawson, S. Kleene, G. Moore, R. Solovay, J. van Heijenoort, éditeurs. Texte avec la version originale en allemand et une traduction en anglais. Volume I, p. 144-195. Oxford University Press, New York, 1986. ISBN 0-19-503964-5.
- [Hilbert, 1990] HILBERT D. *Sur les problèmes futurs des mathématiques : les 23 problèmes*. Première édition Gabay. J. Gabay, Paris, 1990. 60 p. Réimpression de Compte rendu du deuxième congrès international de mathématiques tenu à Paris du 6 au 12 août 1900, Gauthier-Villars, 1902. ISBN 2-87647-037-3. ISSN 0989-0602.
- [Popper, 1978] POPPER K. *La logique de la découverte scientifique*. Reproduction de la première édition française parue en 1973. Payot, Paris, 1978. 484 p. Collection Bibliothèque scientifique. Traduit de The logic of scientific discovery. Première édition anglaise Hutchinson, 1959. Première édition allemande 1934. ISBN 2-228-11391-3.

- [Rescorla et Wagner, 1972] RESCORLA R. ET WAGNER A. A theory of pavlonian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Dans *Classical conditioning II: current research and theory*, Black A. et Prokasy W., éditeurs. Appleton-Century-Crofts, New York, 1972.
- [Saurel, 1992] SAUREL P. *Q-learning et apprentissage par renforcement*. DEA de sciences cognitives, EHESS, Paris, 1992.
- [Turing, 1936] TURING A. M. On computable numbers with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*. 1936, vol. 42, p. 230–265. Corrections, ibid. 1937, vol. 43, p. 544-546. Reproduit dans The undecidable: basic papers on undecidable propositions, unsolvable problems and computable functions. M. Davis, éditeur. p. 116-154. Raven Press, Hewlett (New Jersey), 1965.
- [von Neumann, 1928] VON NEUMANN J. Die Axiomatisierung der Mengenlehre. *Mathematische Zeitschrift*. 1928, vol. 27, p. 669–752. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume I, Logic, theory of sets and quantum mechanics, p. 339-422. Pergamon Press, Oxford, 1961-63.
- [von Neumann, 1951] VON NEUMANN J. The general and logical theory of automata. *Cerebral mechanisms in behavior - the Hixon symposium, Pasadena, 1948*. 1951, p. 1–31. L. Jeffress, éditeur. J. Wiley, New York. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 288-328. Pergamon Press, Oxford, 1961-63.
- [von Neumann, 1992] VON NEUMANN J. *L'ordinateur et le cerveau*. Première édition française. La découverte, Paris, 1992. 130 p. Collection Textes à l'appui/série sciences cognitives. Traduction de The computer and the brain - the Silliman lectures. Première édition américaine, Yale University Press, New Haven, 1958. ISBN 2-7071-2164-9.
- [Watkins, 1989] WATKINS C., 1989. *Learning from delayed rewards*. PhD thesis, Université de Cambridge, Angleterre.

CHAPITRE 1

**L'évolution de la notion
d'axiomatisation et ses relations
avec l'étude du cerveau**

1. QUARANTE ANS D'ÉVOLUTION DE LA NOTION D'AXIOMATISATION À TRAVERS LE PARCOURS DE VON NEUMANN (1903-1943)

L'étude du cerveau par des scientifiques qui n'étaient pas à l'origine des biologistes est très fortement liée à l'évolution de la conception des formalismes et de leur rôle. Au début du siècle, Hilbert a défendu que l'axiomatisation était une méthode susceptible de fonder les sciences et les sciences expérimentales en particulier. Progressivement et avec von Neumann¹ plus spécialement, on observe que l'axiomatisation se transforme en formalisation. La conception des formalismes vient à changer jusqu'à la guerre. À partir de 1943, sous l'impulsion de certains neurobiologistes, on va chercher à appliquer la méthode axiomatique à l'étude du cerveau et de son fonctionnement. Von Neumann jouera aussi un rôle très important dans cette tentative. L'intelligence artificielle (I.A.) sera fondée par des scientifiques qui auront reçu les enseignements de von Neumann. L'I.A. et les sciences cognitives aujourd'hui sont donc également les héritières indirectes de cette approche dans leur conception du rôle des formalismes.

1.1. De l'axiomatisation à la formalisation : von Neumann face à quatre problèmes de Hilbert

1.1.1. Le programme de Hilbert

En 1900, au Second Congrès de Mathématiques qui se tient à Paris, David Hilbert, professeur à l'Université de Göttingen, prononce une conférence, dans laquelle, sur une suggestion de son collègue Hermann Minkowski, il choisit de “deviner le futur” des mathématiques à travers un choix de problèmes. Il propose à cette occasion une liste de 23 problèmes qu'il considère comme fondamentaux pour l'époque à venir [Hilbert, 1990]. De cette manière, il propose non seulement des directions pour l'avancement des mathématiques, mais il précise également quelles relations elles peuvent établir avec les autres disciplines. Hilbert avait d'abord songé à consacrer cette conférence à une étude des relations entre les mathématiques et la physique, pour répondre ainsi à celle tenue quatre ans auparavant par son grand rival Henri

¹1903 est la date de naissance de von Neumann.

Poincaré et dans laquelle il plaideait pour des relations étroites entre ces deux disciplines.

De fait, on constate que c'est une méthodologie et une façon de rendre compte des connaissances que propose Hilbert. "Je pense que partout où se présentent des idées mathématiques, soit en Philosophie (théorie de l'entendement), soit en Géométrie, soit en Physique, le problème se pose de la discussion des principes fondamentaux, bases de ces idées, et de l'établissement d'un système simple et complet d'axiomes ; et cela doit se faire de telle façon que la rigueur des nouvelles définitions et leur applicabilité ne le cèdent en rien aux anciennes définitions arithmétiques" [Hilbert, 1990][p. 8]. Parmi les problèmes proposés, on peut en citer quatre qui ont eu une influence essentielle sur les relations entre les mathématiques et les autres disciplines et qui expliquent *a posteriori* les relations actuelles des sciences cognitives avec la formalisation. Ces différents problèmes ont constitué des difficultés techniques et ont permis en même temps d'amender et de modifier profondément la conception de la notion d'axiomatisation². Ces problèmes sont "Le traitement mathématique des axiomes de la physique" (problème 6), le "Problème de M. Cantor relatif à la puissance du continu" (problème 1), "De la non-contradiction des axiomes de l'Arithmétique" (problème 2), "De la possibilité de résoudre une équation de Diophante" (problème 10).

1.1.2. L'axiomatisation selon Hilbert

Les trois premiers problèmes qui nous intéressent sont directement liés à la conception de Hilbert selon laquelle, "lorsqu'il s'agit de poser les principes fondamentaux d'une science, l'on doit établir un système d'axiomes renfermant une description complète et exacte des relations entre les concepts élémentaires de cette science. Ces axiomes sont en même temps les définitions de ces concepts élémentaires" [Hilbert, 1990][p. 14]. Dès 1899, Hilbert avait lui-même rédigé une présentation des fondements de la géométrie [Hilbert, 1971] selon la méthode axiomatique. Dans ce travail, il ne se contente pas de présenter les axiomes qui fondent la géométrie, il les expose dans l'ordre nécessaire de leur apparition, en vérifiant systématiquement leur indépendance. Il cherche en fait la présentation la plus simple possible de la géométrie. "Ces axiomes ne renferment-ils pas des parties communes superflues que l'on doit supprimer si l'on veut obtenir un système d'axiomes complètement indépendants ?" [Hilbert, 1990][p. 15].

L'exigence de simplicité n'est donc pas seulement esthétique, elle correspond à une volonté d'élucider toutes les difficultés susceptibles de se présenter dans le système. Hilbert cherche à préciser une conception, celle de systèmes formels que

²Hilbert lui-même a changé dans sa conception de l'axiomatisation. On peut dégager deux périodes, une première autour de 1900 et une seconde à partir des années 1920. Nous n'exposons ici que les conceptions de Hilbert de la première période.

l'on pose et qui en ce sens sont relatifs, mais dont on aurait éradiqué toute erreur postérieure possible. Ce type de présentation permet donc selon Hilbert d'expliciter totalement les hypothèses qui sont faites pour obtenir les résultats. Le critère de simplicité doit permettre d'éviter toute possibilité de contradiction interne. "Mais avant tout, parmi tant de questions soulevées par l'examen des axiomes, je regarde comme la plus importante celle-ci : Démontrer que les axiomes ne sont pas contradictoires ; c'est-à-dire démontrer qu'en se basant sur les axiomes l'on ne pourra jamais arriver à des résultats contradictoires au moyen d'un nombre fini de déductions logiques" [Hilbert, 1990][p. 15].

Pour Hilbert, la méthode axiomatique s'oppose aux autres méthodes de fondation et en particulier à la méthode dite "génétique", comme il la nomme, qui permet en particulier de construire les nombres. "La construction de la géométrie est toute différente. Ici, d'habitude on admet l'existence de certains éléments [...], on établit entre eux des relations en posant certains axiomes [...]. Le problème se pose alors de montrer la compatibilité de ces axiomes et leur intégrité, c'est-à-dire qu'il faut prouver que l'application de ces axiomes ne peut pas conduire à des contradictions et que l'ensemble des axiomes permet la démonstration de tous les théorèmes de la géométrie. Nous qualifierons cette méthode d'axiomatique" [Hilbert, 1971][Sur le concept de nombre, appendice vi, p. 256]. Hilbert souhaite distinguer son approche de celle génétique parce que le but, dans l'approche axiomatique est de "dégager [la théorie] des significations concrètes et intuitives sur lesquelles elle a d'abord été construite afin d'en faire clairement apparaître le schéma logique abstrait" [Blanché, 1990]³[p. 55]. Hilbert précise ce que l'on peut accepter comme vrai avec une bonne axiomatisation. Toutes les déductions doivent s'appuyer exclusivement sur les axiomes. Ainsi, le système d'axiomes et les conséquences qu'on en tirera présentent une certaine clôture. Seuls les raisonnements s'appuyant sur un nombre fini d'étapes sont acceptés. Ce dernier point est le germe des définitions de la calculabilité qui seront présentées par Turing, Post et Church. "Aucune affirmation relative à la science dont nous examinons les principes fondamentaux ne sera admise comme exacte, à moins qu'on ne puisse la tirer des axiomes au moyen d'un nombre fini de déductions" [Hilbert, 1990][p. 15].

1.1.3. L'axiomatisation de la physique

C'est dans cette optique que se place le sixième problème de Hilbert, "le traitement mathématique des axiomes de la physique". Une partie de ce travail fut réalisée par von Neumann dans le cadre de la mécanique quantique. Celui-ci est parvenu à unifier la formalisation de la mécanique quantique de Heisenberg, dite <<mécanique des matrices>> et les travaux de Schrödinger, ou <<mécanique ondulatoire>>, grâce

³Blanché poursuit : "or, à cet égard, les premières axiomatiques souffrent encore de bien des imperfections, comme on le voit avec celle de Hilbert".

à une application systématique des travaux sur les opérateurs auto-adjoints dans les espaces hilbertiens. Von Neumann va cependant plus loin que le cadre classique des espaces de Hilbert en développant des résultats nouveaux sur les opérateurs non bornés à l'occasion de ces travaux⁴. C'est dire que la réussite de von Neumann s'appuie à la fois sur les travaux mathématiques de Hilbert et sur la conception hilbertienne des relations entre les mathématiques et la physique.

Les travaux de von Neumann laissent penser qu'en développant les bons outils mathématiques il est possible de fonder une science expérimentale comme la physique à partir d'une axiomatisation. Cette méthodologie permet d'intégrer la sémantique dans l'axiomatique et dans l'outil formel développé. Pour être plus précis, cela veut dire que toute la signification est intégrée soit dans les axiomes, soit dans les outils de démonstration et de déduction. Il est alors superflu et même faux de chercher une interprétation hors des axiomes ou des outils de démonstration. Par exemple, la sémantique de la mécanique quantique doit être recherchée soit dans les axiomes de la théorie, soit dans les axiomes et la sémantique apportés par les mathématiques utilisées et nécessaires concernant les espaces de Hilbert. Le reste n'est que syntaxe et n'est pas signifiant⁵.

1.1.4. Axiomatiser la théorie des ensembles ?

On retrouve cette influence essentielle de Hilbert dans les développements du premier problème qu'il a énoncé, celui qu'il appelle “le problème de M. Cantor relatif à la puissance du continu”. Il inclut en fait dans ce problème l'ensemble des questions soulevées par la construction cantoriennne. On y trouve ensemble l'hypothèse du continu et les questions liées à la notion de bon ordre et à la construction des nombres. Suite à ce que l'on a appelé la crise des fondements et pour éviter la constitution des antinomies, Hilbert souhaite que soit développé un travail d'axiomatisation et de formalisation de la théorie des ensembles. Le but est de fonder les mathématiques elles-mêmes, soit en les axiomatisant, c'est le choix de Hilbert, soit en les réduisant à la logique, c'est le choix de Russell et Frege. Pour Hilbert l'axiomatique est “la méthode [qui] a l'avantage⁶ pour un exposé définitif d'une science, et elle donne à ses bases la sécurité logique indispensable” [Hilbert, 1971][appendice vi, p. 256].

En 1904 Zermelo prend l'axiome du choix comme point de départ pour sa dé-

⁴On trouve ces résultats dans le livre de von Neumann sur la mécanique quantique [von Neumann, 1988] ou, pour une présentation dans un cadre mathématique plus large, dans le livre classique de Rudin [Rudin, 1995].

⁵Cette interprétation du travail de von Neumann reste très schématique. Il est très vraisemblable que sa conception de l'axiomatisation des sciences expérimentales a déjà beaucoup évolué quand il rédige son livre de 1932.

⁶Il s'agit ici d'une traduction du terme allemand *Vorteil*. Une traduction plus adaptée serait : l'axiomatique est “la meilleure méthode pour un exposé définitif d'une science, et elle donne à ses bases la sécurité logique indispensable”.

monstration que tout ensemble peut être bien ordonné [Zermelo, 1904]. En 1908, Zermelo exprime l'intention de son article de "montrer comment toute la théorie créée par Cantor et Dedekind se laisse ramener à quelques définitions et à 7 axiomes qui semblent indépendants les uns des autres" [Zermelo, 1908][p. 261]. Selon Cavailles, "la méthode est celle dont Hilbert avait donné le modèle pour la géométrie" [Cavaillès, 1938][p. 121]. Ces travaux seront ensuite développés, dans le même état d'esprit, par Fraenkel. En effet, l'œuvre de Zermelo était restée inachevée. Il n'avait montré ni la possibilité de déduction de la théorie de l'ordre et des nombres ordinaux à partir des axiomes, ni l'indépendance et la non-contradiction des axiomes. Le mérite des travaux de Fraenkel aura été d'avoir interprété la phrase "propriété décidable au moyen des axiomes de la théorie", dans les termes mêmes de la théorie. On quitte ainsi l'"axiomatisation naïve" de Zermelo. Progressivement, c'est la conception de la notion d'axiomatisation elle-même qui est modifiée, en imposant à l'ensemble de se décrire lui-même en termes formels. Désormais toute l'axiomatique est écrite en termes formels et symboliques. L'intuition et le sens particulier que l'on peut donner aux symboles sont extirpés. Le système axiomatisé devient ainsi progressivement clos et formel. On peut parler désormais de formalisation.

C'est von Neumann cependant qui va achever provisoirement l'axiomatique nécessaire à la construction cantorienne. Il va proposer une nouvelle axiomatique plus complète que celle de Fraenkel, car elle permet de rendre compte de la construction et de certaines propriétés des nombres transfinis. Dans son article de 1925, von Neumann écrit : "le but de ce travail est d'offrir une présentation axiomatique de la théorie des ensembles qui soit logiquement inattaquable" [von Neumann, 1925] [p. 35]. En conclusion du même article, il émet des réserves sur le fait que sa théorie des ensembles soit complètement satisfaisante. Même s'il ne les construit pas directement dans cet article, von Neumann constate que de nouvelles antinomies sont encore possibles dans le cadre de l'axiomatique qu'il propose. Il écrit : "nous ne pouvons pas pour l'instant faire mieux que constater qu'il y a ici des objections contre la théorie des ensembles et qu'il n'y a actuellement aucun moyen connu pour sa réhabilitation" [von Neumann, 1925][p. 56]. L'article de von Neumann de 1928 est exemplaire par sa concision [von Neumann, 1928]. Il présente les axiomes de sa théorie des ensembles en une page. À ce propos, Ulam écrit : "le système de von Neumann constitue la première fondation de la théorie des ensembles sur la base d'un nombre fini d'axiomes possédant la même structure logique simple que par exemple ceux de la géométrie élémentaire. La concision du système d'axiomes et le caractère formel du raisonnement utilisé réalisent l'objectif de Hilbert de traitement des mathématiques comme un jeu fini" [Ulam, 1958][p. 11]. Il ajoute même que "l'on peut deviner ici le germe du futur intérêt de von Neumann pour les machines à calculer⁷ et la "mécanisation" des preuves".

⁷Il faut entendre ici "machines à calculer" dans le sens de procédure de manipulation automatisée de symboles avec un nombre fini de symboles et de manipulations.

Pourtant de nouveaux problèmes se posent. Chaque axiomatique proposée doit faire face à de nouvelles antinomies, construites *a posteriori* et risquant de mettre à mal l'édifice tout entier. Les axiomatiques ressemblent à d'ingénieuses miniatures “qui représentent fidèlement la théorie des ensembles jusque dans tous les détails [...]. Il y paraît toutes les puissances⁸ [...]. Mais aussitôt que l'on met en jeu des instruments plus fins, tout s'évanouit. De toutes les puissances ne restent que le fini et le dénombrable. Cela seul a un sens réel, tout le reste est fiction formaliste” [von Neumann, 1925]⁹[p. 56]. Dans cet article, von Neumann montre que la puissance d'un ensemble est relative à l'axiomatique utilisée. Ceci remet en cause la méthode axiomatique dans sa capacité à ne pas faire appel *a posteriori* à l'intuition. “C'était un avantage de la méthode axiomatique que de réunir, par l'identité de leur structure, une pluralité de systèmes isomorphes. Si, maintenant, les systèmes qu'elle réunit peuvent ne pas être isomorphes, c'est donc qu'elle laisse échapper certaines particularités des structures et qu'elle ne suffit plus à différentier celles-ci. Pour les distinguer, un recours à l'intuition sera nécessaire”[Blanché, 1990]. Seule une démonstration de non contradiction donnerait la sécurité à cette construction. On est passé progressivement d'interrogations sur l'axiomatique à une interrogation sur la logique de cette axiomatique .

En 1938, une nouvelle étape est franchie par Gödel. Celui-ci montre que si la théorie des ensembles est consistante, axiomatisée par l'un des systèmes couramment admis tel que celui de Zermelo-Fraenkel (ZF), alors cette théorie augmentée de l'axiome du choix et de l'hypothèse du continu est encore consistante. En 1963, une nouvelle étape est franchie par Cohen. La consistance de ZF entraîne celle de ZF augmentée de la négation de l'axiome du choix, ainsi que celle de ZF augmentée de l'axiome du choix et de la négation de l'hypothèse généralisée du continu. On aboutit donc à l'indépendance relative de l'axiome du choix par rapport à ZF et de l'hypothèse généralisée du continu par rapport à ZF avec l'axiome du choix. L'essentiel dans cet historique est de constater que, même si la formalisation de la théorie des ensembles est difficile, elle paraît possible. On semble aboutir à une possibilité de fonder les mathématiques par une construction progressive et formalisée. Malgré les difficultés rencontrées et l'évolution de la notion même de formalisme, tout ceci reste parfaitement dans l'optique de Hilbert sur la possibilité de fonder les sciences grâce à la formalisation.

⁸On dit que deux ensembles ont même puissance lorsqu'on peut établir entre leurs éléments une correspondance bi-univoque (c'est-à-dire qu'à tout élément de l'un correspond un et un seul élément de l'autre, et réciproquement). Pour des ensembles finis, avoir même puissance se ramène à avoir le même nombre d'éléments. Pour des ensembles infinis, la plus faible puissance est celle du dénombrable (la suite indéfinie des nombres naturels). La puissance du continu (celle, par exemple des points d'une ligne ou de l'ensemble des nombres réels) est supérieure à celle du dénombrable. Enfin, on peut toujours construire un ensemble dont la puissance surpassé celle d'un ensemble quelconque donné.

⁹cité dans [Cavaillès, 1938][p. 136].

1.1.5. Axiomatiser l'arithmétique : cheminement par la logique ?

Le coup de tonnerre va retentir par contre à propos du deuxième problème de Hilbert au sujet “de la non-contradiction des axiomes de l’Arithmétique”. Dans ce cadre, les premiers résultats de Gödel sont plutôt encourageants. Le 6 juillet 1929, lors de la soutenance de sa thèse, il répond de façon positive à une question de Hilbert et Ackerman datant de 1928 : “Ce que l’on peut établir en logique par le moyen de raisonnements codifiés, par exemple dans le système formel des *Principia mathematica* de Whitehead et Russell, donne-t-il complètement tout ce qui est vrai en logique ?” Cette réponse constitue ce que l’on appelle le théorème de complétude du calcul des prédicats du premier ordre. Ainsi, dans ce cadre particulier, la notion de syntaxe¹⁰ et celle de sémantique se confondent.

À partir de là, Gödel a pu s’attaquer au problème que posait Hilbert sur les relations entre logique et arithmétique et en particulier sur la consistance de l’arithmétique. Gödel obtient plusieurs résultats sur ce sujet en 1930, exposés dans un article fameux [Gödel, 1931] et dont l’essentiel s’énonce sous la forme de deux théorèmes d’incomplétude. Le premier montre que toute formalisation de l’arithmétique et des théories plus puissantes que l’arithmétique, telle la théorie des ensembles, est nécessairement contradictoire ou incomplète. Le second affirme que l’arithmétique, dans un système formel donné, ne permet pas de démontrer que ce système formel est non contradictoire (Théorème XI). Ainsi, jamais une théorie intéressante, c'est-à-dire au moins aussi puissante que l’arithmétique, ne pourra démontrer d’elle-même qu’elle ne conduit pas à une contradiction. Pour conclure, on peut dire que le résultat de 1929 permet d'affirmer que la logique peut se réduire à de la syntaxe, alors que les résultats de 1930 établissent que l’arithmétique et tout système formel plus “puissant” ne peuvent pas être ramenés à de la syntaxe.

Les deux théorèmes d’incomplétude de Gödel mettent fin à l’espoir de Hilbert de ramener les mathématiques de l’infini aux mathématiques du fini dans le cadre d’un système formel particulier et unique. “Je tiens à souligner que le théorème XI [la consistance du système formel P n’est pas prouvable dans P] ne contredit pas le point de vue formaliste de Hilbert. En effet, ce point de vue ne suppose que l’existence d’une preuve de consistance dans laquelle seuls des moyens finis seraient employés et il serait concevable qu’il existe des preuves finies qui ne puissent être montrées dans le formalisme de P ” [Gödel, 1931][p. 198]. Le but était d’établir la consistance des mathématiques par des moyens mathématiques suffisamment élémentaires pour échapper au doute et en particulier plus pauvres que ceux qui sont admis en général en mathématiques. Gödel a démontré que cela n’était pas possi-

¹⁰Dans ce cadre, la syntaxe est l’ensemble des suites finies (voire infinies dénombrables) de symboles bien formées avec un ensemble fini de symboles de départ et un ensemble fini de règles qui autorisent la construction de certaines suites. La sémantique est l’ensemble des propositions vraies ou fausses, c'est-à-dire des suites de symboles bien formées qui peuvent être déduites des propositions initiales vraies ou fausses, avec un nombre fini d’inférences formelles.

ble. Ce résultat technique apporte également des éléments de réponse à une question plus générale et méthodologique que posait Hilbert. “Cet axiome de la possibilité de résoudre tout problème, est-ce une propriété caractéristique et distinctive de la pensée mathématique, ou serait-ce peut-être une loi générale du mode d’existence de notre entendement, à savoir que toutes les questions que se pose notre entendement soient susceptibles d’être résolues par lui ?” [Hilbert, 1990][p. 11]. Après Gödel, on ne peut plus se poser ces questions dans les mêmes termes. L’axiome que pose Hilbert de la possibilité de résoudre tout problème n’est plus aussi intuitif. Gödel a même mis en évidence un cadre formel particulier dans lequel le point de vue de Hilbert ne peut pas être défendu.

Ces résultats sont tout à fait fondamentaux pour comprendre dans quel état d’esprit se sont retrouvés des chercheurs comme von Neumann ou Turing. Le premier avait participé de toutes ses forces à la quête de formalisation générale des sciences sur la base axiomatique, en commençant par les mathématiques et la physique. Le second faisait partie d’une autre génération. Il était trop jeune pour avoir participé à la recherche systématique de formalisation qu’avait proposée Hilbert. Il était préoccupé par la Logique elle-même sous tous ses aspects. Entre les deux, les problèmes s’étaient déplacés de l’axiomatique vers la logique. Et les travaux sur la logique les entraîneront vers le calcul. Pour Turing les résultats de Gödel n’ont pas eu les mêmes conséquences sur sa conception des fondements des sciences. Tous ces résultats ont contribué à reconsidérer complètement le rôle que l’on pouvait attribuer à l’axiomatique comme méthode de fondement.

1.1.6. Les mathématiques du fini pour elles-mêmes : mécaniser les calculs

Le programme de Hilbert a continué cependant, plus ou moins directement, à influencer les relations entre les mathématiques et les autres sciences. Von Neumann aussi poursuit son travail dans l’optique du programme de Hilbert. Il s’attaque en particulier au cinquième problème “de la notion des groupes continus de transformations de Lie, en faisant abstraction de l’hypothèse que les fonctions définissant les groupes sont susceptibles de différentiation”. On considère que von Neumann a quasiment résolu ce problème.

Gödel avait montré que l’on ne pouvait pas réduire les mathématiques de l’infini aux mathématiques du fini. Il fallait alors s’intéresser aux mathématiques du fini pour elles-mêmes. Certaines considérations, liées au dixième problème, à propos “de la possibilité de résoudre une équation de Diophante”, ont alors suscité un vif intérêt. Hilbert demande dans ce problème de “trouver une méthode par laquelle, au moyen d’un nombre fini d’opérations, on pourra distinguer si l’équation est résoluble en nombres entiers rationnels” [Hilbert, 1990][p. 30]¹¹. On trouve ici l’étude de

¹¹Le fameux problème de la décision ou *Entscheidungsproblem* a été formulé dans le livre de 1928 écrit par Hilbert et Ackermann [Hilbert et Ackermann, 1950][p. 112 §12]. Ce problème est, à

problèmes discrets et finis, dont les solutions sont calculables en un temps fini avec un nombre fini d'opérations élémentaires. Ce type de problèmes va amener des mathématiciens comme Turing ou von Neumann à s'intéresser de façon systématique à la notion de calculabilité et aux automates comme machines permettant de faire des calculs "mécaniquement". Hilbert n'avait pas défini précisément ce qu'il entendait par son concept de calculabilité.

Turing va en proposer une définition précise, à la même époque et indépendamment de Post et Church. Après la théorie des ensembles, c'est la logique elle-même que l'on a mise en cause. Pour cela il fallait travailler sur des symboles vides de sens et sur leurs relations. La notion de calculabilité devait être précisée. "Les incertitudes de l'intuition intellectuelle conduisent à répudier [l'évidence des enchaînements logiques] à son tour et à remplacer le raisonnement pensé ou même parlé par un calcul sur des signes, étalés sur une feuille devant le regard" [Blanché, 1990][p. 62].

1.1.7. L'axiomatisation selon von Neumann

Von Neumann obtient un poste de Visiting Professor à l'université de Princeton. À sa création, l'Institute for Advanced Study à Princeton le recrute comme professeur de mathématiques¹². Von Neumann s'installe définitivement à Princeton, pour fuir le nazisme. Il va alors développer ses travaux dans la direction du calcul fini que préconisait Hilbert. C'est à cette époque qu'il éprouve, selon les termes de Ulam, "une véritable fascination pour le problème de la turbulence hydrodynamique" [Ulam, 1958][p. 7]. Il cherche à convaincre les physiciens de cette discipline de l'intérêt des machines à calculer très performantes qui peuvent fournir des solutions numériques à certaines des équations non-linéaires utilisées. Selon Ulam, il pensait que "les travaux numériques étaient la voie la plus prometteuse pour obtenir une intuition du comportement des systèmes [décris par des équations non-linéaires]. Cela le poussa, au départ, à étudier les nouvelles possibilités de calcul sur des machines électroniques" [Ulam, 1958][p. 8].

Von Neumann se montre ouvert aux autres disciplines ; il ne cherche pas pour autant à les réduire à des mathématiques. Selon lui, la méthode axiomatique réductionniste de Hilbert a montré ses limites avec les résultats négatifs de Gödel. Désormais le but est de construire et d'étudier les formalismes. On ne cherche plus toutes les justifications dans les formalismes, ce qui n'est possible que lorsque l'on a

l'époque, un de ceux qui trouvent leur origine dans le problème 10 que Hilbert avait posé en 1900. Mais on ne retrouve pas ce lien explicitement dans [Hilbert et Ackermann, 1950]. Il est à noter que le problème 10 est le seul parmi les 23 qui se pose en termes purement discrets et finis.

¹²Von Neumann fut choisi avec 5 autres professeurs de mathématiques de niveau mondial. Le centre et l'université, situés dans un même bâtiment, "constituaient une des plus grandes concentrations de figures dominantes en mathématiques et en physique que le monde ait jamais connue" [Goldstine, 1993][p. 170]. Princeton prend ainsi la place que Göttingen avait occupée dans les décennies précédentes.

déjà construit le formalisme adéquat. Le témoignage que rapporte Eugène Wigner est à cet égard tout à fait éloquent. Il était présent alors que von Neumann reçut une lettre de Gödel expliquant ses résultats d'incomplétude. À cette occasion von Neumann dit quelque chose comme : “cela change tout mon travail”¹³. Désormais, pour von Neumann, une séparation nette se fait entre les mathématiques et les autres disciplines. Les mathématiques et le calcul peuvent servir à résoudre des problèmes déjà formalisés. D'un autre côté les diverses disciplines sont l'occasion de construire des abstractions particulières qui fourniront des problèmes mathématiques intéressants.

La méthode axiomatique n'est pas abandonnée dans l'esprit de von Neumann, mais elle n'a plus le même statut. Elle reste pertinente pour fonder et formaliser une science, comme il a pu le constater avec les succès obtenus en mécanique quantique. La méthode ne permet plus de prétendre à une réduction de cette science à la logique, ni une réduction de l'ensemble des phénomènes observables au formalisme construit. L'axiomatique permet d'expliciter une problématique dans un formalisme déterminé. Elle n'assure pas pour autant la nécessité de ce formalisme particulier. Pour que ce formalisme soit pertinent, il faut qu'il soit le résultat d'une abstraction qui convienne au phénomène étudié. “Le bienfait de la méthode axiomatique n'est pas d'exclure l'intuition, mais de la contenir et de la refouler sur le terrain étroit où elle est irremplaçable” [Blanché, 1990][p. 90].

C'est désormais dans ce rôle que cherchera à se cantonner von Neumann. Il sera le formalisateur, en permettant simplement aux abstractions de prendre la forme d'un système axiomatique, dont il garantira la consistance et dont il cherchera à déduire les résultats mathématiques intéressants. “Contrairement à la plupart des mathématiciens appliqués qui veulent essentiellement manipuler des équations fournies par un physicien, von Neumann revenait au phénomène initial pour reconsiderer tout à la fois les idéalisations qui avaient été faites et la formulation mathématique” [Goldstine, 1993][p. 175]. Il ne cherchera plus cependant à vérifier la complétude des axiomatiques proposées. Quoi que l'on fasse désormais, il y aura dans les systèmes axiomatiques soit un peu trop, et certaines conséquences ne seront pas observables dans les phénomènes, soit trop peu, et l'axiomatique ne rendra pas compte de certains aspects des phénomènes. L'axiomatique a bien changé de visage. “C'est guidé par les phénomènes naturels que le mathématicien, en fin de compte, choisit les axiomes qui donneront naissance à une théorie efficace” [Cartan, 1943][p. 11]¹⁴.

¹³rapporté par [Heims, 1980] et cité dans [Dupuy, 1994].

¹⁴cité dans [Ullmo, 1969][p. 304].

1.2. Obtenir une définition opératoire de la calculabilité : von Neumann et la machine de Turing

1.2.1. La machine de Turing, la calculabilité et les systèmes formels

À la même époque, les résultats de Turing, en 1936¹⁵, parallèlement à ceux de Church et Post, vont permettre progressivement l'élaboration de machines permettant de manipuler des systèmes formels de façon automatique. Les machines permettant de réaliser des calculs existaient déjà depuis longtemps. Turing propose une machine abstraite avec une structure particulière, qu'il décrit dans son article de 1936 [Turing, 1936]¹⁶. Dans son article de 1948 [Turing, 1948], il appelle ce type de machines Logical Computing Machines ou LCM, machines calculant de façon logique.

Pour cela il propose un dispositif constitué d'une *machine* susceptible de se trouver dans un nombre fini d'états (internes) q_1, q_2, \dots, q_r distincts et appelés ses *m-configurations*. La machine est alimentée avec une *bande* unidimensionnelle, infinie à gauche et à droite, divisée régulièrement en sections, appelées *cases*, dans chacune desquelles peut être inscrit un *symbole*. Dans la case r est inscrit le symbole $S(r)$. À chaque instant, une seule case se trouve dans la machine, c'est la *case inspectée*, dans laquelle est inscrit le *symbole inspecté*. La machine peut tout de même garder une trace de certains symboles inspectés auparavant en modifiant sa *m-configuration* q_n . La liste des comportements possibles est entièrement déterminée par sa *m-configuration* q_n et par le symbole inspecté $S(r)$. Le couple $(q_n, S(r))$ est appelé la *configuration* de la machine. Deux familles de symboles sont utilisées dont les éléments peuvent être inscrits sur la bande. La première famille peut être constituée par exemple des symboles 0 et 1 pour l'écriture des nombres réels en représentation binaire. La deuxième famille, composée d'un nombre fini de symboles, sert de marquage.

On trouve quatre cas de figure pour l'évolution possible de la machine. Ce peut être soit l'inscription d'un symbole sur la case inspectée, si celle-ci est blanche, soit l'effacement du symbole inspecté, la case inspectée devenant blanche, ceci n'étant possible que pour les symboles de la deuxième famille, ceux de la première ne pouvant être effacés, soit encore le changement de case inspectée par déplacement de la bande d'une case vers la droite ou vers la gauche, soit enfin le passage à une autre *m-configuration*. Une table permet de régir le comportement de cet ensemble, c'est-à-dire le passage d'une configuration à une autre. Si le comportement d'une machine est entièrement déterminé par sa seule configuration, Turing parle

¹⁵Turing était anglais, mais il fut étudiant à l'université de Princeton de 1936 à 1938 [Goldstine, 1993][p. 174]. Dès 1935 il avait publié un article qui révisait certains aspects mathématiques d'un article de von Neumann. Celui-ci lui proposa de devenir son assistant mais Turing préféra repartir pour Cambridge.

¹⁶Nous avons essentiellement utilisé pour notre travail la traduction en français par Julien Basch publiée aux éditions du Seuil. Nous en reprenons ici, librement, de larges extraits.

de *machine automatique* ou *a-machine*. Turing évoque le fait que d'autres machines peuvent avoir un comportement qui ne dépend que partiellement de leur configuration. “Un opérateur extérieur doit [alors] intervenir et faire un choix arbitraire pour que la machine puisse continuer son travail” [Turing, 1936][§2]. Turing appelle une telle machine machine à choix ou *c-machine*. “Nous aurions besoin de ces machines pour travailler avec des systèmes axiomatiques” [Turing, 1936][§2]. Par la suite, Turing ne parle plus que des *a-machines* et les appelle simplement machines. Turing distingue également machines cycliques et machines acycliques. Les machines acycliques sont celles qui écrivent un nombre infini dénombrable de symboles de la première famille. Une séquence, c'est-à-dire un ensemble dénombrable de symboles, d'après la description même du mécanisme, est alors dite *calculable* s'il existe une machine acyclique qui la calcule. Un nombre est dit *calculable* s'il existe une machine acyclique qui calcule sa partie décimale.

Il faut chercher plus loin dans cet article l'intérêt selon Turing de sa machine. Turing parle de “machine à calculer universelle” [Turing, 1936][§6]. La machine universelle dont il parle est liée à la spécification d'une table particulière de comportement qu'il explicite. Il exprime cette table dans les termes des modifications élémentaires possibles des configurations de sa structure [Turing, 1936][§4 et §5]. Cette table permet de réaliser comme opérations élémentaires celles de la logique des Principia mathematica, ainsi que des opérations de recopiage et d'effaçage. Des opérations “minimales” peuvent alors être choisies, par exemple, la *négation*, le *ou*, le *et* et l'*identité*. À partir de ces opérateurs logiques, on peut retrouver l'ensemble des opérateurs de la logique classique.

Il propose ensuite un format standard dans lequel il exprime la table précédente. L'expression d'une table dans ce format standard est appelée la description standard (DS) de la machine. La description standard d'une machine, qui est une suite de lettres, peut ensuite être transformée sous forme numérique. “L'entier que cette suite représente sera appelé le nombre descriptif (ND) de la machine. Le ND détermine la DS et la structure de la machine de manière univoque, et on notera M(*n*) la machine dont le ND est *n*” [Turing, 1936][§5].

L'intérêt de cette machine pour Turing est de permettre la conception d'autres machines. “Il est possible d'inventer une unique machine U utilisable pour calculer n'importe quelle séquence calculable. Si on fournit à cette machine une bande au début de laquelle est inscrite la description standard DS d'une machine à calculer quelconque M, la machine U calcule alors la séquence, que M calculerait si elle était construite.” [Turing, 1936][§6]. Turing répète la même chose en d'autres termes dans son article de 1948. “Il est possible de décrire les LCM de façon tout à fait standard, et de donner la description sous une forme qui peut être “comprise” (c'est-à-dire appliquée) par une machine spécifique. En particulier, il est possible de concevoir une “machine universelle”, qui est une LCM, de telle sorte que si la description d'une autre LCM lui est imposée sur son ruban blanc de l'extérieur, et que la machine

(universelle) fonctionne, elle va réaliser les opérations de la machine particulière dont on lui a donné la représentation”[Turing, 1948][p. 6-7].

Il faut alors bien comprendre pourquoi Turing parle de machine universelle. Sa machine ne permet pas de tout faire, mais elle présente simplement une structure standard. Avec les machines de Turing on obtient une certaine universalité. L'universalité de cette machine consiste à pouvoir donner un sens précis à la notion de formalisme. On appelle formalisme un système qui peut être appliqué sur la machine de Turing. “Grâce au travail de A. M. Turing, une définition précise et indubitablement adéquate de la notion générale de système formel peut désormais être donnée”¹⁷. Il ajoute en note : “à mon sens le terme de “système formel” ou de “formalisme” ne devrait jamais être utilisé pour autre chose que cette notion. Dans une lecture à Princeton, j'ai suggéré certaines généralisations transfinies des formalismes ; mais il s'agissait de quelque chose de radicalement différent des systèmes formels dans le sens propre du terme, dont la propriété caractéristique est que raisonner dans leurs termes peut en principe être complètement remplacé par des dispositifs automatiques”. Ainsi la calculabilité proposée par Turing permet de donner un sens précis à la notion de système formel.

1.2.2. Généricité des systèmes formels et machines à réaliser des calculs

En fait on savait déjà construire des machines à calculer, mais elles étaient presque toujours dédiées à un problème particulier. La réussite de Turing est d'avoir proposé une structure qui permet de construire toute machine manipulant un formalisme fini, quel que soit ce formalisme. L'innovation de Turing constituait un changement conceptuel dans la notion de machine à réaliser des calculs. Von Neumann retrouvera ce changement conceptuel et lui associera l'élément qui permet de réaliser les traductions entre systèmes formels. Goldstine rapporte que selon lui “von Neumann fut le premier [...] à comprendre explicitement qu'un ordinateur réalise essentiellement des fonctions logiques, et que les aspects électriques sont secondaires. [...] Avant von Neumann, [...] les gens se concentraient essentiellement sur les aspects d'ingénierie électrique. Ces aspects étaient bien sûr d'importance capitale, mais c'est von Neumann qui, le premier, fournit un traitement logique du sujet, comme s'il s'agissait d'une branche conventionnelle de la logique ou des mathématiques” [Goldstine, 1993][p. 192].

Avec Turing on ne sait cependant pas comment on trouve la traduction qui permet de passer de la table d'un certain système formel à celle d'un autre système formel. Même si la machine de Turing reste une machine abstraite, on pourrait penser retrouver là sous l'angle de la technique une proposition de Wittgenstein : “Tout langage de signes correct doit pouvoir se traduire dans tout autre langage de

¹⁷Citation de Gödel, dans une note ajoutée le 28 août 1963 à son article de 1931 et que l'on peut trouver page 195 du premier tome de ses *Collected papers*.

ce genre”(3.031)[Wittgenstein, 1986]. Cependant ni Wittgenstein ni Turing ne nous disent comment on passe d’un système formel à un autre. Autrement dit on ne sait pas comment il est possible d’exprimer en termes finis de U la description standard DS d’un système formel quelconque. On ne sait pas non plus comment il est possible de passer du formalisme d’une machine $M(n)$ à celui d’une machine $M(m)$.

Auparavant, les machines étaient construites et dédiées à la résolution d’un problème particulier. Elles étaient ce que Goldstine nomme des “special purpose machines”. Le mérite de von Neumann¹⁸ sera de comprendre que les résultats de Turing permettent de traiter un grand nombre de problèmes différents sur la même machine, et qu’il suffit de les traduire dans un formalisme choisi pour cette machine. Ainsi, en même temps, von Neumann résout partiellement le problème de la traduction. Avec ces nouvelles machines, “tout ordinateur qui doit résoudre un problème mathématique complexe doit être <<programmé>> pour cette tâche. Cela veut dire que l’opération complexe consistant à résoudre ce problème doit être remplacée par une combinaison des opérations de base de la machine” [von Neumann, 1992] [p. 17]. Les mêmes machines sont alors capables de résoudre des problèmes différents. Il suffit pour cela d’exprimer leur solution comme combinaison d’opérations élémentaires de la machine effectivement choisie pour être construite. Ainsi, la résolution de problèmes numériques qui était auparavant le travail d’ingénieurs capables de construire les machines adaptées devient un travail “d’informaticien” transcrivant la méthode de résolution du problème dans le langage de la machine.

De cette façon, von Neumann fait la distinction entre la réalisation matérielle de la machine et les “ordres” que l’on peut donner à la machine. Il présente donc ces machines d’une manière qui possède tous les aspects extérieurs d’une axiomatisation par son caractère formel, détaché du substrat, et par son caractère fondateur. La structure de ces ordres devient le programme extérieur. “Notons l’importante différence qu’il y a entre ce mode de contrôle et le contrôle par câblage décrit auparavant. Dans ce dernier, les points de contrôle étaient des objets physiques réels, et leurs connexions par câblage exprimaient la nature du problème. Dans le cas présent, les ordres sont des entités idéales, stockées dans la mémoire, et ce sont par conséquent les contenus de ce segment particulier de la mémoire qui expriment le problème. C’est pourquoi on appelle ce mode de contrôle <<contrôle à stock en mémoire>>” [von Neumann, 1992][p. 29]. Ainsi, en n’ayant plus à contrôler de façon directe la construction d’une nouvelle machine pour chaque nouveau problème, mais en permettant la réalisation dans une seule machine de différentes machines potentielles, la proposition de Turing, révisée et matérialisée par von Neumann, a constitué une véritable révolution.

¹⁸Goldstine [Goldstine, 1993][p. 184-203] explique bien que les innovations portant sur la conception de l’EDVAC sont une véritable œuvre collective et que le mérite de von Neumann est d’avoir permis sa cristallisation grâce à son travail de synthèse [von Neumann, 1945]. Il faut remarquer que ce rapport n’est pas publié dans les six tomes des œuvres complètes de von Neumann.

Turing parle de machines universelles, non pas parce qu'elles peuvent tout faire, mais parce que leur structure est susceptible de correspondre à la matérialisation après traduction de tous les problèmes exprimés dans un langage formel. Turing donne d'ailleurs un exemple de nombre réel définissable et non calculable. "Un nombre définissable n'est cependant pas forcément un nombre calculable, et je présenterai dans la section 8 un nombre définissable qui n'est pas calculable. L'ensemble des nombres calculables est immense, et comparable à bien des égards à celui des réels, mais il est pourtant dénombrable" [Turing, 1936][première page de l'article]. L'ensemble des nombres réels est non-dénombrable, et l'ensemble des nombres réels calculables est dénombrable. Donc pour toute machine avec sa DS, il existe un nombre réel qui n'est pas calculable. Ce premier raisonnement est un raisonnement par l'absurde.

Turing propose également une démonstration directe de la non calculabilité de certains nombres réels. Pour cela il exhibe un nombre particulier définissable et non calculable. Turing utilise pour cela le procédé de la diagonale¹⁹, pour un essai de démonstration de la non-dénombrabilité des nombres calculables. La preuve est nécessairement fausse puisque Turing a montré que les nombres calculables sont en nombre dénombrable. Dans sa preuve de l'erreur présente dans la démonstration s'appuyant sur le procédé diagonal, il montre l'impossibilité de construire une machine particulière qui fournirait un certain nombre réel β . "L'erreur de ce raisonnement réside dans l'affirmation que β est calculable. Cela serait vrai si nous pouvions énumérer les séquences calculables avec des moyens finis, mais ce problème d'énumération des séquences calculables est équivalent à celui qui consiste à déterminer si un nombre donné est le ND d'une machine acyclique, et il n'existe pas de procédure générale pour faire cela en un nombre fini d'étapes" [Turing, 1936][§8].

Après von Neumann, cette propriété d'universalité permet également de résoudre tout problème posé en termes formels avec une seule et même machine. "Dans ce cas, puisque les ordres qui réalisent la fonction de contrôle tout entière sont dans la mémoire, on obtient un plus grand degré de flexibilité que dans tout autre mode de contrôle envisagé précédemment. [...] Il en résulte que toutes sortes de systèmes d'ordres deviennent possibles, qui se modifient sans cesse et qui modifient par conséquent aussi les processus de calcul qui sont également sous leur contrôle. Ainsi des processus²⁰ plus complexes que de simples itérations deviennent possibles" [von Neumann, 1992][p. 29]. D'après les résultats de Turing, cette universalité est nécessairement partielle. Encore faut-il traduire le problème dans les termes de la machine. Aucun système de traduction automatique n'est proposé. La structure

¹⁹Procédé utilisé par Cantor pour montrer la non-dénombrabilité des nombres réels vers 1883-84. Ce procédé avait été introduit et employé de façon intuitive et presque implicite par Du Bois Reymond en 1876. [Cavaillès, 1938][p. 42 et p. 72].

²⁰Ici, le terme de processus est utilisé dans le sens de succession de combinaisons d'opérations élémentaires.

proposée par von Neumann permet cependant de traduire partiellement ou approximativement le problème formel dans les termes de la machine. Cette traduction, c'est le programmeur, plus tard avec l'aide du compilateur, qui la réalise.

1.2.3. Le problème du choix des opérations élémentaires

Un problème formel s'exprime en fonction des opérations formelles élémentaires du système formel dans lequel il s'exprime. Cependant von Neumann précise que, selon les opérations élémentaires de la machine, le problème s'exprimera dans les termes de la machine de façons très différentes. Il explique que même si les machines manipulant les opérations de l'arithmétique et les nombres entiers exprimés en binaire sont les plus nombreuses et permettent sans doute d'exprimer le plus grand nombre de problèmes, elles ne peuvent pas forcément traduire ces problèmes de la "meilleure" façon. Ainsi en est-il de certains problèmes d'équations aux dérivées partielles, pour la solution desquels, des machines capables de réaliser une intégration en un seul pas de temps seront éventuellement plus adéquates. "Pour une classe donnée de problèmes, un ensemble d'opérations de base peut être plus efficace, c'est-à-dire peut permettre l'usage de combinaisons plus simples, moins nombreuses qu'un autre ensemble" [von Neumann, 1992][p. 17].

On voit bien ici les limites qu'il faut impérativement préciser lorsque l'on parle d'équivalence entre des machines ou d'universalité de certaines machines. Pour donner un exemple simple, une machine pouvant utiliser comme opérations élémentaires les opérations arithmétiques devra réaliser un nombre infini d'opérations pour effectuer un calcul d'intégration. De plus, il faudra préciser le schéma d'intégration, par exemple comme somme d'une série de Riemann, c'est-à-dire comme une somme infinie. Au contraire, certaines machines possédant l'intégration comme opération élémentaire pourront réaliser ce calcul d'intégration avec une seule opération élémentaire. Le passage d'un formalisme à un autre n'est pas nécessairement simple. Ces deux machines sont toutes deux simulables sur la machine de Turing, c'est-à-dire qu'il est possible de donner une description standard des deux machines dans les termes de la machine de Turing. Leurs opérations élémentaires n'étant pas identiques, elles sont pourtant très différentes pour un mathématicien qui se pose le problème selon une problématique finie. La première machine n'est pas capable de donner la valeur exacte en un nombre fini d'opérations. La deuxième réalise au contraire l'intégration en un seul pas de temps. Il n'est pas du tout évident en fait de traduire le premier système formel dans les termes du deuxième ou le contraire, mais ce n'était pas le problème qui intéressait Turing.

Von Neumann avait parfaitement conscience de ce mouvement et de la modification progressive des questions pertinentes sur les automates et leur relation à la logique. Il l'explique au Hixon Symposium : "à travers toute la logique moderne, la seule chose qui importe est de savoir si un résultat peut être obtenu en un nom-

bre fini de pas élémentaires ou non. La quantité du nombre de pas nécessaires, de l'autre côté, n'est presque jamais un souci de la logique formelle. [...] En s'occupant des automates, ce constat doit être sérieusement modifié. Dans le cas d'un automate, ce qui importe n'est pas de savoir s'il peut atteindre en fait un certain résultat en un nombre fini de pas, mais aussi combien de tels pas sont nécessaires" [von Neumann, 1951][p. 303]. Le fait de pouvoir exprimer deux automates dans la même structure ne nous assure pas qu'ils ont un comportement identique.

1.2.4. Les difficultés liées à la matérialisation des opérations élémentaires

Ces réflexions ont été faites en tenant compte simplement des opérations élémentaires abstraites autorisées. Elles ne prennent pas en considération les difficultés liées à la matérialisation de ces opérations élémentaires. On pourrait essayer d'argumenter que les opérations logiques sont beaucoup plus simples à matérialiser que les opérations arithmétiques ou algébriques. L'argumentation serait cependant très difficile à étayer. Elle serait sans doute correcte dans certains cas, mais pas en toute généralité, ce qui est le cadre dans lequel on se place habituellement. En effet, il est tout aussi difficile de rendre compte de la somme de la charge de deux condensateurs que de l'intégrale d'un courant dans un circuit intégrateur. La difficulté porte sur le fait que la matérialisation n'a rien à voir avec aucune des opérations, qu'elles soient logiques, arithmétiques ou algébriques. Le problème est d'un autre ordre²¹. Finalement, on ne voit pas pourquoi les opérations algébriques ou arithmétiques seraient plus difficiles à réaliser que les opérations logiques. L'intérêt de considérer des opérations logiques vient donc de leur plus grande généralité. Ainsi le choix d'opérations élémentaires reste aujourd'hui encore un choix fondamental, malgré le développement de la puissance des ordinateurs. On oublie un peu vite aujourd'hui les limitations que les fondateurs connaissaient quand ils parlaient de "machines universelles". "Mais l'affirmation totalitaire du déterminisme universel, comme toute prise de position sur le tout, sort du cadre de la méthode scientifique (c'en est un nouvel exemple après celui de l'entropie universelle) ; elle est étrangère à la science et elle lui est inutile. C'est proprement une hypothèse métaphysique" [Ullmo, 1969][p. 181]. C'est un mouvement du même type que l'on a pu observer entre la conception de la machine universelle due à Turing et les interprétations postérieures qu'on a pu en faire. Le jeu était d'autant plus tentant que Turing lui-même avait utilisé le terme de machine universelle.

²¹Les phénomènes n'ont pas besoin d'observateur pour apparaître. Les opérations sont quant à elles des abstractions humaines. Les deux ne sont donc pas nécessairement de même nature. L'affirmer tout autant que prétendre son contraire seraient défendre une position métaphysique.

1.2.5. Pourquoi privilégier la construction de machines digitales

Von Neumann attribue une autre supériorité aux machines digitales par rapport aux machines analogiques. On trouve des explications sur ce sujet dans [von Neumann, 1992]. Mais les précisions nécessaires au raisonnement se trouvent en fait dans la conférence qu'il a tenue au Hixon Symposium [von Neumann, 1951] [p. 292-296]. La première comparaison porte sur la précision de ces machines. On sait construire des machines digitales avec une précision aussi grande qu'on le souhaite, ce qui n'est pas le cas des machines analogiques. Le second critère porte sur le bruit relatif, lié aux variations du signal effectif par rapport au signal désiré. "Comme nous l'avons fait remarquer ci-dessus, le niveau de bruit relatif d'une machine analogique n'est jamais inférieur à 1 pour 10^5 , et s'élève dans de nombreux cas à 1 pour 10^2 . Dans la machine digitale décimale à 10 chiffres décrite ci-dessus, le niveau de bruit relatif (lié aux arrondis) est 1 pour 10^{10} . Ainsi, l'importance réelle de la procédure digitale tient dans sa capacité à réduire le niveau de bruit computationnel à une quantité qui est complètement inaccessible par une autre procédure (analogique)" [von Neumann, 1951][p. 295]. Ceci ne supprime pas la difficulté qui consiste à choisir correctement les opérations élémentaires pertinentes pour le problème particulier étudié.

1.2.6. La trajectoire de von Neumann dans le sillage du programme de Hilbert

L'essentiel ici était de montrer brièvement que le parcours de von Neumann est celui d'un mathématicien formé par le programme de Hilbert. Il a cherché à le mettre en œuvre de façon quasi-systématique, et en gardant le même état d'esprit. Cette réalisation a d'abord eu lieu en théorie des ensembles, en logique, en physique et devait se poursuivre éventuellement dans les autres disciplines. Après les résultats négatifs de Gödel, la voie la plus prometteuse était celle du calcul. Dans l'esprit de von Neumann, elle doit tenir compte de la matérialisation qui permet le calcul. La théorie des automates offre des potentialités nouvelles. Tous les automates ne sont cependant pas identiques. Ils dépendent en particulier du choix de leurs opérations élémentaires.

2. LES ANNÉES QUARANTE ET L'UNIFICATION DE LA CONCEPTION DU FORMALISME (1943-1956)

La deuxième guerre mondiale, va être l'occasion, volontaire ou non, pour un certain nombre de scientifiques, de mettre leurs connaissances au service des militaires : ils appliquent alors les études théoriques de la décennie précédente à des cas concrets avec comme impératif l'efficacité des réalisations. Le travail est donc essentiellement celui d'ingénieurs qui mettent en pratique des connaissances théoriques. Des équipes nouvelles se constituent, répondant à d'autres contraintes que celles qui sont habituelles dans la communauté scientifique en temps de paix. Ces équipes s'appuient sur des sympathies évidentes mais mettent aussi en relation les personnalités les plus compétentes de chaque discipline pour mener à bien, ensemble, des projets d'envergure dont le but ultime est simple, la victoire des Alliés. Au cloisonnement académique se substituent des équipes pluridisciplinaires dont le travail est intégré, de la conception théorique à la réalisation pratique. La guerre est ainsi l'occasion de rencontres entre scientifiques d'horizons différents.

2.1. Les analogies fondatrices

2.1.1. La causalité circulaire et les mécanismes de rétroaction (feedback) comme concepts unificateurs

Pendant la guerre seront publiés plusieurs articles de Wiener et Rosenblueth, avec pour le plus important d'entre eux la collaboration de Bigelow. Wiener était professeur de mathématiques au MIT. Il avait développé une partie des mathématiques appliquées de la théorie du signal en l'enseignant à de futurs ingénieurs. Wiener, par l'ensemble de ses travaux scientifiques, est d'une importance essentielle. Sa personnalité est aussi universelle que celles de von Neumann ou McCulloch. Rosenblueth, collègue du physiologiste Walter Cannon à la Faculté de Médecine de Harvard, était biologiste de formation. Comme Wiener le précise dans *Cybernétique et société*, c'est à cette époque qu'il prend conscience avec Rosenblueth de "l'unité essentielle de l'ensemble des problèmes ayant trait à la communication, au contrôle et à la mécanique statistique, aussi bien dans la machine que chez l'être vivant"

[Wiener, 1952][p. 287]. La notion d'homéostasie et de stabilisation par rétroaction, c'est-à-dire de *feedback* négatif, est au cœur de cette prise de conscience.

Ce thème est également au centre des conférences Macy qui prennent pour titre officiel *Circular causal and feedback mechanisms in biological and social systems*. Dans l'article de 1943, *Behavior, purpose and teleology* [Wiener et al., 1943], Wiener, Rosenblueth et Bigelow s'interrogent sur la notion de comportement¹ et les classifications que l'on peut en faire en cherchant à éviter toute forme de mentalisme ou de psychologisme. Le but est de classifier les modifications des sorties d'une boîte noire. Une méthodologie se met donc en place pour étudier différents types de systèmes *a priori* très différents, comme les systèmes balistiques, la physiologie du corps humain ou les comportements liés à une activité de l'esprit.

La même année, le biologiste McCulloch et le mathématicien Pitts proposent un modèle formel du cerveau, dans leur article *A logical calculus of the ideas immanent in nervous activity* [McCulloch et Pitts, 1943]. Ils cherchent à expliquer, par l'intermédiaire d'un modèle formel et s'appuyant sur des faits neurobiologiques, les relations entre les capacités logiques et le fonctionnement cérébral. McCulloch et Pitts veulent rendre compte du système nerveux en termes logiques. McCulloch a eu un parcours intellectuel pour le moins sinueux. "Ayant un intérêt majeur pour la philosophie et les mathématiques, je suis arrivé en psychologie à cause du problème suivant : comment quelque chose comme les mathématiques peut-il apparaître – et alors de quelle sorte de chose s'agit-il ? Pour cette raison, je me suis tourné progressivement vers la psychologie, puis j'ai été amené à la neurophysiologie, parce que je ne parvenais toujours pas à trouver les variables pertinentes" [von Neumann, 1951][p. 319].

Depuis longtemps McCulloch, comme médecin et neuropsychiatre, réfléchissait sur les relations entre la logique et le fonctionnement du système nerveux. La question de la fondation neurophysiologique des nombres le hantait, comme le met en évidence son article *What is a number, that a man may know it, and a man that he may know a number?* [McCulloch, 1961]. La rencontre de Pitts lui permit de formaliser ses conceptions et de les présenter sous forme axiomatique [McCulloch et Pitts, 1943]. Il s'explique lui-même au Hixon Symposium, au cours de la discussion sur l'exposé de von Neumann : "la tentative de construire une théorie dans une discipline comme celle-là de sorte qu'elle puisse être soumise à vérification est ardue. Assez bizarrement j'ai commencé avec le mauvais point de vue, vers 1919, en essayant de construire une logique pour les verbes transitifs. Cela se transforma en un problème de logique modale, et ce n'est pas avant d'avoir vu l'article de

¹Les termes de comportement, pensée, intelligence, esprit, cerveau, systèmes nerveux sont utilisés à l'époque de la cybernétique sans avoir jamais été proprement définis, mais en évitant tout psychologisme. Ces termes seront cependant interchangeables dans certains contextes. En particulier on ne sait jamais très bien s'ils renvoient à des aspects physiques, chimiques, biologiques ou psychologiques.

Turing que je commençai à pouvoir m'en sortir et qu'avec l'aide de Pitts je formulai le calcul logique nécessaire. Nous pensions que ce que nous faisions (et je pense que nous avons pas mal réussi), c'était traiter le cerveau comme une machine de Turing ; c'est-à-dire comme un dispositif qui pourrait réaliser les fonctions qu'un cerveau doit réaliser, si ce n'est se tromper et avoir une psychose" [von Neumann, 1951][p. 319]².

On peut sans doute considérer que quatre pas successifs ont été franchis par McCulloch avant sa rencontre avec Pitts. Le premier consiste à assimiler les objectifs des travaux de Boole et à considérer que les pensées fonctionnent comme les nombres. Le deuxième est méthodologique et relève de la démarche de Hilbert avec une fondation axiomatique des connaissances scientifiques. Le troisième pas est celui que fit franchir Shannon, en montrant que l'algèbre de Boole peut être matérialisée. Le quatrième pas consiste à comprendre que l'universalité des machines de Turing permet de se contenter d'un fondement axiomatique avec des unités simples qui sont biologiquement plausibles. Pour McCulloch, cette unité biologique, c'est le neurone. Sa rencontre avec Pitts va lui permettre d'articuler correctement ces quatre éléments. Pour se convaincre de l'importance de la démarche "logique", c'est-à-dire axiomatique, dans cet article, il convient de rappeler les trois références que font les auteurs. Elles vont respectivement à Carnap, puis Hilbert et Ackermann [Hilbert et Ackermann, 1950], et enfin Whitehead et Russell. On peut noter à ce propos que lorsque Pitts, âgé de 18 ans, rencontra Russell à Chicago en 1938, il fut stupéfait par sa connaissance parfaite des *Principia mathematica*. Russell organisa une rencontre avec Carnap. Lors de leur première entrevue, Pitts fit remarquer à Carnap des erreurs qu'il avait observées dans son dernier livre.

2.1.2. L'analogie entre machines et organismes vivants

La première question qui se pose est de savoir pourquoi ces scientifiques ont décidé qu'ils allaient essayer de modéliser l'esprit. La réponse est loin d'être triviale. Pour eux, au moins dans un premier temps, cela signifie qu'il s'agit de modéliser le cerveau et son fonctionnement. Cette assimilation de l'étude de l'esprit à celle du cerveau est justifiée par McCulloch dans une communication au Hixon Symposium en 1948, au titre explicite : *Why the mind is in the head* [McCulloch, 1951].

Les motivations sont très loin d'être les mêmes pour les uns et les autres. L'origine de ces travaux est très certainement la prise de conscience d'une analogie entre les machines et les organismes vivants. C'est dans ce sens que l'on doit interpréter la citation de Wiener sur l'origine de la cybernétique³. Heims rapporte qu'"une nouvelle convergence d'intérêt entre von Neumann et Wiener est apparue [pendant la guerre], très différente de celle pour la théorie ergodique : l'exploration des possibilités d'analogie entre les dispositifs techniques, comme les ordinateurs et

²cité par Pignon [von Neumann, 1992].

³voir ci-dessus.

les systèmes nerveux, ou plus généralement les organismes vivants. Le travail de von Neumann dévoile une prise de conscience de telles analogies dans son *First draft of a report on the logical design of a prospective new computing machine* du 30 juin 1945 [von Neumann, 1945]. Dans ce “brouillon”, il emprunte un formalisme imaginé à l'origine pour décrire l'organisation logico-formelle du cerveau humain” [Heims, 1980][p. 182]⁴.

2.1.3. Les hésitations de von Neumann ...

Au-delà de ces analogies, ces motivations répondent sans doute en partie à un besoin métaphysique que la science est censée combler. On trouve chez McCulloch une interrogation sur la nature des nombres et des relations qu'ils entretiennent avec les hommes. Pour Wiener, les motivations sont moins claires. Il semble qu'au départ la force des analogies qu'il a décelées soit sa motivation principale. Il veut savoir jusqu'où elles peuvent le mener. Von Neumann était dubitatif sur l'intérêt pour les mathématiques d'étudier l'esprit et le cerveau. Comme l'écrit Heims, “au cours de l'année 1946, von Neumann avait émis des doutes sur la faisabilité de créer des modèles logico-mathématiques féconds, que ce soient des modèles de cerveaux d'organismes même très simples, ou des modèles des formes générales de la communication ou du contrôle mettant en jeu des systèmes nerveux” [Heims, 1980][p. 203].

Il hésite à l'époque à emprunter la voie proposée par Schrödinger dans son ouvrage de réflexion *What is life?*, et à étudier la vie et ses différents mécanismes, en particulier ceux de réplication. Schrödinger avait insisté dans son livre sur l'intérêt des travaux de Delbrück. Von Neumann est tenté par la voie ouverte par les biophysiciens comme Delbrück, avec ses expériences sur le bactériophage “des organismes vivants, qui, quoique primitifs, se reproduisent. La complexité d'un tel organisme semblait mathématiquement abordable, et les interprétations des observations pour les étudier en détail étaient disponibles, en particulier la diffraction par rayon X et la microscopie électronique” [Heims, 1980][p. 204]. “Dans sa lettre polémique, von Neumann met Wiener au défi de réfuter la conclusion selon laquelle le bactériophage est un sujet beaucoup plus prometteur pour la recherche mathématique que le système nerveux” [Heims, 1980][p. 204].

Il est clair que l'étude du cerveau, pour von Neumann, doit être essentiellement l'occasion de mettre en œuvre ses talents de mathématicien proche des préoccupations des expérimentateurs. Le but pour lui est de formaliser la discipline. C'est désormais ainsi qu'il conçoit son rôle de mathématicien. À chaque formalisation nouvelle qu'il a pu proposer est associée la création d'une nouvelle branche des mathématiques. L'étude du cerveau et des systèmes nerveux peut en être également l'occasion. La formalisation constitue peut-être également pour von Neumann la

⁴On trouve des passages de ce rapport et en particulier du vocabulaire et des images employées par von Neumann dans [Goldstine, 1993][p. 192-197].

meilleure façon de rendre publiques les théories et les conceptions éventuellement floues que l'on peut avoir d'un phénomène. Von Neumann ne cherchera pas vraiment à formaliser les disciplines liées à l'étude de l'esprit. Il se contentera d'aborder la formalisation d'une certaine classe d'automates, ceux que l'homme construit, et ainsi de mieux les comprendre. Dans son intervention au Hixon Symposium, il propose des lignes directrices pour la construction d'une formalisation logique des automates artificiels. Il a pris bien soin auparavant d'expliquer pourquoi les cerveaux, quoiqu'étant des automates par postulat, sont bien loin d'être semblables aux automates construits par l'homme. Aussi les directions que propose von Neumann ne doivent pas être comprises comme permettant de créer une formalisation adaptée au comportement du cerveau, mais concernent seulement les automates artificiels [von Neumann, 1951].

2.1.4. Les outils mathématiques et la méthodologie de la physique comme point commun

Au cours des conférences Macy se sont confrontées différentes conceptions sur le fonctionnement du cerveau. On peut considérer que l'enjeu à l'époque est la définition d'une approche commune pour étudier des phénomènes dont les similarités n'avaient pas été perçues auparavant. Cette approche s'appuie sur un certain nombre d'outils mathématiques nouveaux construits dans les années trente et que nous avons en partie évoqués précédemment. La question de fond, telle que posée par Wiener, est de savoir s'il est concevable que les outils mathématiques utilisés en physique permettent de formaliser l'étude de l'esprit et de la conscience. On constate en fait que le point commun entre les participants à ces conférences n'est pas le sujet d'étude, mais certains outils d'analyse. Comme le souligne Dupuy, "tous ont en commun un certain rapport à la science qui privilégie la modélisation mathématique" [Dupuy, 1985][p. 48].

Aussi l'enjeu de ces conférences est-il une véritable confrontation entre des conceptions au départ très diverses du rôle de la modélisation en sciences. Le point commun entre les participants aux conférences Macy était d'attribuer un rôle important aux formalismes dans les explications scientifiques. Au Hixon Symposium, le neurophysiologiste Lashley déclare : "ce qui nous réunit ici, c'est la conviction que nous partageons tous, je crois, qu'il est possible en dernière instance de décrire les phénomènes de l'esprit et du comportement au moyen des concepts des sciences mathématiques et physiques"⁵.

⁵Hixon Symposium [p. 112] cité par [Dupuy, 1994].

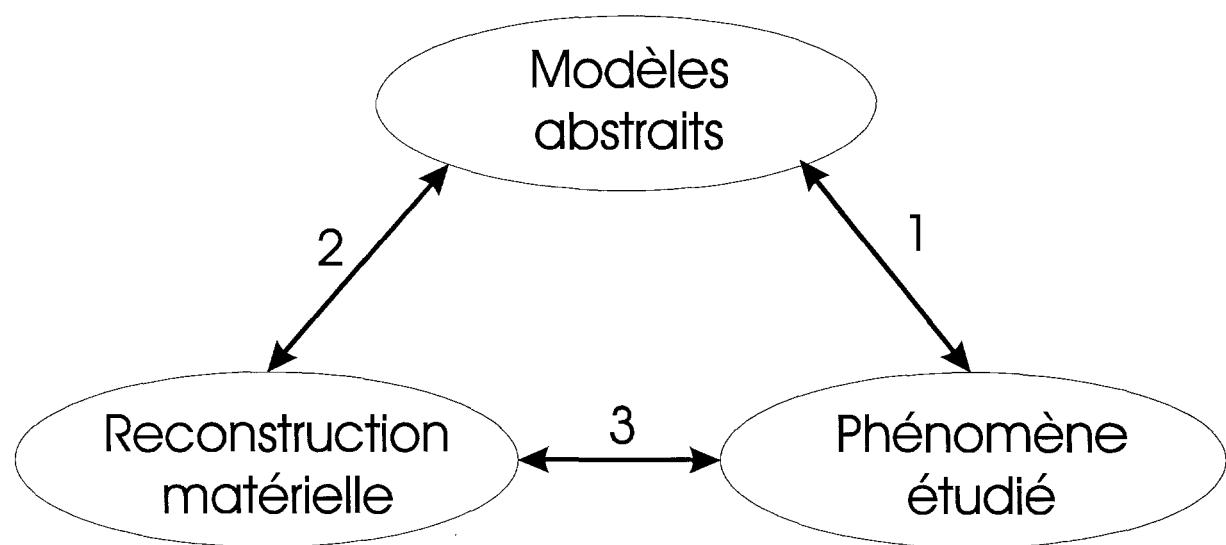


FIGURE 1 : *les trois analogies et les trois identifications*

2.2. L'analogie comme méthodologie scientifique

Au-delà de cette volonté commune d'appliquer les outils des sciences dites exactes, on peut dégager, au cours des années quarante, trois conceptions différentes du rôle de la formalisation. À l'époque les différents protagonistes avaient des positions diverses sur le rôle des outils provenant de la physique et des mathématiques. On peut distinguer trois approches. Pour les besoins de l'argumentation, les citations seront prises à l'un ou à l'autre, mais on constate en lisant attentivement les textes que les différentes positions ont pu être défendues par les uns et les autres selon le point de vue qu'ils choisissaient, celui du mathématicien, du logicien, de l'ingénieur, du physicien, voire du neurophysiologiste. Pour les plus brillants d'entre eux, le point de vue pouvait être pluridisciplinaire.

Ces trois approches sont à l'époque les différentes facettes du groupe des cybernéticiens. Elles s'appuient sur une triple analogie entre les trois termes de la figure 1 : analogie du phénomène étudié avec le modèle abstrait, analogie du modèle abstrait avec sa matérialisation par l'ordinateur, analogie⁶ du phénomène étudié avec sa reconstruction matérielle sur l'ordinateur. On discerne une position que l'on peut appeler “de mathématicien”, et deux que nous nommerons “d'ingénieurs”. Ces dénominations sont évidemment en partie arbitraires.

2.2.1. L'approche méthodologique du “mathématicien”

Le “mathématicien” s'inspire du biologique pour définir des théories mathématiques qu'il étudie pour elles-mêmes, avec un regard plus ou moins distrait dirigé vers les phénomènes auxquels cette théorie est censée correspondre. Il travaille à partir d'une analogie entre le phénomène étudié et le modèle abstrait. Le rôle du mathématicien dans ce cadre est de manipuler des symboles. “Le progrès réalisé par l'axiomatique consiste en une claire séparation de l'intuitif et du logique : d'après l'axiomatique, seuls les faits logiques et formels forment l'objet de la science mathématique, mais non l'élément intuitif qui peut s'y rattacher”⁷. C'est en partie l'attitude que suit von Neumann. À plusieurs reprises, au Hixon Symposium, comme dans l'introduction de son livre *The computer and the brain*, il se présente d'abord comme un mathématicien. “Comme je ne suis ni neurologue ni psychiatre, mais mathématicien, le travail qui suit demande quelques explications et justifications. Il s'agit d'une tentative pour comprendre le système nerveux du point de vue d'un mathématicien” [von Neumann, 1992][p. 13].

⁶À l'époque les différentes analogies jouent dans les deux sens. On observe un va-et-vient constant entre les deux termes de chaque analogie. Il serait possible d'en donner des exemples. Ainsi les deux termes de chaque analogie se renforcent mutuellement en empruntant le vocabulaire ou les représentations l'un à l'autre.

⁷Citation d'Einstein, tirée de *La géométrie et l'expérience* et rapportée dans [Blanché, 1990][p. 99].

Le propre du mathématicien lorsqu'il aborde une autre discipline n'est pas de pratiquer des mathématiques, mais de construire des mathématiques adaptées, de les fabriquer. Ainsi von Neumann va au-delà d'un simple travail sur un système formel. Il veut lui-même établir le lien entre l'intuitif et le formel. "En premier lieu, il est exagéré de décrire ce travail comme "une tentative pour comprendre". Ce n'est qu'un ensemble un peu organisé de spéculations sur la façon dont il faudrait procéder à une telle tentative. Autrement dit, j'essaie de deviner, dans le brouillard où nous sommes encore à leur propos, lesquelles des perspectives offertes par les mathématiciens semblent *a priori* prometteuses et lesquelles semblent ne pas l'être" [von Neumann, 1992][p. 13]. L'attitude n'est donc pas du tout celle du mathématicien qui plaque des outils qu'il maîtrise sur un phénomène qu'il ne connaît pas. Il s'agit bien au contraire de chercher à construire des outils formels, des modèles abstraits, qui sont le mieux possible adaptés aux phénomènes étudiés.

Ces outils abstraits sont construits par analogie avec les propriétés décelées et mises en évidence par l'expérimentateur du phénomène étudié. On trouve aussi cette attitude chez Wiener, le mathématicien appliqué, qui cherche à reconstruire pour chaque fonctionnalité un modèle abstrait, traduit dans un "langage" matérialisable par la suite. "Wiener avait son psychologue, Edwin Boring, qui lui fournissait des listes de fonctions psychologiques considérées d'un point de vue behavioriste, la charge revenant à Wiener d'en concevoir les "équivalents" électriques ou électroniques – "équivalents" voulant dire capables des mêmes transformations d'input en output" [Dupuy, 1994][p. 84]. Pour faire cela, Wiener réalise d'abord une description abstraite, un modèle formel de la fonction en termes électriques. Il construit ensuite le système électrique matériel qui lui correspond.

Cette attitude, Wiener la précise déjà dans un texte de 1945, écrit avec Rosenblueth [Wiener et Rosenblueth, 1945]. Il définit entre autres la notion de modèle abstrait ou formel. "Le modèle formel est l'assertion symbolique en termes logiques d'une situation idéalisée relativement simple partageant les propriétés structurelles du système factuel originel" [Wiener et Rosenblueth, 1945][p. 317]. On suppose de plus que cette situation simple est faiblement reliée au reste de l'univers, et que les caractéristiques de son comportement qui nous intéressent sont en quelque sorte autonomes. On sépare ainsi une partie de l'univers du reste de l'univers. On peut alors abstraire un modèle formel qui correspond au comportement observé de cette petite partie du monde simplifiée et considérée comme autonome. "La composition d'un modèle simple pour une boîte fermée presuppose que certaines variables sont seulement faiblement couplées avec le reste de celles qui relèvent du système" [Wiener et Rosenblueth, 1945][p. 319].

2.2.2. L'approche méthodologique du “premier ingénieur”

Celui que nous appelons le “premier ingénieur” s’inspire du biologique pour construire des réalisations matérielles performantes qu’il considère pour elles-mêmes sans en tirer nécessairement des conclusions sur la nature du phénomène dont il s’inspire. Il réalise une analogie entre le phénomène étudié et sa reconstruction matérielle. On retrouve en particulier cette attitude chez Wiener lorsqu'il réalise des prothèses pour handicapés. Ce travail constitue une construction matérielle. Celle-ci est partielle et s’appuie sur une analogie entre les fonctionnalités de la prothèse et celles de l’organe étudié.

Cette primauté de la reconstruction matérielle, antérieure à une abstraction formelle, se retrouve déjà en germe dans l’article de Wiener et Rosenblueth de 1945 [Wiener et Rosenblueth, 1945]. “Le modèle formel n’a pas besoin d’être parfaitement compris ; le modèle matériel sert alors à le compléter” [Wiener et Rosenblueth, 1945][p. 317]. Le modèle formel abstrait perd déjà ici une partie de son hégémonie.

2.2.3. L'approche méthodologique du “second ingénieur”

Le “second ingénieur” construit matériellement le modèle abstrait sur un support physique, désormais l’ordinateur. Dans ce travail il raisonne par analogie entre le modèle abstrait et sa reconstruction matérielle. Cette attitude est en partie nouvelle à l’époque. Elle constitue le mouvement inverse de l’abstraction réalisée au départ par von Neumann. Celui-ci, en travaillant comme consultant sur l’ENIAC, a compris que l’on pouvait retrouver des modèles formels et plus précisément la logique à partir de ces machines qui en constituent une matérialisation. Le travail effectué par von Neumann est donc une abstraction. En même temps, il propose l’idée de séparation, aujourd’hui évidente, entre les instructions qui constituent le *software* et la structure qui constitue le *hardware*. Il réalisera cette idée avec le projet EDVAC, puis avec JOHNNIAC.

Désormais, inversement, l’ordinateur comme matérialisation de la logique peut offrir la possibilité de matérialiser tout énoncé, toute abstraction présentés de manière formelle. Ceci est possible grâce à l’universalité de la machine abstraite de Turing. On retrouvera ce type de raisonnement chez Ashby qui reconstruit matériellement des éléments dont il s’est fixé le jeu logique *a priori*. Ce sera également la méthodologie privilégiée de l’intelligence artificielle (I.A.) qui se fixe par exemple des règles “accouchées” des experts d’une discipline, qu’elle “implémente” par la suite dans un ordinateur. Le système expert sur l’ordinateur est considéré comme une matérialisation de la logique de l’expert.

2.3. De l'analogie à l'identification

Avec l'apparition de l'ordinateur, les sciences et plus particulièrement la cybernétique, l'intelligence artificielle et les sciences cognitives sont confrontées concrètement à la valeur qu'elles attribuent à une analogie et à une métaphore. À la triple analogie que nous venons de mettre en évidence va succéder progressivement une triple identification ou réduction. Celle-ci ne fut pas toujours partagée par l'ensemble des participants aux conférences Macy. Cette triple réduction peut être considérée comme un triple choix métaphysique sur la nature des objets étudiés. En ce sens nous dirons que ces choix sont ontologiques. Ils dépassent l'affirmation de simples résultats et ne peuvent pas par nature être démontrés dans le cadre d'une activité scientifique.

2.3.1. La réduction du modèle abstrait à sa réalisation matérielle physique

La première réduction est celle du modèle abstrait à sa réalisation matérielle physique. Cette réduction nécessite deux étapes. La première est de dire que le modèle abstrait peut s'exprimer dans les termes de la machine de Turing. Ceci est possible depuis les travaux de Turing. Il a ouvert la voie, avec ses travaux sur l'équivalence entre les machines logiques abstraites. Von Neumann l'a brillamment suivie avec le projet EDVAC et la réalisation effective de machines qui semblent potentiellement universelles. Désormais on considère que la machine de Turing peut être matérialisée. Ceci consiste à oublier d'une part que les machines construites sont des machines de Turing finies et d'autre part que les machines sont constituées de matière et ne manipulent pas les abstractions que sont la logique ou les nombres. Cette possibilité de matérialisation des machines logiques abstraites est donc, en fait, une position ontologique et non un choix méthodologique. On raisonnera alors indifféremment sur les machines logiques abstraites ou sur leur matérialisation, à chaque étape, selon les besoins.

Nous avons déjà dit les limites que von Neumann lui-même attribuait aux machines réelles. Wiener ne prend pas toujours les mêmes précautions. Il écrit : “la science d'aujourd'hui est opérationnelle, c'est-à-dire considère chaque proposition comme essentiellement concernée par des expériences possibles ou des processus observables. En accord avec cela, l'étude de la logique doit se réduire à l'étude de la machine logique” [Wiener, 1961]¹⁸. En fait il faut comprendre deux choses dans cette phrase. Tout d'abord l'affirmation que la science ne s'intéresse qu'à des faits partagés, considérés comme publics. Implicitement, Wiener veut préciser que la science rejette les études qui relèveraient de la métaphysique. Il est paradoxal qu'il en conclut une prise de position ontologique en identifiant la logique avec la machine logique et en ne distinguant pas la machine logique abstraite et sa réalisation

¹⁸cité par [Lévy, 1987].

matérielle. Si seules les constructions matérielles importent, tout travail théorique en vue de l'abstraction et non de la matérialisation devient inutile. Pourtant la science a toujours travaillé sur des abstractions, comme Wiener le faisait lui-même remarquer dans *The role of models* [Wiener et Rosenblueth, 1945].

La position que décrit Wiener a évidemment des incidences méthodologiques. À la suite de cette identification entre la logique et sa matérialisation, on comprend que l'intelligence artificielle se soit intéressée essentiellement à la logique et aux systèmes formels, ou à leur matérialisation, sans vraiment tenir compte des différences qu'il pouvait y avoir entre ces deux points de vue. Wiener, en affirmant, peut-être involontairement, qu'il faut se méfier des abstractions et préférer leur réalisation matérielle, ouvre la voie à une méthodologie qui oubliera de faire des abstractions à partir de phénomènes étudiés de façon précise et expérimentale. Ceci va tout à fait à l'encontre de ce qu'il disait en 1945 [Wiener et Rosenblueth, 1945][p. 320]. À l'époque il insistait sur l'importance de "l'élaboration de modèles théoriques afin d'expliquer les faits observés – en d'autres mots la recherche scientifique de modèles abstraits avec une structure équivalente à celle d'une expérience donnée". Il précisait : "aucune partie substantielle de l'univers n'est suffisamment simple pour pouvoir être saisie et contrôlée sans abstraction" [Wiener et Rosenblueth, 1945][p. 316].

2.3.2. La réduction du phénomène étudié au modèle abstrait

La deuxième réduction est celle du phénomène étudié au modèle abstrait. Elle fait son apparition avec les travaux de McCulloch et Pitts. Désormais, le cerveau est considéré comme une machine logique, c'est-à-dire la matérialisation de la logique. Le cerveau peut être identifié à son modèle abstrait. Si des distinctions sont possibles, elles sont affaire de détail, mais non de substance. Selon eux l'essentiel, c'est-à-dire les fonctionnalités du cerveau, se retrouve dans le modèle logique.

2.3.3. La réduction du phénomène étudié à sa reconstruction matérielle sur un ordinateur

La troisième réduction est celle du phénomène étudié à sa reconstruction matérielle sur un ordinateur. En effet, le cerveau est la matérialisation d'une machine logique abstraite. Or toutes les machines logiques abstraites sont équivalentes, d'après les résultats de Turing. De plus la machine logique abstraite de Turing peut être matérialisée. Donc le cerveau peut être identifié à un ordinateur. Comme pour les deux réductions précédentes, ce raisonnement ne constitue pas une simple analogie, mais bien une prise de position ontologique. Puisque le cerveau et l'ordinateur sont tous les deux des matérialisations de la logique, on peut les identifier quant à leurs fonctionnalités. Le substrat est différent, mais leur potentialité fonctionnelle est identique. On obtient ainsi la troisième identification annoncée.

Cette identification est d'autant plus importante que, méthodologiquement, elle permet de réaliser des reconstructions matérielles de phénomènes étudiés, sans la construction préalable des abstractions adéquates susceptibles de permettre une compréhension minimale du phénomène ou de sa reconstruction. Une "nouvelle science" sans abstraction devient possible. Elle permet de laisser croire que l'on réalise des expériences et que l'on travaille sur des objets concrets parce que l'on manipule des réalisations matérielles. Ce qui est manipulé n'est pourtant pas le phénomène étudié. Cette attitude est d'autant plus séduisante que la simulation sur ordinateur permet de modifier les paramètres sans la moindre difficulté.

On oublie de préciser alors que l'expérimentation porte sur la reconstruction et non sur le phénomène. "Le but de l'enquête est de comprendre et de contrôler une part de l'univers" [Wiener et Rosenblueth, 1945][p. 316]. La tentation est grande de se détourner d'une compréhension du phénomène naturel pour lui préférer une compréhension du phénomène artificiel, cette autre part de l'univers. La question se pose de savoir ce que nous souhaitons comprendre *in fine*, et si une attitude concentrée sur nos propres reproductions nous satisfait vraiment. Changeons-nous d'objectif en cours de route ? Voulons-nous comprendre, non plus les phénomènes dégagés initialement, mais leur reconstruction, tout à fait différente et toujours aussi difficile à analyser ?

2.3.4. La triple réduction comme possible position ontologique au-delà des nécessités méthodologiques

On constate ainsi que la deuxième et la troisième réduction, qui constituent des prises de position ontologiques, ont également des implications méthodologiques. En effet l'identification entre le cerveau et le modèle logique abstrait, et celle entre le cerveau et l'ordinateur, se font par le biais d'identifications en terme de potentialités fonctionnelles. Pour analyser et comprendre un phénomène, la science distingue des comportements. Il faut comprendre cette pratique comme une nécessité méthodologique et non comme une position sur le statut ontologique du phénomène étudié. Les deux positions précédentes vont plus loin. Elles prônent en même temps une conception du cerveau, du modèle logique abstrait et de l'ordinateur, qui les considère exclusivement selon les différents comportements qu'ils peuvent présenter.

L'ensemble de ces trois réductions n'est autre que la reprise par des scientifiques d'une proposition du Tractatus : "le disque, le phonographe, la pensée musicale, les notes, les ondes sonores, tous se trouvent les uns les autres dans cette relation interne de représentation qui existe entre le langage et le monde, la structure logique leur est commune à tous" (4.014) [Wittgenstein, 1986]. Ce qui permet ces trois analogies, c'est tout d'abord la perception d'une structure logique sous-jacente similaire et potentiellement commune aux trois termes. Ceci permet de montrer *a posteriori* à quel point la démarche des cybernéticiens est imprégnée des travaux

et de l'état d'esprit qui s'étaient développés depuis *Les fondements de la géométrie* de Hilbert [Hilbert, 1971]. La triple réduction va encore plus loin, chez les cybernéticiens, et correspond à un raisonnement que l'on trouvait déjà chez Wittgenstein en 1918. Comme le note Lévy [Lévy, 1985a][p. 272], à propos de Wittgenstein, “c'est moins d'un parallélisme entre un ordre physique et un ordre logique qu'il s'agit, que d'une équivalence entre états de choses par l'intermédiaire de leur contenu logique commun”. Il faut entendre ici le terme équivalence dans son sens fort, non celui d'analogie, mais celui d'identification. Les cybernéticiens mathématiciens auraient sans doute utilisé le terme d'isomorphie, sans autre précision et recouvrant alternativement ces deux notions.

2.3.5. La prise en main par McCulloch ou la triple réduction comme position ontologique qui s'impose

Von Neumann ne participe pas à la sixième conférence Macy de mars 1949. Wiener et von Neumann n'apparaissent pas avec le groupe des cybernéticiens à la huitième conférence Macy en 1951, ni aux suivantes. Wiener rompt avec Pitts et McCulloch en 1952. La même année, par dépit suite à cette séparation, Pitts brûle tous ses papiers. On peut ajouter à cela la disparition prématurée de von Neumann en 1957⁹, ainsi que les engagements de Wiener pour une science avec conscience. De plus en plus, désormais, les travaux de McCulloch peuvent être lus selon cette triple identification. Tous ces éléments vont favoriser l'apparition d'une conception relativement unifiée de la notion de modèle et des relations entre le modèle, les phénomènes et la reconstruction matérielle. Cette conception s'appuie implicitement sur la triple identification mise en évidence précédemment. McCulloch en fut le moteur. Cette position devait selon lui s'accompagner d'une étude conjointe, par les mêmes chercheurs, des modèles formels, des éléments neurobiologiques ainsi que des constructions matérielles. Selon lui, cette condition, cette concentration sur ces trois aspects en même temps avait seule permis la cristallisation de l'article de 1943. Aussi c'est bien malgré lui que l'identification qu'il proposait fut comprise comme une justification permettant l'étude d'un seul des trois termes identifiés.

On peut avancer l'idée que von Neumann et Wiener étaient les figures de proue des premières conférences Macy. Heims écrit : “*von Neumann and Wiener were the star performers*” [Heims, 1980][p. 203]. Tous deux étaient prolifiques, on voyait peu von Neumann, mais on discutait beaucoup ses idées et les propositions qu'il avait pu faire lors de ses courtes apparitions aux conférences précédentes. Wiener apparaît comme un fouineur et un “touche-à-tout” qui n'hésitait pas à lancer des idées, même dans des disciplines qu'il ne connaissait pas, quitte à les voir rejetées au cours de la discussion. McCulloch, en comparaison, a une stature beaucoup plus académique. Il est l'organisateur et préside ce qui ressemble à “une foire d'empoigne” intellectuelle,

⁹Turing meurt en 1954. Il ne participait cependant pas aux conférences Macy.

ce qui correspond d'ailleurs à l'esprit de la fondation Macy. On comprend alors que, von Neumann et Wiener n'apparaissant plus, les trois dernières conférences voient la personnalité de McCulloch devenir prépondérante.

C'est en un certain sens sa position, déformée par diverses interprétations, qui va devenir prédominante. Dupuy écrit : "Wiener raisonne en mathématicien appliqué, il lui suffit d'établir un isomorphisme mathématique pour conclure à l'analogie. C'est une prise de position ontologique qu'en regard défend McCulloch" [Dupuy, 1994] [p. 42]. Il faut souligner que McCulloch n'est pas mathématicien. Aussi, lorsque les mathématiciens Wiener et von Neumann seront partis, plus aucun frein ne s'opposera aux interprétations du rôle des modèles que propose implicitement McCulloch. "Le modèle, pour McCulloch, n'est pas un simple instrument de calcul, dont la valeur serait purement pragmatique, du type "est-ce que ça marche ?" Il a une réalité ontologique" [Dupuy, 1994][p. 43]. De ce point de vue, la confrontation que furent les conférences Macy va déterminer pour longtemps le rôle que les diverses sciences enfantées par la cybernétique vont attribuer aux systèmes formels et donc aux mathématiques. Le risque d'une position ontologique comme celle de McCulloch étant d'amener certains scientifiques à se contenter de travailler sur un seul des trois termes de la figure 1, par exemple le modèle. Un tel état d'esprit est justement contraire à celui, pluridisciplinaire, que préconisait McCulloch. De ce point de vue on peut regretter que les personnalités fortes que furent von Neumann ou Wiener n'aient pas été plus claires, à cette époque, sur les distinctions entre des choix philosophiques et personnels et des choix de méthodologie scientifique.

2.4. Ontologie et conception de l'intelligence chez Turing

Les conceptions de Turing permettent d'illustrer ces positions ontologiques et de montrer que leurs implications, qui peuvent sembler éventuellement naïves ou caricaturales, sont en fait la recherche des conséquences ultimes d'une position métaphysique qui s'affirme. Turing attribue une grande importance aux automates. Pour lui, ils permettent de matérialiser la logique. Nous montrerons que l'on retrouve simultanément chez Turing les trois identifications.

2.4.1. La réfutation par Turing de cinq arguments opposés à la possibilité de construire des machines montrant un comportement intelligent

Turing "propose d'étudier la question de savoir s'il est possible pour une machine de montrer un comportement intelligent. Il est généralement admis sans argumentation que ce n'est pas possible" [Turing, 1948][p. 107]. Turing considère que ses machines permettent de rendre compte de l'intelligence humaine. Son premier objectif est "anti-métaphysique". Il n'accepte pas que l'on refuse *a priori* la comparaison entre l'homme et la machine. Selon lui les attaques contre sa conception s'appuient

généralement sur cinq arguments¹⁰. Trois sont d'ordre métaphysique. Le premier est “un refus d'admettre la possibilité que le genre humain puisse avoir un quelconque rival en termes de facultés intellectuelles” [Turing, 1948][p. 107]. Le deuxième est “une croyance religieuse selon laquelle toute tentative pour construire de telles machines est une sorte d'irrévérence prométhéenne” [Turing, 1948][p. 107]. Le troisième est que “si jamais la machine peut montrer un comportement intelligent, il ne faut le regarder comme rien d'autre qu'une projection de l'intelligence de son créateur” [Turing, 1948][p. 108]. Pour Turing, les deux premiers arguments “sont purement émotionnels et n'ont pas vraiment besoin d'être réfutés” [Turing, 1948][p. 108]. Il précise pourtant plus loin que “ces arguments ne peuvent pas être complètement ignorés, parce que l'idée d’“intelligence” est elle-même émotionnelle plutôt que mathématique” [Turing, 1948][p. 108].

Pour comprendre le projet de Turing, il faut aller plus loin dans l'analyse. Ces trois arguments contre Turing recoupent en fait deux points. Le premier point est constitué de préjugés, liés à une conception judéo-chrétienne de la nature de l'homme et de sa position dans le monde. L'homme est considéré comme un être à part, choisi. De par sa relation privilégiée à la divinité, il ne peut être comparé facilement aux autres éléments qui remplissent le monde. De même que l'homme ne pouvait pas “descendre du singe”, à l'époque de Darwin, son intelligence n'est pas comparable à celle des machines, à l'époque de Turing. Le deuxième point, prométhéen, selon les termes de Turing, est lié au fait que certains domaines du monde ne doivent pas être trop étudiés et doivent rester en dehors du champ de la connaissance rationnelle. On se protège, en dernier ressort, en affirmant que si les machines pouvaient présenter des comportements intelligents, cette connaissance et cette capacité ne seraient que le reflet de notre propre intelligence. Tous ces arguments se regroupent derrière une croyance selon laquelle s'interroger sur l'intelligence humaine et ne pas lui attribuer une place privilégiée serait diminuer l'homme. Ainsi, la position de Turing est d'abord “anti-métaphysique”, dans le sens où elle refuse dogmatiquement un certain nombre de préjugés sur la place prépondérante de l'homme dans l'univers. Turing soupçonne apparemment que ceux qui affirment l'impossibilité de son projet ont une conception anthropomorphique de l'intelligence.

Les deux autres arguments sont liés aux capacités des machines. Le quatrième argument est “le caractère très limité des machines qui ont été utilisées jusqu'en des temps très récents (i.e jusqu'en 1940)” [Turing, 1948][p. 107]. On oppose tout d'abord à Turing une conception des machines que lui-même juge naïve. “La nature d'une machine est de “faire la même chose toujours et encore, aussi longtemps qu'elle fonctionne”” [Turing, 1948][p. 108]. C'est selon Turing une version bien simplifiée de ce que peuvent faire les machines. Comme nous l'avons vu, les machines abstraites qu'il a définies ont une certaine universalité potentielle. Les machines peuvent donc

¹⁰Nous nous sommes appuyés sur l'article de Turing de 1948 [Turing, 1948]. Ces différents arguments et d'autres sont repris et discutés dans l'article de 1950 [Turing, 1950].

présenter des suites d'états très différents, c'est-à-dire des comportements très sophistiqués. Ainsi, si l'on peut comparer désormais l'homme et les machines, cela provient plutôt d'une élévation de la machine que d'un abaissement de l'homme. “[Von Neumann] retenait [des travaux de Turing] le caractère proprement paradoxal que le comportement d'une machine n'est pas la chose banale que l'on croit, puisqu'aucune machine ne peut le prévoir complètement” [Dupuy, 1994][p. 155].

Turing avait l'habitude qu'on lui oppose comme cinquième argument les limitations que Gödel a apportées à la logique. “Récemment, le théorème de Gödel et des résultats liés (Gödel 1931, Church 1936, Turing 1937) ont montré que si l'on essaye d'utiliser les machines avec comme objectif de déterminer la vérité ou la fausseté des théorèmes mathématiques et si l'on ne veut pas tolérer un résultat faux occasionnel, alors quelle que soit la machine elle sera incapable dans certains cas de donner une quelconque réponse” [Turing, 1948][p. 108]. Le résultat de Gödel nous apprend que certains énoncés d'un domaine formalisé ne peuvent être décidés qu'en faisant appel à des connaissances qui ne relèvent pas de cette formalisation. Turing réfute l'utilisation de cet argument contre la possibilité de l'intelligence des machines. Il affirme qu'en fait ce qui doit être remis en cause c'est essentiellement l'hypothèse selon laquelle une machine ne devrait jamais se tromper et ne pourrait tolérer un résultat faux occasionnellement. Turing considère que les hommes eux-mêmes, du fait de leur intelligence, donnent également des résultats faux¹¹. D'ailleurs cette notion de résultat vrai ou faux ne peut pas être donnée dans l'absolu. Cette propriété est liée à la non-complétude des systèmes formels, dont la complétion éventuelle nécessite l'introduction de systèmes plus sophistiqués. Ainsi Turing demande que l'on accepte que les machines fassent des erreurs comme les hommes. Ces erreurs ne doivent pas dénoter une inconsistance, mais une incomplétude du système formel employé.

Pour Turing, l'argument le plus important provenant des résultats de Gödel est lié au fait que les raisonnements s'appuyant seulement sur la logique sont insuffisants pour correspondre à l'intelligence humaine. En ce sens, il devait sans doute considérer que les limitations de sa machine logique étaient la logique elle-même. Pour aller plus loin, il fallait donc être capable d'expliquer comment on peut passer d'un système formel S_i à un système formel S_{i+1} qui le complète. Ceci explique sans doute l'intérêt de Turing à la fin de sa vie pour la “morphogénèse”, qui doit certainement être considérée comme l'ébauche d'une tentative de compréhension de l'émergence¹² d'un système S_{i+1} à partir d'un système S_i .

¹¹On est loin ici des considérations sans doute un peu rapides de McCulloch lorsqu'il affirme qu'avec Pitts ils avaient conçu leur modèle “comme un dispositif qui pourrait réaliser les fonctions qu'un cerveau doit réaliser si ce n'est se tromper et avoir une psychose” [von Neumann, 1951][p. 319].

¹²Nous ne faisons ici qu'effleurer le problème.

2.4.2. Comment tester, selon Turing, l'intelligence du comportement des machines ?

Selon Turing, l'homme est la meilleure preuve qu'une intelligence mécanique est possible. Ainsi, la position de Turing est avant tout une position métaphysique qui attribue un statut particulier au cerveau et à l'intelligence humaine. En ce sens, il a parfaitement raison de réfuter tout argument s'appuyant sur les résultats de Gödel. Ceux-ci portent sur des limitations d'ordre logique. Ils ne statuent aucunement sur la nature des choses, même s'ils obligent à laisser de côté une ontologie trop naïve.

Pour son projet, Turing propose une méthode qui correspond à une certaine conception de l'intelligence. Comme nous l'avons déjà écrit, Turing "propose d'étudier la question de savoir s'il est possible pour une machine de montrer un comportement intelligent" [Turing, 1948][p. 107]. La question que pose Turing est donc formulée en termes fonctionnels. Ce que veut étudier Turing, ce sont les comportements intelligents, que l'on assimilera bien vite à l'intelligence, mais qui n'en sont que la manifestation. Cette conception qui consiste à étudier des comportements et non les structures matérielles qui les permettent correspond à la position que Wiener, Rosenblueth et Bigelow défendaient dans leur article de 1943 [Wiener *et al.*, 1943]. Turing ne fait pas la différence entre comportements intelligents et intelligence. Selon lui, l'intelligence ne vaut que par ses manifestations. Ainsi, ce qui pouvait être au départ une proposition méthodologique consistant à étudier l'intelligence en termes d'entrées/sorties ou de comportements devient une prise de position ontologique, l'intelligence étant définie par l'ensemble des comportements qu'elle permet.

De plus Turing a une conception formelle de l'intelligence, du moins dans son article de 1948. Il propose un certain nombre de domaines où l'on pourrait tester l'intelligence des machines. "Nous sommes alors confrontés au problème de trouver des branches de la pensée adaptées pour que la machine y exerce ses capacités. Les domaines suivants me semblent avoir certains avantages :

- i) différents jeux, par exemple les échecs, ... le bridge, le poker
- ii) l'apprentissage des langues
- iii) la traduction des langues
- iv) la cryptographie
- v) les mathématiques" [Turing, 1948][p. 117].

Il est à noter que tous ces domaines proposés par Turing font appel à une connaissance formalisée, formalisable ou plus ou moins symbolique. Leur intérêt est qu'ils semblent correspondre à des activités bien différenciées, qui semblent pouvoir être définies et effectuées sans faire intervenir d'autres activités. Elles semblent de

plus ne pas nécessiter une forme particulière de matérialisation. Tout ceci permet de raisonner sur elles en termes d'entrées/sorties. Il faut tout de suite rappeler que les domaines formalisables comme l'arithmétique ne sont pas des champs scientifiques dont la véracité des propositions provient exclusivement de la logique. En ce sens, le formalisable ne peut pas être réduit au logique. Nous l'avons déjà vu avec les résultats de Gödel.

On peut donc énoncer deux critiques aux tests que propose Turing. La première est liée au fait qu'un système ne pourra pas décider avec le seul formalisme qu'il peut manipuler, tous les énoncés bien formés avec ce formalisme. Nous avons vu que Turing a en fait déjà répondu à cette critique, lorsqu'on lui oppose des arguments liés aux résultats de Gödel. Elle ne met cependant pas en cause la possibilité de construire des machines intelligentes. En effet on retrouve aussi cette incapacité chez l'homme. Eux non plus ne peuvent pas décider la validité de tous les énoncés qu'ils sont susceptibles de formuler.

La seconde critique est liée à la définition implicite que Turing fait de l'intelligence. Pour lui toute intelligence se manifeste dans des activités qui s'énoncent dans des systèmes formels ou symboliques. L'intelligence est alors essentiellement syntaxique et si elle prend sens ce sera une simple propriété supplémentaire qui n'est pas requise *a priori*. Le sens viendra du passage d'un niveau syntaxique à un autre. On peut tout à fait contester cette limitation de la nature de l'intelligence. Conduire une voiture, manger une pizza, courir après un bus constituent des activités de la vie quotidienne que l'on peut considérer comme mettant en jeu l'intelligence humaine. Certaines activités mettent en jeu une intelligence formelle. On peut cependant contester que toutes les formes de l'intelligence humaine soient de ce type. Il est possible de soutenir que l'intelligence humaine ne se réduit pas à des activités formelles et symboliques. On peut défendre que toute une partie de notre intelligence est étroitement liée à notre incarnation particulière. Alors, même en raisonnant en termes formels, l'intelligence ne se réduit pas à la manipulation de symboles au niveau S_{i+1} qui seraient indépendants des syntaxes inférieures qui auraient permis de la constituer, de S_0 à S_i .

Il faut cependant expliquer, ne serait-ce que rapidement, pourquoi Turing défend un point de vue qui considère l'intelligence comme essentiellement formelle. Ce sont les succès de la cryptographie, de la traduction automatique¹³ et de la sémantique statistique pendant la guerre qui ont permis cet engouement. "Ce que la sémantique statistique a de plus surprenant, c'est l'absence d'un effort de compréhension du texte traduit. Cette approche paraît, rétrospectivement, si incroyable que l'on peut se demander comment des gens intelligents ont pu penser qu'elle tenait debout" [Haugeland, 1989][p. 169]. Haugeland propose, pour l'expliquer, trois facteurs que

¹³Succès à l'époque de la guerre. Les doutes, puis les limitations, viendront avec les travaux de Bar-Hillel en particulier, qu'il exposera dans son livre de 1960. Haugeland [Haugeland, 1989] donne quelques détails et les références.

nous reprenons ici. Le premier est lié aux progrès de la cryptographie statistique. Weaver cite en particulier un cas d'un message codé à partir du turc, retranscrit en turc par quelqu'un qui n'avait jamais parlé cette langue. Le deuxième facteur serait la possibilité développée à la fin des années trente, par Turing en particulier, de passer automatiquement et sans erreur, par une simple transformation, d'un langage formel à un autre¹⁴. Le troisième facteur est le développement de la théorie de l'information de Shannon et Weaver, qui consiste à étudier la transmission de manière statistique. Le cas du passage d'un langage formel à un autre est alors considéré comme un cas pur, sans bruit au cours de la transformation du signal, alors que les langues naturelles seraient des cas imparfaits, les imperfections intervenant du fait de la matérialisation. La cryptographie peut être considérée comme une activité intelligente et ne met pourtant en œuvre, contre toute attente, que des activités formelles. On peut alors supposer que toutes les activités mettant en jeu l'intelligence sont de même nature, c'est-à-dire formelles¹⁵.

2.4.3. Interpréter dogmatiquement les résultats de Turing et des cybernéticiens

Sous les traits de Turing, mais aussi chez certains membres les plus reconnus du cercle des conférences Macy, on constate que la cybernétique constitue aussi une prise de position ontologique. Cette position se retrouve à deux niveaux. Le premier consiste à s'appuyer sur une crédibilité scientifique pour défendre des thèses qui *in fine* ne sont pas obtenues de faits d'expérience, mais d'une position métaphysique *a priori*. On trouve déjà sous la plume de Pierre Cossa [Cossa, 1957] une critique de ce mode d'argumentation. Le second consiste à dire qu'en réduisant le monde à des modèles et en les reconstruisant, en décrivant les phénomènes sous l'angle exclusif de systèmes perturbés par des entrées et des sorties et en occultant volontairement la structure de ces objets, les cybernéticiens vont bien au-delà de leur rôle scientifique. Ils prennent position sur la nature des choses et du monde. Ils considèrent déjà que l'essentiel se situe dans les comportements possibles des objets, leur structure et le substrat qui les permettent étant annexes.

Cette attitude peut être féconde quand elle se pose comme hypothèse dont on cherche à tirer les conséquences. Quand elle devient position métaphysique sur laquelle on n'est pas prêt à revenir, on aboutit à une certaine forme de dogmatisme qui ne permet plus d'étudier d'autres hypothèses s'appuyant sur des positions différentes. On ne peut sans doute taxer ni les cybernéticiens ni Turing de dogmatisme. Cependant, certaines de leurs positions seront après eux défendues comme si elles étaient devenues des vérités absolues qui ne doivent pas être remises en cause.

¹⁴Nous avons vu que cette interprétation des résultats de Turing était un peu rapide.

¹⁵Cette conception purement formelle de l'intelligence est déjà remise en cause implicitement par Turing lui-même [Turing, 1950] ; voir à ce propos [Andler, 1996].

2.5. Les remises en cause de la méthodologie formaliste classique et les premières tentatives pour la dépasser

2.5.1. Le rôle des modèles selon Wiener et les premières difficultés rencontrées pour modéliser le cerveau/esprit

Le texte de Wiener et Rosenblueth de 1945 [Wiener et Rosenblueth, 1945] est particulièrement intéressant, car il permet de connaître, de façon précise, les positions de Wiener avant les premières rencontres Macy. Wiener commence par préciser les objectifs de la recherche scientifique, à savoir “obtenir une compréhension et un contrôle d'une partie de l'univers” [Wiener et Rosenblueth, 1945][p. 316]. Il ajoute qu’“aucune partie substantielle de l'univers n'est suffisamment simple pour être saisie et contrôlée sans abstraction” [Wiener et Rosenblueth, 1945][p. 316]. Il affirme ainsi deux choses ; d'une part que le but de la science est de comprendre un phénomène et non de le reproduire à l'identique ; d'autre part que la compréhension scientifique est partielle et doit faire référence à un phénomène bien délimité. On peut cependant constater que cette compréhension d'un phénomène limité doit être totale. Pour cela, il explique que les modèles d'abord simples se complexifient pour rendre de mieux en mieux compte du phénomène. Il est enfin très important de constater que Wiener insiste sur la notion d'abstraction. Celle-ci “consiste à remplacer la partie de l'univers considérée par un modèle similaire mais de structure plus simple” [Wiener et Rosenblueth, 1945][p. 316]. Wiener distingue alors deux types de modèles, des modèles matériels d'un côté et des modèles formels ou intellectuels de l'autre.

Tout ceci semble bien évident. Les choses ont pourtant bien changé depuis que l'on utilise des ordinateurs de façon systématique. Wiener ajoute que “les problèmes sont souvent approchés [...] du factuel à l'abstrait” [Wiener et Rosenblueth, 1945] [p. 317]. “L'ajout progressif de terminaux¹⁶ et de variables amène progressivement des modèles théoriques plus élaborés : de là une hiérarchie dans ces modèles, de ceux qui sont relativement simples et très abstraits à des modèles plus complexes avec des structures théoriques plus concrètes” [Wiener et Rosenblueth, 1945][p. 319]. Wiener considère ici que progressivement on va faire un lien de plus en plus étroit entre les modèles et les données expérimentales obtenues sur le phénomène étudié. Il est d'ailleurs vraisemblable que c'est par un argument de ce type qu'il est parvenu à convaincre von Neumann de l'utilité de travailler dès cette époque sur les systèmes nerveux. En effet, comme nous l'avons déjà fait remarquer, von Neumann était relativement sceptique au départ et aurait préféré formaliser d'autres disciplines. “Ce qui troublait von Neumann, c'était l'état de la neurophysiologie expérimentale et sa relation à la théorie” [Heims, 1980][p. 204].

¹⁶Il ne s'agit évidemment pas ici de terminaux informatiques, mais de paramètres dont Wiener ne précise pas la nature.

L’argumentation de Wiener laisse présager que l’on peut construire d’abord des modèles simples, puis que le “progrès” permettra de les améliorer au fur et à mesure. Mais nous montrerons par la suite que, pour des phénomènes “complexes” et pour les phénomènes cognitifs en particulier, ceci est quasiment impossible. Si le modèle initial n’est pas suffisamment sophistiqué, il est quasiment impossible de l’améliorer *a posteriori*. C’est aussi une confrontation à la complexité de l’“esprit” que permirent les conférences Macy. Cette méthodologie “classique” de la modélisation, qu’évoque Wiener en 1945, va montrer ses limites avec l’étude de phénomènes comme le fonctionnement du cerveau/esprit humain.

Toute la difficulté porte sur le fait que les modèles simples doivent déjà être pertinents pour pouvoir être complexifiés. “Une analogie grossière n’est pas féconde scientifiquement” [Wiener et Rosenblueth, 1945][p. 318]. “La composition d’un modèle simple pour une boîte fermée suppose qu’un certain nombre de variables sont seulement faiblement couplées avec le reste des variables appartenant au système” [Wiener et Rosenblueth, 1945][p. 319]. Wiener a réussi à reproduire des comportements qu’on lui rapportait. Mais il semblait bien aux participants des conférences Macy que ces comportements particuliers ne permettaient pas de construire des premiers modèles “simples” que l’on pourrait complexifier et qui pourraient rendre compte un tant soit peu des particularités de l’esprit. C’est certainement avec cet état d’esprit qu’il faut interpréter les réactions violentes à la présentation par Ashby de ses automates au cours de la neuvième conférence Macy. Ceux-ci n’étaient pourtant que la mise en œuvre de certaines conséquences des propositions que les cybernéticiens avaient eux-mêmes formulées.

2.5.2. Les prémisses de la mise en paragon

Von Neumann est sans doute celui qui le mieux a pris conscience de l’échec de cette tentative de capter l’esprit dans une optique modélisatrice classique. Il envoie comme message à discuter en son absence à la sixième conférence Macy : “ 10^{10} neurones, traités comme des relais simples, sont tout à fait insuffisants pour rendre compte des capacités humaines”¹⁷. Ce doute n’est pas seulement le résultat d’un calcul mais également un questionnement sur la méthodologie. Pitts reprendra cette estimation et montrera qu’elle est incorrecte car von Neumann ne tient compte que des éléments en jeu et non des états possibles de chaque neurone, c’est-à-dire de l’ensemble des configurations possibles du réseau. Von Neumann insiste ici sur les éléments constituants, leur structure et leur nombre. Ceci est bien différent de l’approche en termes d’entrées/sorties prônée en 1943 par Wiener.

Von Neumann prend conscience que la complexité du cerveau n’est pas seulement comportementale, mais liée aussi en bonne partie à sa complexité matérielle. Il est à ce sujet particulièrement intéressant de comparer les descriptions du système nerveux

¹⁷ cité dans [Dupuy, 1994][p. 55].

proposées respectivement par von Neumann au Hixon Symposium [von Neumann, 1951] et par Louis de Broglie [de Broglie, 1953]. On constate que le premier considère le système nerveux comme un objet d'une complexité extrême, à la fois par le nombre des éléments, par les différents types de réactions en jeu et par l'enchevêtrement des relations entre les différents éléments. En comparaison, la description de de Broglie donne l'impression étrange que le système nerveux est un simple réseau électrique. Le texte de de Broglie laisse accréditer l'image selon laquelle les cybernéticiens ne connaissaient pas la biologie détaillée du système nerveux. Ceci est totalement erroné. En particulier von Neumann discute en détail l'importance respective des activités de type humoral et neural du système nerveux et s'interroge sur la pertinence de ne tenir compte que de l'activité nerveuse.

Aussi un modèle formel du cerveau ne peut plus suffire, pour von Neumann : il lui faut également un modèle matériel. Le billet envoyé à la sixième conférence peut donc être interprété comme un appel direct aux neurophysiologistes et aux biologistes, pour qu'ils rappellent l'importance de la matérialisation particulière que constitue le cerveau. Cet appel, von Neumann l'avait déjà fait dans l'introduction à sa conférence au Hixon Symposium, en 1948, dans laquelle il avait présenté sa théorie des automates auto-reproducteurs. Hélas, les biologistes n'ont pas vraiment relevé le gant et n'ont pas apporté les éléments qui permettraient de produire une formalisation pertinente. Comme nous l'avons déjà dit, von Neumann était à l'époque tout à fait surpris du manque d'explication théorique globale et pertinente du fonctionnement du cerveau. En un certain sens, sa prestation du Hixon Symposium est une reculade volontaire. Puisqu'on ne lui offre pas de quoi formaliser le fonctionnement du cerveau, il propose une théorie des automates qui se reproduisent. Selon le mot de Wiener, "une analogie grossière n'est pas féconde scientifiquement" [Wiener et Rosenblueth, 1945].

Certainement la conviction de von Neumann à la fin de sa vie est-elle la nécessité d'étudier les sophistications du cerveau dans leur matérialité. Par sa complexité, le cerveau n'autorise que la création d'un modèle complexe. Von Neumann ne voit pas comment rendre compte en termes formels de la complexité qu'il perçoit. Pour lui les mathématiques de son époque sont dans l'incapacité d'en rendre compte et il ne voit pas comment en créer de nouvelles qui offrent cette possibilité. Ainsi, pour lui, seuls les modèles matériels complexes peuvent rendre compte de phénomènes matériels complexes. Toute la difficulté ensuite est de ne pas tomber dans les pièges décelés par Lewis Carroll et décrits dans *Sylvie et Bruno*¹⁸, où il a montré que la seule échelle complètement satisfaisante pour une carte d'un pays donné est ce pays lui-même. Cette conception extrême du modèle est déjà présente chez Wiener comme le modèle absolu, à la fin de son article sur le rôle des modèles [Wiener et Rosenblueth, 1945]. On retrouve également des variations autour de ce thème dans certaines nouvelles de Borgès. On risque, si on n'y prend garde, de perdre deux des impératifs que

¹⁸exemple cité dans [Wiener et Rosenblueth, 1945].

Wiener avait fixés à la science, abstraire et comprendre.

Il semble que von Neumann avait bien compris ce qui était en jeu, les contraintes et les conséquences que cette position imposait. Au Hixon Symposium, à Pasadena, en 1948 [von Neumann, 1951], il présente sa formalisation de cette autocréation des automates. Veut-il par là théoriser ce dernier maillon qui lie l'homme à la machine ? Est-ce la dernière axiomatique moqueuse d'un mathématicien qui a vu les limites de ses théories ? Peut-être a-t-il compris, à la fin de sa vie, que toute sémantique pouvait avoir sa version formalisée, aussi utopique soit-elle ? Cette théorie est extrême, à la fois dans son objet, qui est l'élaboration d'automates devenant indépendants de leur concepteur, et dans sa méthodologie. En effet elle nécessite une construction matérielle de ces automates, ce qui oblige à aller au-delà de la méthodologie scientifique axiomatique, telle qu'elle était conçue dans les travaux de Hilbert. Nous reviendrons sur ce point dans les chapitres suivants, car la constitution d'une autre méthodologie, après l'axiomatique, concerne en premier lieu les sciences cognitives. “L’utopie de la créature mécanique renouvelle, en l’inversant, le thème de la création de l’homme par Dieu. Le mythe d’une nouvelle chute originelle possible de sa créature à lui, tout en l’effrayant, le fascine et le flatte, car c’est lui qui jouerait alors le rôle de Dieu auquel on désobéit. Il est au fond déçu quand on le ramène à la raison, et quand on souligne – ce qui est pourtant l’évidence même – qu’une machine à calculer électronique n’est pas plus un cerveau surhumain qu’une machine à roue dentée, et qu’un avion sans pilote ou un projectile autoguidé n’a rien en lui-même de plus particulièrement maléfique que le régulateur d’un chauffage d’appartement. Il est déçu quand il est contraint de s’apercevoir qu’il lui est impossible de couper les liens qui relient ses créatures à sa propre vie, quand il est contraint de constater que sa technique échoue vraiment à créer, de même que sa science à vraiment comprendre” [Ruyer, 1954][p. 30-31]. Cette citation de Ruyer est une critique anticipée et sans doute un peu rapide de la méthodologie que von Neumann cherche à mettre en place et qu'il n'a pas totalement explicitée. Cette nouvelle méthodologie, que recherche von Neumann, Wiener l'a lui-même bien comprise. Cela explique certains travaux de la fin de sa vie sur la création d'objets construits par l'homme et qui doivent prendre leur autonomie. “L'autonomisation du modèle se confond avec la modélisation de l'autonomie. Griserie et paradoxe de fabriquer un automate, c'est-à-dire un être qui ne tienne sa loi que de lui-même. Comble de la volonté de puissance que de se vouloir la cause d'un être qui soit cause de soi, inconditionnée ; et simultanément perte de soi dans le miroir tendu par une créature faite à son image” [Dupuy, 1994][p. 156].

2.5.3. La première voie compatible avec les résultats de von Neumann : le nécessaire retour à l'observation et à l'expérimentation neuroanatomiques

Au-delà de cette “griserie”, il importe de poursuivre le chemin pour chercher à comprendre comment fonctionnent le cerveau et l'esprit. McCulloch et Pitts à la fin de leur vie ont suivi deux voies distinctes qui correspondaient toutes deux à des propositions qu'avait faites von Neumann. La première possibilité était d'étudier les détails neuro-anatomiques du système nerveux et du cerveau, c'est celle que suivra Pitts au cours d'une fin de vie tourmentée. Cette voie avait été particulièrement mise en évidence par von Neumann et le neurophysiologiste Ralph Gerard. Au cours de la septième conférence Macy, celui-ci, selon les termes de Dupuy, “mit solennellement le groupe en garde contre les dangers d'une formalisation mathématique qui irait beaucoup plus vite que l'observation et l'expérimentation” [Dupuy, 1994][p. 87]. McCulloch lui-même l'avait compris lors des questions qu'il posa suite à l'intervention de von Neumann au Hixon Symposium. Il tire les conclusions que von Neumann n'avait pas formulées explicitement. “Aucune des théories ne nous dit comment une opération particulière est effectuée, pas plus qu'elles ne nous disent avec quelle sorte de système nerveux particulier elle est effectuée, pas plus qu'elles ne nous disent dans quelle partie d'un ordinateur elle est effectuée. Ceci signifie que vous [von Neumann] nous obligez à étudier l'anatomie et à exiger de l'anatomiste des éléments qu'il nous a rarement donnés dans des détails suffisants. J'ai enseigné la neuro-anatomie quand j'étais à l'école de médecine, mais jusqu'il y a un an ou deux, je n'étais pas en mesure d'interroger un quelconque neuro-anatomiste sur le détail précis d'une structure quelconque. Je n'avais pas de raison physiologique pour souhaiter ce type d'information. Maintenant on commence à en avoir besoin” [von Neumann, 1951][p. 320].

On comprend également en lisant cette remarque que McCulloch avait parfaitement conscience d'une hypothèse fondamentale que Pitts et lui avaient faite en construisant leur modèle. Il savait que les neurones ne sont pas le niveau biologique fondamental, mais il faisait l'hypothèse que ce niveau est suffisant pour rendre compte du fonctionnement du cerveau. Les neurones sont en quelque sorte les atomes que McCulloch avait choisis pour rendre compte des faits psychologiques sur une base physiologique. Ceci n'empêchait pas McCulloch de s'intéresser à des travaux qui seraient partis d'autres hypothèses. En particulier, il avait lu avec intérêt les travaux de l'autrichien von Foerster qui rendait compte du caractère de “tout ou rien” de l'influx nerveux avec le formalisme de la mécanique quantique. Suite à cela, il le convia à participer dès 1949 à la sixième conférence puis aux suivantes.

2.5.4. La deuxième voie compatible avec les résultats de von Neumann : le développement de logiques alternatives

La deuxième possibilité était de prendre la suite du constat, formulé par von Neumann, que nous n'avons pas les outils logiques nécessaires pour construire des modèles du système nerveux satisfaisants. En fait von Neumann va plus loin, car il dit que ces outils sont manquants pour fournir une théorie des automates construits par l'homme. Or il a bien pris soin dès le début de montrer que le cerveau a certainement un fonctionnement beaucoup plus sophistiqué que les automates tels qu'ils sont construits. Ceci ne retire rien au fait que von Neumann considère que le cerveau est un automate particulier.

Toujours dans son intervention au Hixon Symposium, il précise quelles propriétés minimales on doit exiger d'une logique pour qu'elle puisse rendre compte de certaines propriétés des automates. "C'est pourquoi la logique des automates diffèrera de notre système actuel de logique formelle en deux aspects essentiels. Premièrement la longueur effective des "chaînes de raisonnement", c'est-à-dire des chaînes d'opérations, devra être considérée. Deuxièmement les opérations de la logique (syllogismes, conjonctions, disjonctions, négations, etc., c'est-à-dire dans la terminologie habituelle pour les automates, les formes variées d'ouverture de porte, de coïncidence, d'anti-coïncidence, de blocage, etc., d'actions) devront toutes être traitées comme des procédures admettant des exceptions (dysfonctionnements) avec des probabilités faibles mais non-nulles" [von Neumann, 1951][p. 304]. Von Neumann précise plus loin : "il y a de nombreux indices qui nous font penser que ce nouveau système de logique formelle va se déplacer progressivement vers une autre discipline qui a été peu reliée avec la logique par le passé. Il s'agit de la thermodynamique, dans sa forme première présentée par Boltzmann" [von Neumann, 1951][p. 304]. Von Neumann a même proposé les prémisses d'une nouvelle logique dans un article de 1956, publié quelques mois avant sa mort [von Neumann, 1956]. Comme l'indique le titre de l'article, il cherche à construire une logique probabiliste qui permette de faire la synthèse d'organismes liés à partir de composants indépendants.

McCulloch avait bien compris quels étaient les obstacles que von Neumann avait mis en évidence. "McCulloch en fut ébranlé et il consacra les dernières années de sa vie à des recherches sur des logiques alternatives, mais sans beaucoup de succès, semble-t-il" [Dupuy, 1994][p. 65]. Ces considérations sur la nécessité de construire de nouvelles logiques furent certainement initiées par von Neumann. Leur justification relève du bon sens. Elle s'appuie sur des connaissances simples de neurobiologie fonctionnelle. L'argumentation provient de "la fiabilité fonctionnelle du système nerveux. Comment concevoir des réseaux qui demeurent logiquement stables (c'est-à-dire calculent les mêmes fonctions) quand leurs seuils connaissent, non pas de petites perturbations, mais d'énormes variations ? Ces énormes variations de seuils se produisent dans les états comateux, les anesthésies générales, en cas d'absorption de grandes quantités d'alcool, etc., sans que les principales fonctions vitales, com-

mandées par le système nerveux, comme la respiration, soient trop perturbées” [Lévy, 1985a][p. 235].

McCulloch poursuivra le travail en s’intéressant à diverses logiques. Avec Pitts, il s’intéresse d’abord aux logiques probabilistes du type de celle développée par von Neumann. Avec S. Winograd et J.D. Cowan, il cherche à mettre en relation, de manière théorique, fiabilité fonctionnelle et redondance logique. Dans ce cadre, il développe de nouvelles notations des opérations logiques plus adéquates pour la conception et le développement des réseaux. Dans les années soixante il étudie en particulier les logiques non-aristotéliciennes. “Il pensait qu’une simple logique des classes n’était pas capable de rendre compte de la complexité du fonctionnement du système nerveux (et donc de l’esprit)” [Lévy, 1985a][p. 236]. Il revient également à des logiques “stoïciennes”, qui intègrent le temps et auxquelles il s’était intéressé avant de lire l’article de Turing et de concevoir avec Pitts son modèle de 1943. “Cette recherche nouvelle est étroitement liée dans son esprit au problème de l’abduction” [Lévy, 1985a][p. 236].

3. L'INTELLIGENCE ARTIFICIELLE ET LES SCIENCES COGNITIVES, QUARANTE ANS DE CONFRONTATION À UNE ONTOLOGIE FORMALISTE (1956-1996)

3.1. Les héritiers de von Neumann, l'intelligence artificielle et les sciences cognitives

McCarthy organise un séminaire d'été à Dartmouth, en 1956. Il crée à cette occasion le terme d'intelligence artificielle. Newell et Simon y présentent leur premier programme intelligent qui démontre des théorèmes. Minsky présente dans une première ébauche un article qui constitue un programme de recherche pour l'intelligence artificielle. McCarthy commence à travailler sur le langage informatique Lisp. Une nouvelle discipline est née. L'étude du cerveau ou de l'esprit sous toutes ses formes prend un nom et un nouveau visage, liés à une technologie¹. “Alors que la cybernétique dans sa seconde phase conservait l'ambition de modéliser l'intelligence naturelle et maintenait le contact avec la neurologie, l'intelligence artificielle se délivrait de cette contrainte et liait son sort au développement des ordinateurs” [Dupuy, 1994][p. 58]. Les chercheurs qui ont initié ce programme avaient tous été d'une façon ou d'une autre en relation étroite avec certains des cybernéticiens issus des disciplines les mieux formalisées. Ainsi, Shaw, collaborateur de Newell et Simon pour leur projet, leur donne accès au Johnniac. Shaw ayant déjà participé à l'élaboration de l'ENIAC, il avait donc fréquenté von Neumann pendant plus de dix ans. Newell avait suivi les cours de von Neumann à Princeton. Shannon va se passionner pour les automates de von Neumann dès son intervention au Hixon Symposium. Minsky et McCarthy, qui s'étaient liés d'amitié pendant leurs études de mathématiques à Princeton, vont devenir stagiaires de Shannon. Ainsi, les quatre fondateurs de l'intelligence artificielle ont été en contact indirect mais essentiel avec les aspects formels et matériels de la théorie des automates qu'avait conçue von Neumann.

Pendant les années soixante à quatre-vingts, l'intelligence artificielle s'est fixé

¹Pour une introduction historique et une présentation plus complète de l'intelligence artificielle, on pourra consulter les deux articles de Jean-Louis Le Moigne, dans l'ouvrage *Intelligence des mécanismes, mécanismes de l'intelligence*.

comme principal objectif la résolution de problèmes et la construction d'experts artificiels. Il semblait qu'atteindre ces objectifs permettrait également de comprendre l'intelligence humaine. L'optique principale consistait à développer les idées que Turing avait proposées, filtrées comme il convenait. On cherchait en fait à rendre compte des facultés humaines, à partir des présupposés ontologiques de la cybernétique, réinterprétés, déformés et poussés à leur extrême. Le programme consistait à rendre accessibles et à reconstruire les activités humaines où l'intelligence humaine, affirmée comme complètement formelle, est en jeu. On continuait à supposer le soubassement logique de l'esprit. "La thèse de la BOnde Vieille Intelligence Artificielle² n'est pas que les processus qui sous-tendent l'intelligence peuvent être décrits symboliquement (caractéristique également partagée par les ouragans et les embouteillages), mais qu'ils sont symboliques (ce qui les rend très différents des ouragans et des embouteillages)" [Haugeland, 1989][p. 113-114]³. L'explicitation des capacités fonctionnelles particulières en termes logiques devenait un simple exercice de style. Ces présupposés ontologiques ancrés profondément dans la méthodologie de ces chercheurs explique leur foi aussi bien qu'un aveuglement complet quant à l'évaluation de la difficulté des problèmes à résoudre. Ceci explique pourquoi Minsky considère que la vision par ordinateur est un problème simple dont il confie la résolution à trois étudiants comme travail d'été.

Les premiers successeurs des cybernéticiens et des conférences Macy, comme Ashby, ont rapidement imposé les directions pertinentes et la méthodologie autorisée pour étudier le cerveau et l'esprit avec les outils formels que les sciences proposaient. "Avec Ashby, on est dans le pur *a priori* mathématique. Le projet interdisciplinaire de la cybernétique, l'interaction avec ceux qui, concrètement, interrogent le vivant ou l'esprit et entrent en "conversation" avec les objets du monde, de difficile et limité qu'il était, est maintenant devenu radicalement impossible" [Dupuy, 1994][p. 165].

Arbib résume bien l'état d'esprit de ces années là. Il écrit : "au milieu des années soixante, beaucoup de chercheurs ont abandonné l'étude intégrée des cerveaux et des machines pour développer l'intelligence artificielle comme une fin en soi, c'est-à-dire la programmation d'ordinateurs pour mettre en évidence certains aspects de l'intelligence humaine, mais en insistant sur l'obtention de références pour les performances, plutôt que sur la saisie des mécanismes qui permettent aux hommes d'être eux-mêmes intelligents. Certains chercheurs ont essayé d'utiliser les concepts de l'I.A. pour modéliser la cognition humaine en se servant de programmes informatiques, mais ils étaient si dominés par la métaphore "l'esprit est un ordinateur" que nombre d'entre eux ont défendu que l'esprit devait partager avec les ordinateurs des années 60 la propriété d'être séquentiel, d'exécuter des séries d'opérations les

²Le terme de BOVIA est souvent utilisé avec ironie par les détracteurs de l'Intelligence Artificielle.

³Haugeland fait également référence à certains travaux de Dreyfus qui mettent en évidence cette prise de position ontologique. On trouvera la référence de Dreyfus, datant de 1965 dans [Haugeland, 1989] [p. 114].

unes après les autres. Après les années 60 vinrent les années 70 et cette tendance s'est poursuivie" [Arbib, 1987][préface p. vii]. Arbib poursuit en insistant sur ses positions personnelles qui étaient tout autres. "J'ai toujours insisté sur le fait que la métaphore "l'esprit est un ordinateur" ne devait pas se lire comme une réduction du cerveau au niveau de notre technologie actuelle de l'ordinateur" [Arbib, 1987][préface p. vii]. On constate bien dans les propos d'Arbib qu'une des analogies que nous avions soulignées précédemment a été prise à la lettre. Elle est réduction et comme telle a eu des incidences méthodologiques importantes. De ce point de vue, l'essentiel est de constater que l'on a donné la primauté à l'utilisation de l'ordinateur. De plus on a constaté un retour à une spécialisation des disciplines tout à fait contraire à la conception pluridisciplinaire d'une méthodologie de la modélisation de l'esprit telle que McCulloch l'avait développée.

Les positions ontologiques précédentes vont également permettre la constitution d'une discipline plus large qui s'intéressera aux diverses formes que peut prendre la cognition. "Mais la leçon à en tirer va plus loin : si l'intelligence artificielle n'a vraiment pas grand-chose à voir avec la technologie informatique mais dépend bien plus des principes abstraits d'organisation mentale, les distinctions entre l'I.A., la psychologie et même la philosophie de l'esprit semblent se dissoudre. Que l'on étudie ces principes fondamentaux à l'aide d'outils et de techniques empruntés à l'informatique, selon les méthodes de la psychologie expérimentale ou en termes philosophiques traditionnels, le sujet reste le même. Un gigantesque mariage interdisciplinaire semble donc imminent ; en fait un certain nombre d'enthousiastes ont déjà dit "oui" et ont baptisé le nouveau domaine unifié science cognitive. À en croire les bans de mariage, l'intelligence artificielle et la psychologie, de même que certaines branches de la philosophie, de la linguistique et de l'anthropologie, ne sont plus désormais que des "sous-disciplines" d'une discipline unique qui englobe l'étude de la connaissance, de l'intelligence et de l'esprit, c'est-à-dire de la manipulation des symboles" [Haugeland, 1989][p. 10-11].

3.2. Les limites de l'intelligence artificielle et l'offensive anti-computationnelle selon Searle

Pourtant la traduction automatique dans les années 60 et les systèmes experts à partir de la fin des années 80, pour ne citer qu'eux, ont montré des limites. La science réussissait à envoyer des hommes sur la lune. Elle échouait à réaliser des machines qui feraient face à des situations que tout homme affronte sans difficulté apparente. On a alors commencé à découvrir que l'incapacité principale consistait à réaliser des systèmes artificiels performants dans des environnements complexes et variables. Progressivement on a assisté à une remise en cause de la logique comme mode unique de fonctionnement de l'intelligence humaine. On s'est rendu compte par exemple que les "règles de raisonnement" dont on devait accoucher les experts ne

fonctionnaient pas toujours selon un schéma logique et qu'elles n'étaient ni forcément explicitables, ni tout à fait cohérentes dans un cadre purement logique.

L'attaque lancée contre l'I.A. classique a donc porté en bonne partie contre l'ontologie sous-jacente et implicite, héritée de la cybernétique. La remise en cause des positions classiques dans les années 70-80 doit se comprendre comme une attaque ontologique contre les positions implicites des scientifiques qui considéraient les trois identifications précédentes comme acquises et évidentes.

La conception selon laquelle ni le monde ni l'esprit ne se réduisent aux systèmes symboliques se trouve déjà chez Wittgenstein. La conception de Wittgenstein, ou du moins de celui que l'on a coutume d'appeler le second Wittgenstein, ne se réduit donc pas du tout à l'impression que peut laisser le *Tractatus* [Wittgenstein, 1986]. Pour lui, selon les termes de Pierre Lévy, "le monde strictement calculable et communicable se réduit à une mince pellicule de l'être, celle des événements, parcourue par le jeu des opérations et des traductions" [Lévy, 1985b][p. 277]. Ainsi le rôle de la logique devient tout autre. Celle qui devait nous montrer les structures du monde, car elle semblait leur correspondre, n'est plus considérée que comme un ensemble d'énoncés tautologiques qui ne nous apprennent rien sur le monde. "Les propositions de la logique, ainsi vidées de leur sens proprement logique — comme celles de la géométrie l'étaient de leur sens proprement géométrique — deviennent donc des formes pures : simples tautologies, comme l'entendra Wittgenstein, c'est-à-dire énoncés qui ne disent rien sur le réel mais qui, pour cette raison, demeurent valables, quel que soit le contenu concret qu'on y verse" [Blanché, 1990][p. 71].

C'est également une position ontologique que préconise Searle. À l'ontologie défendue par l'I.A., Searle oppose une contre-ontologie qui refuse que la pensée se réduise à la manipulation symbolique. "Une tendance dominante de la philosophie de l'esprit et de la science cognitive fut de supposer que le calcul⁴ est une caractéristique intrinsèque du monde et qu'on peut, en quelque sorte, éliminer la conscience et l'intentionnalité, soit en faveur de quelque chose d'autre, soit parce qu'elles sont relatives à l'observateur, ou les réduire à quelque chose de plus fondamental comme le calcul. Dans ce livre je soutiens que c'est exactement l'inverse : la conscience et l'intentionnalité sont intrinsèques et inéliminables, et le calcul - hormis les rares cas où il est bel et bien effectué par un esprit conscient - est relatif à l'observateur" [Searle, 1995][p. 16]. Searle parle de caractéristiques intrinsèques du monde et indépendantes de l'observateur. Il affirme un statut particulier de ces caractéristiques. L'attitude de Searle est donc bien une prise de position ontologique, car elle statue sur la nature des choses indépendantes de tout observateur. Pour s'opposer à l'ontologie souvent implicite de l'I.A., Searle défend explicitement une autre position ontologique.

⁴Searle utilise le terme de computation qu'il faut comprendre dans le sens de Turing, c'est-à-dire de manipulation symbolique.

De même que les positions ontologiques de l'I.A. entraînaient une conception méthodologique spécifique, la position de Searle a également des incidences méthodologiques. “Parce que c'est une erreur que de supposer que l'ontologie du mental est objective, c'est une erreur que de supposer que la méthodologie d'une science de l'esprit doive s'occuper exclusivement du comportement objectivement observable” [Searle, 1995][p. 43]. C'est bien pour contrer une position ontologique que Searle attaque une conception méthodologique qui lui est attachée. Searle attaque ici l'ensemble des conceptions réduisant l'esprit à des comportements. La méthodologie en termes d'entrées/sorties que proposaient Wiener, Rosenblueth et Bigelow dans [Wiener *et al.*, 1943] est donc complètement rejetée parce qu'elle prétend épuiser les phénomènes liés à l'esprit. De même, par ses affirmations, Searle refuse la conception de Turing [Turing, 1948] selon laquelle l'intelligence se réduit à des comportements intelligents. Searle pense qu'une autre méthodologie est envisageable qui ne se contente pas de chercher à comprendre l'esprit par l'intermédiaire des comportements observables. Il écrit : “c'est une erreur que de supposer que nous ne connaissons l'existence des phénomènes mentaux chez les autres qu'en observant leur comportement” [Searle, 1995][p. 45]. Searle n'est cependant pas très précis lorsqu'il s'agit de proposer et de mettre en place une autre méthodologie s'appuyant sur ses propres conceptions ontologiques.

En fait on assiste à une attaque générale contre le formalisme, l'axiomatique et plus généralement la manipulation symbolique en sciences sur plusieurs plans. La pensée est un des domaines privilégiés dans lesquels cette contre-ontologie qui refuse la réduction aux formalismes s'est développée. Searle en est un de ses représentants les plus farouches.

3.3. En résumé

Finalement les sciences cognitives s'appuient sur un socle où certaines options théoriques ont déjà été prises quant au statut de la formalisation. Ce choix, c'est celui de Turing, déformé et simplifié par ses successeurs. Ce que nous avons voulu montrer ici, c'est qu'un autre choix était envisageable dès l'époque de von Neumann. On observe que ces deux attitudes, qui constituent deux types de relations entre les formalisations et les sciences auxquelles on les applique, existent encore à l'heure actuelle. Grossièrement, on trouve d'un côté des scientifiques qui se contentent d'analogies et de l'autre des scientifiques qui sont convaincus que les relations observées sont plus que de simples analogies.

Le premier courant est un choix volontaire qui consiste à considérer les disciplines auxquelles on applique la formalisation comme une possibilité de proposer des problèmes intéressants. Cette option est généralement celle des mathématiciens⁵.

⁵On peut citer comme exemple les travaux de Morel ou de Mumford, dans le domaine de la

Selon une autre version de ce choix, on étudie les relations entre le formalisme et les phénomènes d'un point de vue simplement analogique. C'est généralement celle des ingénieurs qui cherchent une source d'inspiration pour construire des outils puissants, et dont l'objectif est essentiellement la performance de l'outil qu'ils ont développé⁶. Le point commun à tous dans ce courant est de considérer le phénomène comme un point de départ anecdotique et une simple source d'inspiration. En retour, ces personnes n'ont généralement pas la prétention d'affirmer quoi que ce soit sur l'analyse du phénomène lui-même, à partir des outils qu'ils ont développés.

Le second courant, quant à lui, assimile le formalisme et sa réalisation matérielle. Il reste aujourd'hui encore tout à fait majoritaire en intelligence artificielle. On constate que ce postulat légitime en bonne partie les travaux dans cette discipline. Il justifie que l'on porte son attention exclusivement sur la machine qui permet la matérialisation, l'ordinateur.

Il revient aux sciences cognitives de réfléchir de nouveau aux relations qu'elles souhaitent établir entre les phénomènes qu'elles étudient et les systèmes formalisés. Ce que nous souhaitons vivement est une réhabilitation du modèle. Cependant, nous voulons proposer une notion de modèle qui, comme en physique, s'intègre dans une démarche générale et ne se limite ni à la construction de systèmes abstraits intéressants mathématiquement, ni à des constructions matérielles sur ordinateurs qui auraient pour fondement une confusion entre l'abstraction et sa matérialisation. Ainsi, aux origines des sciences cognitives, on trouve deux voies différentes pour les relations entre les formalismes et l'étude du cerveau. Cependant, plus en amont, ces deux démarches s'inscrivent dans un projet général d'Hilbert de relation entre les systèmes formels et les phénomènes. Tout en restant dans une optique qui recherche la formalisation, nous voulons montrer ici qu'une troisième voie est possible qui attribue à la formalisation un rôle tout aussi important mais qui permette des relations plus étroites mais moins globales avec les phénomènes étudiés.

On peut considérer que les sciences cognitives sont nées en bonne partie de ce désir d'unifier les travaux sur l'esprit et le cerveau humain, en allant au-delà des limitations que les études formalisées avaient pu montrer. Toute la difficulté consiste alors à redéfinir le rôle des formalismes. Ce rôle doit pouvoir être tenu indépendamment de toute position ontologique. Ceci afin de pouvoir tolérer les trois identifications précédentes, sans pour autant les présupposer.

vision, qui se placent bien dans cette optique.

⁶C'est par exemple le cas des travaux de Faugeras, concernant également la vision.

RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 1

- [Andler, 1996] ANDLER D. *Turing : pensée du calcul, calcul de la pensée*. CRÉA, École Polytechnique, Palaiseau, 1996. Rapport technique 9611.
- [Arbib, 1987] ARBIB M. A. *Brains, machines and mathematics*. Deuxième édition. Springer, New York, 1987. 202 p. Première édition 1964. ISBN 0-387-96539-4 3-540-96539-4.
- [Blanché, 1990] BLANCHÉ R. *L'axiomatique*. Première édition Quadrige. Presses Universitaires de France, Paris, 1990. 108 p. Collection Quadrige. Première édition 1955. ISBN 2-13-043141-0. ISSN 0291-0489.
- [Cartan, 1943] CARTAN H. Sur le fondement logique des mathématiques. *La revue scientifique*. 1943, vol. LXXXI, no 3216, p. 3–11.
- [Cavaillès, 1938] CAVAILLÈS J. *Remarques sur la formation de la théorie abstraite des ensembles*. Première édition. Hermann, Paris, 1938. Volumes 606 et 607 des Actualités scientifiques et industrielles. Deux tomes, I/ Préhistoire, la création de Cantor (107 p.), II/ Dedekind, les axiomatisations (Zermelo, Fraenkel, von Neumann) (51 p.).
- [Cossa, 1957] COSSA P. *La cybernétique : du cerveau humain aux cerveaux artificiels*. Deuxième édition. Masson, Paris, 1957. 102 p. Collection Évolution des sciences.
- [de Broglie, 1953] DE BROGLIE L. Sens philosophique et portée pratique de la cybernétique. *Nouvelle revue française*. 1953, p. 60–85.
- [Dupuy, 1985] DUPUY J.-P. L'essor de la première cybernétique. *Cahiers du CRÉA*. 1985, no 7, p. 9–139.
- [Dupuy, 1994] DUPUY J.-P. *Aux origines des sciences cognitives*. Première édition. La découverte, Paris, 1994. 188 p. Collection Textes à l'appui/série sciences cognitives. ISBN 2-7071-2200-9.
- [Gödel, 1931] GÖDEL K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*. 1931,

- vol. 38, p. 173–198. Reproduit dans les Collected works de Kurt Gödel. S. Feferman, J. Dawson, S. Kleene, G. Moore, R. Solovay, J. van Heijenoort, éditeurs. Texte avec la version originale en allemand et une traduction en anglais. Volume I, p. 144-195. Oxford University Press, New York, 1986. ISBN 0-19-503964-5.
- [Goldstine, 1993] GOLDSTINE H. H. *The computer: from Pascal to von Neumann*. Cinquième édition. Princeton University Press, Princeton (New Jersey), 1993. 378 p. Première édition 1972. ISBN 0-691-02367-0.
- [Haugeland, 1989] HAUGELAND J. *L'esprit dans la machine : fondements de l'intelligence artificielle*. Première édition française. O. Jacob, en collaboration avec les éditions du Seuil, Paris, 1989. 256 p. Traduction de Artificial intelligence, the very idea. MIT Press, Cambridge (Mass.), 1985. ISBN 2-7381-0054-6.
- [Heims, 1980] HEIMS S. J. *John von Neumann and Norbert Wiener: from mathematics to the technologies of life and death*. MIT Press, Cambridge (Mass.), 1980. 547 p. ISBN 0-262-58056-X.
- [Hilbert, 1971] HILBERT D. *Les fondements de la géométrie*. Première édition Dunod. Dunod, Paris, 1971. 311 p. Traduction de Grundlagen der Geometrie. Première édition allemande Teubner, 1899.
- [Hilbert, 1990] HILBERT D. *Sur les problèmes futurs des mathématiques : les 23 problèmes*. Première édition Gabay. J. Gabay, Paris, 1990. 60 p. Réimpression de Compte rendu du deuxième congrès international de mathématiques tenu à Paris du 6 au 12 août 1900, Gauthier-Villars, 1902. ISBN 2-87647-037-3. ISSN 0989-0602.
- [Hilbert et Ackermann, 1950] HILBERT D. ET ACKERMANN W. *Principles of mathematical logic*. Chelsea, New York, 1950. 172 p. Traduction de la seconde édition de Grundzüge der theoretischen Logik, Springer, Berlin, 1938. Première édition allemande 1928.
- [Lévy, 1985a] LÉVY P. L'œuvre de Warren McCulloch. *Cahiers du CRÉA*. 1985, no 7, p. 211–255.
- [Lévy, 1985b] LÉVY P. Wittgenstein et la cybernétique. *Cahiers du CRÉA*. 1985, no 7, p. 257–285.
- [Lévy, 1987] LÉVY P. *La machine univers : création, cognition et culture informatique*. La découverte, Paris, 1987. 220 p. Collection Sciences et société. ISBN 2-7071-1661-0.
- [McCulloch, 1951] MCCULLOCH W. S. Why the mind is in the head. *Cerebral mechanisms in behavior - the Hixon symposium, Pasadena, 1948*. 1951, p. 42–111. L. Jeffress, éditeur. J. Wiley, New York. Reproduit dans les Collected works

de Warren McCulloch intitulés *Embodiments of mind*. p. 72-141. MIT Press, Cambridge (Mass.), 1965.

[McCulloch, 1961] MCCULLOCH W. S. What is a number, that a man may know it, and a man, that he may know a number? (the ninth Alfred Korzybski memorial lecture). *General Semantics Bulletin*. 1961, no 26 et 27, p. 7-18. Reproduit dans les Collected works de Warren McCulloch intitulés *Embodiments of mind*. p. 1-18. MIT Press, Cambridge (Mass.), 1965.

[McCulloch et Pitts, 1943] MCCULLOCH W. S. ET PITTS W. H. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*. 1943, vol. 5, p. 115-133. Reproduit dans les Collected works de Warren McCulloch intitulés *Embodiments of mind*. p. 19-39. MIT Press, Cambridge (Mass.), 1965.

[Rudin, 1995] RUDIN W. *Analyse fonctionnelle*. Première édition française. Édissience International, Paris, 1995. 412 p. Traduction de Functional analysis. Première édition américaine, McGrawHill, New York, 1973. ISBN 2-84074-113-X.

[Ruyer, 1954] RUYER R. *La cybernétique et l'origine de l'information*. Flammarion, Paris, 1954. 236 p. Collection de la Bibliothèque de philosophie scientifique.

[Searle, 1995] SEARLE J. *La redécouverte de l'esprit*. Première édition française. Gallimard, Paris, 1995. 364 p. Collection nrf essais. Traduction de The rediscovery of mind. Première édition américaine MIT Press, Cambridge (Mass.), 1992. ISBN 2-07-073641-5.

[Turing, 1936] TURING A. M. On computable numbers with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*. 1936, vol. 42, p. 230-265. Corrections, ibid. 1937, vol. 43, p. 544-546. Reproduit dans The undecidable: basic papers on undecidable propositions, unsolvable problems and computable functions. M. Davis, éditeur. p. 116-154. Raven Press, Hewlett (New Jersey), 1965.

[Turing, 1948] TURING A. M. Intelligent machinery. 1948. Rapport rédigé pour le National Physical Laboratory à la fin de l'été 1948. Première publication dans Cybernetics: key papers. C. Evans et A. Robertson, éditeurs. Baltimore, University Park Press, 1968. Reproduit dans les Collected works d'Alan Turing. D. Ince, éditeur. Mechanical intelligence, p. 107-128. North-Holland, Amsterdam, 1992. ISBN 0-444-88058-5.

[Turing, 1950] TURING A. M. Computing machinery and intelligence. *Mind*. 1950, vol. 59, no 236, p. 433-460. Reproduit dans les Collected papers d'Alan Turing. D. Ince, éditeur. Mechanical intelligence, p. 133-160. North-Holland, Amsterdam, 1992. ISBN 0-444-88058-5.

[Ulam, 1958] ULAM S. John von Neumann 1903-1957. *Bulletin of the American Mathematical Society*. 1958, vol. 64, no 3, part II, p. 1-49.

[Ullmo, 1969] ULLMO J. *La pensée scientifique moderne*. Flammarion, Paris, 1969. 315 p. Volume 92 de la Collection Champs philosophie.

[von Neumann, 1925] VON NEUMANN J. Eine Axiomatisierung der Mengenlehre. *Journal für die reine und angewandte Mathematik*. 1925, vol. 154, p. 219-240. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume I, Logic, theory of sets and quantum mechanics, p. 34-57. Pergamon Press, Oxford, 1961-63.

[von Neumann, 1928] VON NEUMANN J. Die Axiomatisierung der Mengenlehre. *Mathematische Zeitschrift*. 1928, vol. 27, p. 669-752. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume I, Logic, theory of sets and quantum mechanics, p. 339-422. Pergamon Press, Oxford, 1961-63.

[von Neumann, 1945] VON NEUMANN J. *First draft of a report on the EDVAC*. 1945. Rapport technique. Texte non reproduit dans les Collected works de John von Neumann.

[von Neumann, 1951] VON NEUMANN J. The general and logical theory of automata. *Cerebral mechanisms in behavior - the Hixon symposium, Pasadena, 1948*. 1951, p. 1-31. L. Jeffress, éditeur. J. Wiley, New York. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 288-328. Pergamon Press, Oxford, 1961-63.

[von Neumann, 1956] VON NEUMANN J. Probabilistic logics and the synthesis of reliable organisms from unreliable components. Dans *Automata studies*, Shannon C. et McCarthy J., éditeurs, p. 43-98. Princeton University Press, Princeton (New Jersey), 1956. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 329-378. Pergamon Press, Oxford, 1961-63.

[von Neumann, 1988] VON NEUMANN J. *Les fondements mathématiques de la mécanique quantique*. Première édition Gabay. J. Gabay, Paris, 1988. 338 p. Réimpression de la traduction publiée par F. Alcan, 1946. Première édition allemande, Springer, Berlin, 1932. ISBN 2-87647-047-0.

[von Neumann, 1992] VON NEUMANN J. *L'ordinateur et le cerveau*. Première édition française. La découverte, Paris, 1992. 130 p. Collection Textes à l'appui/série sciences cognitives. Traduction de The computer and the brain - the Silliman lectures. Première édition américaine, Yale University Press, New Haven, 1958. ISBN 2-7071-2164-9.

- [Wiener, 1952] WIENER N. *Cybernétique et société*. Les deux rives, Paris, 1952. 296 p. Traduction de Cybernetics and society: The human use of human beings.
- [Wiener, 1961] WIENER N. *Cybernetics or control and communication in the animal and the machine*. Deuxième édition. MIT Press, Cambridge (Mass.), 1961. 212 p. Texte comprenant l'édition originale de 1948, avec une préface et deux chapitres supplémentaires.
- [Wiener et Rosenblueth, 1945] WIENER N. ET ROSENBLUETH A. The role of models in science. *Philosophy of science*. 1945, vol. 12, p. 316–322.
- [Wiener *et al.*, 1943] WIENER N., ROSENBLUETH A. ET BIGELOW J. Behavior, purpose and teleology. *Philosophy of science*. 1943, vol. 10, p. 18–24.
- [Wittgenstein, 1986] WITTGENSTEIN L. *Tractatus logico-philosophicus. Suivi de Investigations philosophiques*. Gallimard, Paris, 1986. 378 p. Volume 109 de la collection Tel. Première édition anglaise 1918. ISBN 2-07-070773-3.
- [Zermelo, 1904] ZERMELO E. Beweis daß jede Menge wohlgeordnet werden kann. *Mathematische Annalen*. 1904, vol. 59, p. 514–516.
- [Zermelo, 1908] ZERMELO E. Untersuchungen über die Grundlagen der Mengenlehre. *Mathematische Annalen*. 1908, vol. 65, p. 261–281.

CHAPITRE 2

**Comment les sciences cognitives
conçoivent l'étude des processus**

INTRODUCTION DU CHAPITRE 2

L'objectif de ce chapitre est de discuter et d'expliquer certaines caractéristiques des sciences cognitives. Nous considérons *a priori* qu'il est possible de définir les sciences cognitives comme une discipline à part entière. Elle possèderait de ce fait une certaine autonomie par rapport aux disciplines qui l'ont formée initialement. Le but est alors de mieux cerner ses caractéristiques. Ceci étant fait, nous pourrons comprendre les contraintes imposées à la modélisation et le rôle que les formalismes peuvent prendre dans cette discipline. Nous avons déjà vu quelles raisons ont expliqué historiquement l'apparition des sciences cognitives. Il reste à comprendre quels sont les objectifs et les moyens que se donnent ces sciences. Comment des connaissances issues de disciplines différentes vont pouvoir être prises en compte simultanément et de façon étroite ?

Les ouvrages portant sur les sciences cognitives insistent essentiellement sur la diversité des études portant sur la cognition et peu sur leurs points communs ou sur ce qui fait l'unité de ces sciences. Les sciences cognitives sont alors présentées en extension, par l'ensemble des travaux qui la composent. Il est difficile de concilier une vision synthétique et la présentation de recherches suffisamment détaillées.

Sans développements mathématiques ou formels, les travaux sont vite considérés comme insuffisants. Pourtant, une science qui débute a souvent besoin de réfléchir sur ses objectifs, les moyens qu'elle se donne et les contraintes qu'elle s'impose. Avant de développer des modèles formels, il faut préciser dans la langue naturelle l'ensemble des hypothèses qui sont faites. Qu'est-ce que la cognition ? Peut-on identifier le cerveau, l'esprit et la pensée ? La cognition se limite-t-elle aux mécanismes du cerveau, ou faut-il l'étendre davantage, par exemple au système hormonal, ou au corps dans son ensemble ? Les questions de ce type, fondamentales pour la discipline, sont nombreuses. Des réponses même partielles sont un préalable nécessaire à toute formalisation. Certains scientifiques, comme Newell¹, ont pris conscience qu'une synthèse était possible. Le cas reste cependant isolé et le cadre proposé par Newell est loin d'être unanimement accepté. Pour Gerald Edelman une conception synthétique devient même une nécessité pour étudier la cognition. Il écrit par exemple que "toute explication [des catégories perceptuelles, des représentations, de la mémoire, etc...] doit tenir compte de l'organisme biologique dans son ensemble, de

¹Son objectif dans ses William James Lectures de 1987 [Newell, 1990] est de montrer que l'on peut concevoir la cognition à partir d'un unique ensemble de mécanismes. Newell illustre sa position avec la présentation de son architecture SOAR, un système d'intelligence artificielle qui doit rendre compte de certaines caractéristiques de la cognition humaine. SOAR est censé être capable d'opérer en temps réel, condition nécessaire pour un comportement intelligent. SOAR est capable, selon l'auteur, de résoudre des tâches nécessitant des réponses immédiates et la reconnaissance d'objets. On peut critiquer l'effectivité des résultats obtenus. Il faut noter cependant la volonté affichée de considérer la cognition comme pouvant être étudiée pour elle-même.

son ancrage originel dans l'évolution, de son développement et des relations avec son environnement” [Reeke et Edelman, 1987][p. 44]. Aucune alternative unifiée ne vient cependant prendre la relève et on ne sait pas très bien actuellement, pourquoi, ni comment des travaux si divers peuvent être considérés et présentés dans un domaine de recherche unifié. Nous avons déjà vu que des raisons historiques expliquent la situation actuelle². Il s'agit désormais d'essayer d'établir quels chemins permettent une éventuelle synthèse.

Le point commun affirmé des différentes approches est lié à des propriétés particulières de la cognition, objet(s) d'étude des sciences cognitives. Le sens du terme **cognition** n'est quant à lui quasiment jamais précisé, mais considéré comme bien connu de tous. Dans son paragraphe introductif intitulé “la définition introuvable” de son article *Calcul et représentation : les sources*, qui sert d'introduction au recueil [Andler, 1992], Andler cherche à délimiter le champ des sciences cognitives. Il propose comme définition, avec quelques réserves, “les sciences cognitives ont pour objet de décrire, d'expliquer et le cas échéant de simuler les principales dispositions et capacités de l'esprit humain – langage, raisonnement, perception, coordination motrice, planification ...”. Dans la suite de son article, Andler propose une désignation de la discipline par une description des deux projets actuellement concurrents et de certains de leurs présupposés : le cognitivisme et le connexionnisme. Cette analyse comparée des écoles et des programmes de recherche différents ne permet pourtant pas de se faire une idée claire des limites que l'on fixe à la discipline, ni de son objet d'étude. Une présentation qui intégrerait une liste de sujets semblerait également réductrice. Il est très difficile de définir l'intelligence ou la vie par une liste de propriétés qui se voudrait exhaustive. De même une telle liste ne suffirait pas à définir la cognition.

²Voir le chapitre 1, ou le Rapport & Documents du CREA n°9612.

1. DIVERSITÉ ET UNITÉ DES SCIENCES COGNITIVES

On peut considérer, dans un premier temps, que les sciences cognitives ont pour objet d'étude le fonctionnement du cerveau/esprit humain. Cette première définition permet déjà d'expliquer la diversité des disciplines concernées.

1.1. La diversité des disciplines concernées et l'organisation autour de thèmes

1.1.1. La diversité des disciplines concernées

Si l'on considère que l'objet d'étude des sciences cognitives est le cerveau/esprit, il apparaît immédiatement que de nombreuses disciplines autonomes s'y intéressent déjà et vont donc bien sûr en relever. Les sciences cognitives sont alors à la jonction de six disciplines : l'intelligence artificielle, la linguistique, la logique et les mathématiques, la philosophie, la psychologie et les neurosciences.

Le livre rédigé sous la direction de Daniel Andler [Andler, 1992] est lui-même divisé en six chapitres correspondant aux disciplines en jeu dans les sciences cognitives. Dans ce livre, les différentes parties s'articulent autour d'une problématique propre aux sciences cognitives, qui fait appel en fait à plusieurs disciplines déjà consacrées. On y retrouve, pour mémoire, *le cerveau, problème et modèle* (1), *la modularité, hypothèse psychologique* (2), *langage et communication* (3), *la formation des concepts et des théories* (4), *cerveau, esprit et représentation mentale* (5), *individu et société* (6).

Le cerveau/esprit est au cœur de l'ouvrage, comme il semble au centre des sciences cognitives. Les travaux vont à la fois porter sur le cerveau comme objet physique, mais aussi comme support pour des mécanismes mentaux. Le cerveau est déjà étudié, de façon privilégiée, comme objet physique et lieu de réactions physico-chimiques, par la biologie et les neurosciences. Les mécanismes issus du fonctionnement du cerveau, comme le langage ou la formation des concepts, relèvent également des sciences cognitives. Celles-ci font appel à la linguistique, à la psychologie et à la philosophie du langage et de la connaissance (épistémologie) qui ont depuis longtemps mis ces problèmes en évidence et au cœur de leurs réflexions.

La linguistique, la psychologie et la philosophie ne se réduisent pas pour autant aux aspects qui intéressent les sciences cognitives. Ces disciplines conservent leur unité et leur autonomie. On constate simplement que les sciences cognitives, en constitution, mettent au cœur de leurs interrogations des problèmes qui ont déjà été posés par d'autres disciplines, même si cela a pu se faire en d'autres termes. Les sciences cognitives ne peuvent simplement pas se permettre de délaisser des réflexions plus anciennes qui ont pu porter sur des problèmes identiques. Les tentatives de reconstruction artificielle et matérielle des propriétés attribuées au cerveau relèvent également des sciences cognitives. On peut donc intégrer dans ce domaine certains aspects de l'intelligence artificielle. Les mathématiques, enfin, sont nécessaires pour élaborer des modèles formels. Six domaines scientifiques sont ainsi mis en évidence afin d'étudier le cerveau/esprit. Les sciences cognitives se trouvent alors à la jonction de ces disciplines.

On peut noter cependant que cet ouvrage fait également mention de la relation entre l'individu et la société¹. Ainsi, certains phénomènes sociaux pourront relever des sciences cognitives. Certains se demandent, par exemple, si la signification linguistique est réductible à l'individu et ne doit pas tenir compte des relations interindividuelles. Dans ce contexte, les sciences sociales ont aussi un rôle à jouer pour étudier la cognition. L'objet d'étude des sciences cognitives, le cerveau/esprit, doit donc être compris dans un sens large. Il ne s'agit pas seulement d'étudier le cerveau/esprit mais aussi les cerveaux/esprits. Une définition qui se contenterait d'évoquer le cadre biologique correspondrait à une attitude déjà orientée et réductrice. On éliminerait de cette façon presque immédiatement les sciences sociales du champ des sciences cognitives. Les sciences cognitives font en fait appel à un nombre de thèmes importants qu'il est difficile de synthétiser. Issues d'un renouveau nécessaire d'une analyse scientifique de l'intelligence, les sciences cognitives constituent actuellement un agrégat. Le désir de fédérer certains aspects de ces disciplines se fait sentir. Il est cependant difficile à l'heure actuelle d'affirmer que les discours très hétérogènes qui s'en réclament sont effectivement en train d'être synthétisés.

1.1.2. L'organisation autour de thèmes

Le numéro spécial du courrier du CNRS de 1992 [CNR, 1992] montre bien la diversité des thèmes étudiés qui relèvent des sciences cognitives. Ce numéro constitue un bon panorama de l'actualité de la recherche en France en sciences cognitives. Ce panorama est lui-même assez proche de la situation générale des sciences cognitives en Europe et dans le monde². La description se situe ici à un niveau sociologique.

¹On peut citer, entre autres, l'article de Sperber, dans la partie 6 de l'ouvrage [Andler, 1992], intitulé *les sciences cognitives, les sciences sociales et le matérialisme*.

²Il faut noter cependant que la situation semble changer rapidement dans le monde anglo-saxon. Désormais de nombreux centres de recherche de sciences cognitives aux Etats-Unis et même en

Ceci est tout à fait nécessaire, si on souhaite esquisser les directions que la communauté des chercheurs peut choisir, aujourd’hui, pour orienter ses travaux. Parmi les questions que se posent naturellement ces chercheurs, il faut noter, en particulier, celles relatives à la méthodologie utilisée en sciences cognitives. Dans ce cadre, nous nous intéresserons plus particulièrement au rôle de la modélisation, et plus généralement des systèmes symboliques.

Dans cette revue [CNR, 1992], on trouve de nombreux thèmes transversaux qui font intervenir plusieurs disciplines, sans jamais s'épuiser dans une seule d'entre elles. Sont cités *la parole et la lecture*, *le raisonnement*, *les différentes étapes de la perception à l'action*, *la mémoire*, *l'attention et l'apprentissage*, *le développement cognitif*. Ceci montre bien que, dans la pratique, les recherches s'organisent autour de thèmes et de projets susceptibles, sans doute, de fédérer des équipes venant d'horizons différents. Autour de chacun de ces thèmes des projets de construction localisés se constituent.

On peut donc noter qu'il est reconnu que les sciences cognitives sont pluridisciplinaires et font appel à des domaines relevant à la fois des sciences “exactes”, des sciences du vivant et des sciences “humaines”. Le champ couvert est donc très large. Les sciences cognitives ne peuvent et ne doivent alors pas se définir par une appropriation de ces différentes disciplines. La diversité des équipes de recherche correspond simplement à la diversité des thèmes qui relèvent des sciences cognitives et aux multiples points de vues que l'on peut en avoir.

1.2. L'étude de processus, élément principal de l'unité des sciences cognitives

1.2.1. La cognition comme ensemble de processus

Il est facile aujourd’hui d’insister sur la diversité des sciences cognitives. Il est par contre beaucoup plus difficile de mettre en évidence l’unité de ces sciences. Il reste en effet à montrer quels liens existent entre tous ces travaux. Toutes ces études cherchent, *in fine*, à savoir comment il est possible pour des êtres vivants d’appréhender le monde qui leur est extérieur. Ils relèvent donc d’une étude de la cognition. Le terme de **cognition** se définit très largement, comme *l’ensemble des moyens, des méthodes qui permettent à un être vivant de connaître le monde*. La difficulté de cette définition est d’utiliser des termes généraux dont il est difficile de limiter l’acception. Par exemple, nous ne précisons pas ce que nous entendons par “être vivant” et nous ne délimitons pas le “champ du vivant”. De même nous ne savons pas ce que “connaître” veut dire, ni ce qu’est “le monde”. Il est alors difficile

Angleterre organisent leurs projets pluridisciplinaires autour de problématiques liées à l’étude de la cognition.

de savoir si nous étudions ou non un “moyen de connaître le monde”, c'est-à-dire d'affirmer que nous sommes dans le champ d'étude des sciences cognitives. Nous nous limiterons pour l'instant à une compréhension intuitive du mot cognition à partir de la définition précédente.

Un point important ressort de cette définition, malgré ses insuffisances. Lorsque nous parlons d'un “moyen de connaître le monde”, nous évoquons quelque chose d'actif, de changeant. La cognition est toujours considérée comme une activité, comme un acte. Un outil n'est pas en lui-même un moyen de connaître le monde. Par contre l'action d'agir avec cet outil le sera éventuellement. Ce “moyen de connaître le monde” nécessite l'utilisation de cet objet, mais il n'est pas cet objet.

Ainsi, lorsque l'on cherche à définir la cognition, on constate qu'elle ne constitue pas un objet d'étude comme les autres. En effet, nous l'avons définie comme un processus, plutôt que comme un objet. Nous entendons par processus une activité par opposition à un objet. Pour être plus précis, nous appelons **objet** *l'abstraction d'un phénomène statique, stable, ou persistant à l'échelle à laquelle on l'observe* et par **processus**, *l'abstraction d'un phénomène dynamique*³. D'après ce que nous venons d'écrire, il faut donc déjà un observateur pour savoir si on parle d'un phénomène statique, stable, persistant ou d'un phénomène dynamique⁴. Étudier des phénomènes et les abstraire comme des objets ou des processus peut donc constituer soit un choix méthodologique, soit une position ontologique.

Le terme de processus est très souvent utilisé en sciences cognitives, à la fois par des mathématiciens, des chercheurs en intelligence artificielle, des psychologues et des philosophes. L'ingénieur-mathématicien von Neumann utilise le terme de processus dans le sens d'une succession d'opérations élémentaires. Le philosophe Searle emploie le terme à la fois pour faire référence à des faits biologiques aussi bien que “mentaux”. Il écrit “nous sommes parvenus à mieux comprendre le caractère biologique des processus qui caractérisent les organismes vivants” [Searle, 1985][p. 31]. Il utilise le terme de processus plus loin à propos du cerveau. “En bref, pour dissiper les mystères, comprenons les processus. Nous ne les comprenons pas encore parfaitement, mais nous comprenons leur caractère général, nous comprenons que certaines activités spécifiques électro-chimiques se produisent parmi les neurones ou les modules de neurones, et peut-être ailleurs dans le cerveau, et que ce sont ces processus qui engendrent la conscience” [Searle, 1985][p. 31]. Les philosophes emploient également le mot dans l'expression “processus mentaux”. Le théoricien

³Il y a en fait un double usage du mot abstraction. Dans cette phrase, pour être correcte il faudrait plutôt dire qu'un processus et un objet sont les résultats d'une abstraction et non l'abstraction elle-même. Selon l'usage et pour ne pas alourdir le texte, nous emploierons par la suite le terme **abstraction** pour désigner soit l'abstraction elle-même, soit le résultat de l'abstraction. Le contexte permettra de préciser le sens choisi.

⁴Nous ne prenons position ici ni sur la possibilité d'étudier des phénomènes “bruts” ni sur le statut des phénomènes.

de la vision artificielle David Marr écrit : “la vision est le *processus*⁵ qui permet de découvrir, à partir des images, ce qui est présent dans le monde et où cela se trouve” [Marr, 1982][p. 3]. On peut encore citer le psychologue Jean-François Le Ny qui, dans son article de l’Encyclopædia Universalis, utilise le terme de processus pour définir l’apprentissage. Ainsi ce mot est employé en sciences cognitives par toutes les disciplines qui y participent.

On pourrait très bien définir les sciences cognitives comme l’étude des processus cognitifs⁶. Personne n’a pour l’instant essayé de préciser ce que le terme de “cognitifs” pouvait ajouter de spécifique à ces processus.

Nous nous intéressons pour l’instant aux contraintes que l’étude de ces processus impose à ces sciences. Dans ce but nous insistons sur la façon dont les chercheurs en sciences cognitives abordent l’étude des processus.

Les phénomènes dont sont abstraits les processus sont dynamiques. Ainsi, de façon presque naturelle, les sciences cognitives vont s’intéresser à des abstractions qui ont des propriétés dynamiques. *Le terme de processus ne présuppose pas non plus le type de dynamique* qui est en jeu. Nous l’avons vu, von Neumann utilise le terme dans le sens d’une succession de combinaisons d’opérations élémentaires. “Il en résulte que toutes sortes de systèmes d’ordres deviennent possibles, qui se modifient sans cesse et qui modifient par conséquent aussi les processus de calcul qui sont également sous leur contrôle. Ainsi des processus plus complexes que de simples itérations deviennent possibles” [von Neumann, 1992][p. 29]⁷. Il faut alors comprendre ces opérations comme abstraites, par exemple des opérations logiques ou arithmétiques.

Le terme de processus ne présuppose pas non plus la forme sous laquelle on va présenter ce processus. On peut le faire en termes mathématiques ou formels, mais aussi bien dans la langue naturelle ou sous forme de dessins. Les mathématiques ne sont donc pas par définition la voie obligatoire pour expliciter les processus. Ce terme fait partie du vocabulaire des mathématiques, dans un sens souvent précis et technique comme celui de “processus markovien”. Cela ne nous oblige pas pour autant à présenter les processus exclusivement sous une forme mathématique.

⁵C'est David Marr qui souligne.

⁶C'est d'ailleurs le choix du Nouveau Petit Robert dans son édition de 1995 qui donne la définition suivante (article **cognitif** [p. 399]) ; **Sciences Cognitives** : ensemble des sciences qui concernent la connaissance et ses processus.

⁷Citation déjà mentionnée, chapitre 1, section 1.2, Obtenir une définition opératoire de la calculabilité : von Neumann et la machine de Turing, section 2 §7.

1.2.2. Les sciences cognitives, les processus et les composants

Pour les sciences cognitives, les phénomènes dynamiques dont on abstrait les processus se constituent sur des phénomènes statiques, que l'on abstrait comme des objets

Pour les sciences cognitives, les phénomènes dynamiques dont on abstrait les processus se constituent sur des phénomènes statiques, que l'on abstrait comme des objets. Les phénomènes statiques n'ont en eux-même aucune importance. On n'en retient que ce qui apporte quelque chose au phénomène dynamique. Cette hypothèse est certainement l'hypothèse centrale des sciences cognitives aujourd'hui. Elle est partagée par tous les chercheurs⁸. Il s'agit alors généralement d'une position ontologique. Il est cependant possible de défendre la même hypothèse en la considérant simplement comme une position méthodologique. Nous verrons par la suite certaines implications de cette position considérée comme méthodologique. Varela défend cette hypothèse centrale dans une position ontologique et il écrit à ce propos : “les systèmes vivants sont des <<machines>>, avec les conséquences suivantes : premièrement, cela implique une approche opposée à celle de l'animisme ; deuxièmement, nous insistons sur le fait qu'un système vivant est défini par son organisation et donc qu'on peut l'expliquer comme on explique n'importe quelle organisation, c'est-à-dire en termes de relations et non à partir des propriétés de ses composants” [Varela, 1989][p. 40].

Les phénomènes dynamiques dont on abstrait les processus sont donc constitués sur des **composants**. Ces composants sont eux-mêmes abstraits comme des objets. D'une part ces composants sont des phénomènes statiques et d'autre part on n'a pas besoin de tenir compte de leurs propriétés. Les propriétés éventuellement pertinentes devront se retrouver dans la description de l'organisation elle-même. Les phénomènes dynamiques pour les sciences cognitives sont donc liés aux modifications de l'organisation de phénomènes statiques, les composants. Le phénomène dynamique dont on abstrait le processus ne peut se manifester comme phénomène dynamique que lorsque des phénomènes statiques lui sont associés que l'on abstraira eux-mêmes comme des objets.

Dans cette conception seuls certains processus sont étudiés, les processus “émergents”. Les **processus émergents** sont alors *les processus abstraits de phénomènes dynamiques qui se constituent à partir de phénomènes statiques*. Le connexionnisme est l'apôtre de cette conception des processus. En effet, les processus étudiés dans le cadre du connexionnisme sont abstraits de phénomènes dynamiques particuliers, des “phénomènes de réseau” ou des “phénomènes de système”. La conception du connexionnisme s'appuie sur des données générales de biologie selon lesquelles la conscience serait un processus abstrait à partir d'un phénomène dy-

⁸On doit pouvoir trouver des exceptions, mais elles sont rares et ne feraient, sans doute, que confirmer ce qui vient d'être dit.

namique qui apparaîtrait comme un produit de l’assemblage du fonctionnement des neurones considérés comme des unités élémentaires. Les processus émergents dans le cadre du connexionnisme sont alors possibles grâce au nombre excessivement élevé de ces neurones. L’étude des gros systèmes d’unités simples en interaction est le cœur de ce que l’on appelle aujourd’hui la “complexité”. En ce sens, le connexionnisme et les sciences cognitives sont en relation étroite avec la “complexité”. Ainsi le connexionnisme et les sciences cognitives sont très liés à la complexité parce qu’ils s’appuient essentiellement sur une explication des processus cognitifs à partir du fonctionnement d’un nombre très élevé d’unités très simples en interaction. La complexité est alors une autre facette des sciences cognitives que nous regarderons plus en détail dans le chapitre 3.

Pour les sciences cognitives, l’étude des processus prime celle des objets et en particulier celle des composants

L’étude de processus n’est pas une spécificité des sciences cognitives. Ce qui les caractérise, c’est d’étudier exclusivement les processus et de laisser de côté ce qu’elles appellent les composants, parce qu’ils sont eux-mêmes statiques. On trouve des processus dans d’autres domaines scientifiques. En physique par exemple, l’étude des transitions de phase [Le Bellac, 1988] est bien l’étude d’un phénomène dynamique. On le traduit comme processus, par exemple comme l’apparition d’une singularité particulière d’une fonction mathématique.

Dans le cas des transitions de phase, cependant, l’ensemble des composants particuliers qui permettent le processus est le véritable objet d’étude. Ce qui intéresse le physicien avant tout dans le cas d’une transition magnétique, c’est l’étude du matériau lui-même, par exemple le nickel. Cela se traduit dans les questions qui ont été posées sur ce métal. On a d’abord essayé d’établir théoriquement que ce métal pouvait présenter plusieurs phases distinctes dans certaines conditions. C’est seulement ultérieurement que l’on a commencé à se demander comment le matériau passait d’une phase à une autre. Or un matériau comme le nickel est considéré comme un objet. L’étude des transitions de phase n’est donc pas faite pour elle-même, comme l’étude d’un phénomène dynamique dont on chercherait à abstraire un processus. Elle constitue avant tout l’analyse d’une des propriétés comportementales de certains matériaux. Le véritable objet d’étude est donc ce métal lui-même considéré comme un objet. Ainsi, la physique étudie des processus, qui sont des abstractions de phénomènes dynamiques, mais qui ne constituent pas l’objet d’étude premier et qui doivent être considérés comme des propriétés particulières des objets étudiés.

Au contraire, en sciences cognitives, ce sont les processus et les phénomènes dynamiques qui nous intéressent, et non un objet ou un phénomène statique qui lui serait sous-jacent. Ainsi ce ne sont pas les neurones biologiques qui intéressent

le chercheur en sciences cognitives, contrairement au biologiste. Le chercheur en sciences cognitives va chercher, par exemple, à comprendre comment l'apprentissage, comme propriété du réseau, est possible avec des neurones.

Le chercheur en sciences cognitives s'intéressera aux propriétés dynamiques observées sur les réseaux de neurones biologiques. Il cherchera à expliquer comment le processus d'apprentissage se constitue à partir des neurones. Il portera finalement peu d'intérêt au substrat, aux composants qui permettent l'apprentissage, c'est-à-dire aux neurones eux-mêmes et à leur milieu dans le cerveau. Les composants sont considérés comme un moyen de constituer le phénomène dynamique, mais il semble que le processus d'apprentissage aurait pu être matérialisé sous la forme d'un autre phénomène dynamique semblable au premier et constitué avec d'autres composants. Ainsi ce qui distingue les sciences cognitives des autres sciences est d'avoir comme objet d'étude principal et premier des processus et non des objets. Cette attitude revient presque naturellement à considérer que le processus n'est pas complètement contraint par le substrat, par les composants qui le permettent.

Une autre définition de la notion de tournant cognitif

Lorsque l'on veut étudier un objet, on va d'abord l'analyser dans sa stabilité, puis on va observer à quels phénomènes dynamiques il est lié. Dans un domaine de recherche, dès que l'on a suffisamment analysé l'objet d'étude de façon statique, on commence à en étudier les aspects dynamiques. On peut citer de nombreux exemples, comme celui des métaux. Ceux-ci étaient d'abord étudiés de façon statique, en analysant leur composition et leur structure. On étudiait par exemple les relations spatiales entre les différents éléments atomiques qui constituent un morceau de métal. On s'intéresse plutôt aujourd'hui aux propriétés dynamiques de ces métaux et en particulier à la façon dont on peut passer d'une phase du métal à une autre. On peut se demander dans quelle mesure ce n'est pas également le sens que l'on peut donner au terme de **tournant cognitif** que rencontraient certaines disciplines. Le tournant cognitif serait alors le moment que rencontre toute discipline lorsqu'elle passe de l'étude des phénomènes statiques à celle des propriétés dynamiques.

Il faut noter que les sciences cognitives vont encore un peu plus loin. Elles considèrent que ces phénomènes dynamiques ne sont pas déterminés par leur substrat. Ainsi elles vont étudier les processus pour eux-mêmes en se désintéressant du substrat particulier à partir duquel ils ont été abstraits.

1.2.3. Les processus : abstraction, généralisation, rematérialisation

Il reste encore à préciser comment on étudie ces processus. Il s'agit d'abord de voir comment ces processus sont effectivement abstraits. De plus nous devons préciser comment on généralise le domaine d'application de ces processus. Enfin

il faut préciser comment la généralisation du domaine permet d'envisager la rematérialisation de ces processus.

Des expériences particulières aux processus : l'abstraction des processus

Pour illustrer cette démarche nous partirons de l'exemple de l'apprentissage. Dans son article de l'Encyclopædia Universalis sur l'apprentissage, Le Ny [Le Ny, 1990] précise que "l'étude méthodique, particulièrement expérimentale, des phénomènes d'apprentissage, [...] a permis d'identifier des processus de caractère général". Ce que l'on qualifie donc d'apprentissage n'est pas un phénomène particulier étudié mais *le processus dégagé, abstrait, avec ses caractéristiques générales*. L'apprentissage est alors identifié à un processus ou à un ensemble de processus. Les phénomènes deviennent alors inutiles. On peut se focaliser sur la construction de *processus génériques*.

Pour chacun des thèmes qui relèvent actuellement des sciences cognitives, on peut noter le même type de raisonnement représenté sur la figure 1. Nous l'illustrerons par le thème de l'apprentissage, représenté figure 2. On part de quelques expériences caractéristiques du thème étudié [Vauclair, 1992]. Nous rapportons ici la description d'une expérience mise au point par Jacques Vauclair et Joël Fagot à Marseille : "un babouin est assis face à un moniteur d'ordinateur sur lequel s'affichent un curseur et une cible colorée. Un minimanche, installé sur un support à hauteur des mains, s'interpose entre l'écran et le singe" [Vauclair, 1992][p. 164]. Ainsi ces expériences sont des phénomènes construits, par exemple le comportement d'un primate que l'on oblige par un système de punitions et de récompenses à réaliser une tâche. "Il faut compter environ quinze jours d'entraînement, à raison de trois heures par jour, pour que le babouin maîtrise la tâche consistant à manipuler le minimanche de telle sorte que les mouvements de ce dernier, isomorphes à ceux du curseur, conduisent à intercepter une cible en mouvement sur l'écran. Naturellement, les rencontres curseur-cible produites par le babouin sont suivies de la distribution d'une récompense alimentaire" [Vauclair, 1992][p. 164]. On abstrait à partir de ces expériences particulières des processus généraux qui leur correspondent (flèche 1). "Initialement, ce dispositif a été conçu pour l'étude des asymétries de performances manuelles dans l'atteinte des cibles. Il m'apparaît cependant que cette situation expérimentale pourrait aussi contribuer à une meilleure connaissance des opérations cognitives sous-tendant la reconnaissance de soi" [Vauclair, 1992][p. 165]. Le processus mis en évidence au cours d'une expérience sera par exemple caractérisé par une augmentation de la vitesse de réponse du primate. Puis on identifie le thème, l'apprentissage à l'ensemble des processus mis en évidence (flèche 2). L'apprentissage sera alors défini, entre autres, comme la capacité à améliorer sa vitesse de réponse, lorsque l'on se retrouve de façon répétée dans des situations similaires. Ainsi les thèmes étudiés en sciences cognitives seront des processus plus que des expériences particulières.

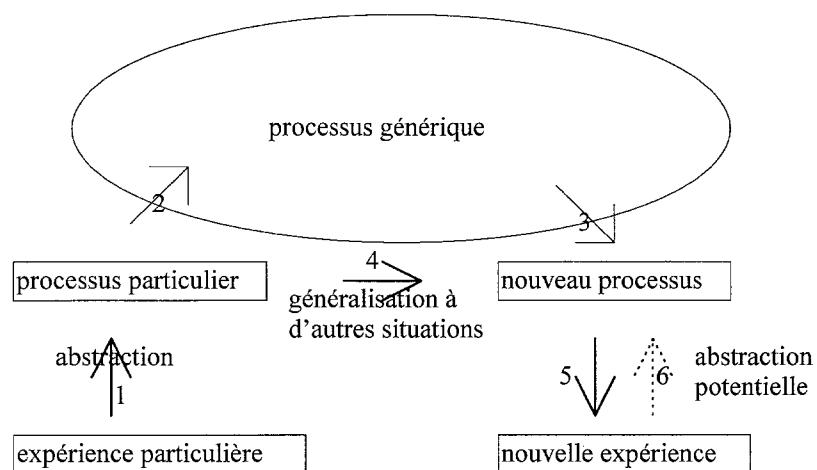


FIGURE 1 : relations entre une expérience, les processus mis en évidence et le thème abordé en sciences cognitives

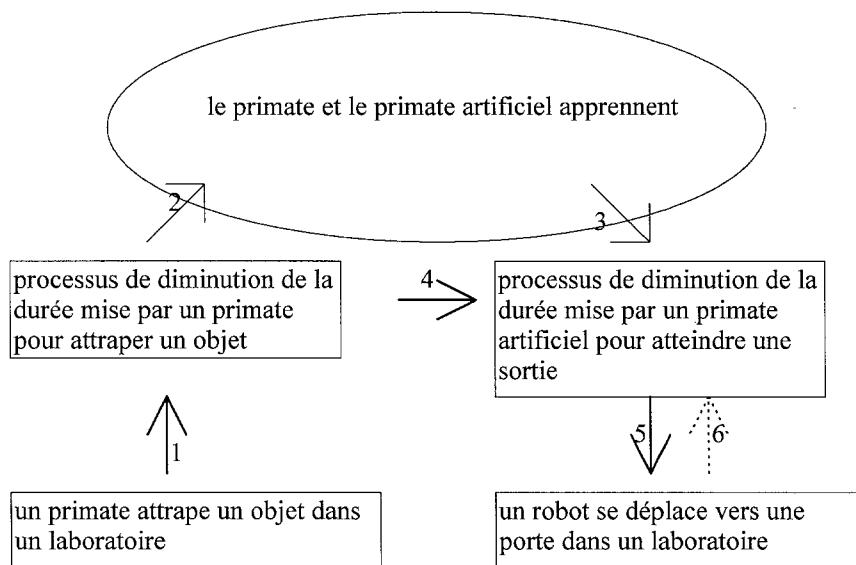


FIGURE 2 : exemple dans le cas de l'apprentissage des relations entre une expérience, les processus mis en évidence et le thème abordé

Chapitre 2 -- Figures 1 et 2

On cherche ensuite d'autres situations où ces processus pourront également être mis en évidence. On se demandera par exemple dans quels contextes on observe qu'un agent améliore ses capacités et nous donnerons le nom d'apprentissage à ces situations. On constate par exemple que l'ensemble des traits caractéristiques des processus d'apprentissage auraient pu également être abstraits à partir d'expériences réalisées sur des primates artificiels (flèche 6). On peut alors intégrer dans le même thème "d'apprentissage", l'amélioration du comportement d'un primate artificiel ou d'un primate (flèche 3). On sent bien qu'il y a eu glissement du sens du terme apprentissage, lorsque l'on est passé du primate au primate artificiel (flèche 4). L'étude de la cognition naturelle est identifiée à celle de la cognition artificielle, la première n'est plus la mesure de la seconde. On ne sait plus alors si la référence doit être l'apprentissage du primate ou celui du primate artificiel.

Cette méthodologie particulière, on la retrouve également dans d'autres disciplines ou sur d'autres sujets, comme l'étude des transitions de phase. Après avoir étudié la transition de phase de l'eau, on a découvert des transitions magnétiques pour certains métaux. Par un mouvement de généralisation, d'abstraction et de théorisation, on a ensuite considéré certaines propriétés en essayant de les formuler de façon indépendante du métal considéré. On abstrayait ainsi de ces phénomènes des processus généraux qui pouvaient caractériser la notion de transition de phase (flèches 1 et 2). On a ensuite trouvé un ensemble d'autres matériaux qui possèdent le même comportement dans certaines situations bien précises. On pouvait donc appliquer ces idées, abstraites à partir d'un métal particulier, à d'autres métaux, à partir desquels semble-t-il, la même méthode d'abstraction aurait pu être appliquée (flèche 6). On peut alors identifier la notion de transition de phase aux processus communs abstraits à partir de ces différentes expériences. Il est alors possible de considérer le nouveau processus abstrait comme une *simple généralisation du domaine d'application de ce processus* (flèche 4). On retrouve ici certains éléments de la méthode présentée sur la figure 1. La justification qui permet de considérer la généralisation comme licite est le fait que les substrats partagent d'autres propriétés que ces propriétés dynamiques. Par exemple, dans le cas des transitions de phase, tous les substrats en jeu ont une structure cristalline et on les classe tous comme des métaux. La possibilité d'apparition d'une transition de phase ne constitue alors qu'un critère supplémentaire pour classer les métaux. Ainsi, en physique, le rapprochement entre des phénomènes dynamiques par l'intermédiaire d'un processus commun est également justifié par le rapprochement de substrats qui partagent d'autres propriétés pas nécessairement dynamiques.

Dans le cas des sciences cognitives, on ne trouve pas de lien aussi serré entre les différents phénomènes mis en relations, puisqu'on abandonne définitivement le substrat. En effet, quasiment par définition du domaine, la seule chose qui doit impérativement relier deux phénomènes, en sciences cognitives, est la capacité d'en rendre compte par un processus identique. Seuls les processus rapprochent les

phénomènes rassemblés et non les substrats. La méthodologie mise en évidence sur la figure 1 pose alors de très grosses difficultés. En effet, les liens entre l'expérience particulière et le processus mis en évidence sont souvent très distendus. Les caractéristiques du processus abstrait ne semblent plus du tout caractériser le phénomène. Par exemple, lorsque l'on parle d'une transition de phase magnétique, on sait déjà *a priori* que l'on parle d'un métal qui va donc avoir certaines propriétés. Lorsque l'on parle de processus en sciences cognitives, le seul liant entre deux phénomènes est ce processus lui-même et on ne sait donc rien dire d'autre *a priori*, pour l'instant en tous cas, sur les phénomènes qui sont en jeu. Une souris apprend, un chat apprend, un robot apprend, un primate artificiel apprend, mais pourquoi pas aussi une plante, une boule de billard, etc..., si on est capable de créer des situations d'où il serait⁹ possible d'abstraire les caractéristiques mises en évidence des processus d'apprentissage.

Lorsque l'on cherche à voir comment le processus mis en évidence peut être abstrait à partir d'autres phénomènes, la correspondance entre les phénomènes reliés semble souvent très superficielle. Dans le cas des sciences cognitives, la généralisation du domaine d'application du processus semble abusive (flèche 4). Elle met presque toujours en avant le scientifique qui doit décider, le plus souvent à partir de critères qui ne sont pas précisés, si les deux phénomènes reliés renvoient bien au même processus. Ainsi c'est *in fine* le scientifique qui décide si oui ou non il y a apprentissage du primate artificiel. Il décide d'employer le même terme pour le comportement du primate artificiel que pour caractériser celui du primate alors que ces comportements sont, en toute généralité, manifestement différents et que seuls certains aspects peuvent être mis en correspondance.

Ceci est critiquable dans la mesure où lorsque l'on cherche à reprendre le terme d'apprentissage à partir de la psychologie, on tend alors à le réduire aux mesures qui sont faites, ou à ses tendances qualitatives, comme l'accroissement d'un paramètre. En tout état de cause, quand on veut utiliser le même terme dans un domaine radicalement différent, on se contente aujourd'hui d'une définition de l'apprentissage qui se réduit à une mesure de certaines de ses manifestations. Encore faut-il noter que les spécificités de l'apprentissage issues de la psychologie et qui justifient le terme pour les systèmes artificiels sont rarement explicitées. Dans une telle démarche, le **psychologisme**¹⁰, c'est-à-dire *l'intervention de la subjectivité de chaque scientifique particulier dans l'interprétation, la justification et la validation d'un acte scientifique*, fait un retour en force dans la méthodologie. La difficulté est liée au fait que bien souvent on ne précise pas assez quels critères permettent de rapprocher différents phénomènes dynamiques et de considérer qu'ils permettent d'abstraire le

⁹Il est éventuellement inutile de réaliser effectivement l'expérience. Dans certains cas, on se contente d'une "expérience de pensée".

¹⁰C'est le terme technique employé habituellement dans ce contexte, par Popper [Popper, 1978], mais aussi avant lui. J'utilise le terme selon l'usage et non comme une dévalorisation de la psychologie.

même processus. On se contente généralement de faire appel à l'intuition du scientifique.

Les processus comme abstractions permettent la généralisation de leur domaine d'application

Il nous faut à présent préciser certaines propriétés des processus. Nous les avons simplement définis pour l'instant comme des abstractions de phénomènes dynamiques. Cependant, puisque les processus sont des abstractions et non des phénomènes, ils constituent également des idéalisations et vont donc acquérir une certaine indépendance par rapport à l'expérience particulière dont ils ont été abstraits, ce qui justifie la possibilité de généralisation de leur domaine d'application (flèche 4).

Les processus étudiés sont observés lorsqu'ils sont effectivement matérialisés. Dans ce cas, ils redeviennent des phénomènes. Il convient cependant de bien faire la distinction entre les phénomènes étudiés, leur abstraction sous forme de processus, puis la "rematérialisation" de ces processus, par exemple sous forme d'une simulation. Cette simulation, comme le phénomène, relèvera de la matérialité. Le processus au contraire relève de l'abstraction. Nous avons déjà noté, dans le premier chapitre, certaines difficultés liées à une indifférenciation de ces différents aspects.

En s'intéressant à des processus, les sciences cognitives, comme toutes les sciences, font le choix de travailler sur des abstractions. La masse, l'espace, ..., en physique sont aussi des notions essentiellement abstraites et donc construites. Ceci n'empêche pas d'ailleurs de défendre des positions réalistes. La bonne abstraction permettant par exemple la découverte, l'invention, au sens archéologique, de la structure réelle pertinente. Le danger serait cependant de considérer méthodologiquement cette abstraction comme première et précédant le phénomène. Ce n'est pas ce que font les sciences habituellement, qui définissent des abstractions à partir de phénomènes, de faits, de conceptions théoriques... La difficulté pour les sciences cognitives est d'abord conceptuelle. On a tendance à considérer la lecture ou l'apprentissage comme des phénomènes, alors que l'on parle déjà d'abstractions. La physique a déjà rencontré ce type de problèmes, lorsque l'on a cherché à définir une notion abstraite de masse ou d'espace alors que la masse et l'espace semblaient être des phénomènes que l'on comprenait intuitivement et que l'on appréhendait immédiatement.

En physique habituellement les abstractions dont on parle sont considérées comme des objets (l'électron, la masse, ...). La matérialisation de ces objets peut être considérée comme stable pour l'observateur. Au contraire la matérialisation d'un processus sera dynamique. Il est beaucoup plus difficile, dans des cas dynamiques, de continuer à faire la distinction entre abstraction et phénomène, car le phénomène dynamique semble déjà nous échapper.

L'ordinateur permet la rematérialisation des processus

On comprend bien alors le rôle essentiel que joue l'ordinateur dans la conception actuelle des sciences cognitives. L'ordinateur permet la matérialisation d'une abstraction sur un substrat différent. On appelle **substrat** *l'ensemble des phénomènes statiques qui permettent la constitution du phénomène dynamique que l'on abstrait comme processus.* Les substrats sont alors abstraits eux-mêmes par les sciences cognitives comme des objets. Les substrats n'ont donc pas de propriété dynamique qui interférerait avec le phénomène dynamique auquel on s'intéresse et que l'on abstrait comme processus. En effet, toutes les propriétés dynamiques qui entrent en jeu sont considérées comme partie prenante du phénomène dynamique lui-même et n'interviennent plus dans les phénomènes statiques qui permettent ce phénomène dynamique. En ce sens, ces substrats sont alors interchangeables. Nous entendrons par **rematérialisation** (flèche 5) *le fait de matérialiser une abstraction sur un substrat différent de celui qui a permis initialement cette abstraction.* On parle de multiréalisabilité ou de multiréalisation des abstractions.

La rematérialisation permet, au moins potentiellement, d'obtenir un nouveau phénomène à partir du même processus. L'attitude qui consiste à parler dans ce cas de rematérialisation est légitime si l'on pense que l'on aurait pu obtenir la même abstraction en partant des phénomènes observés sur le nouveau substrat (flèche 6). Dans le cas des sciences cognitives, le premier substrat est biologique et la rematérialisation se fait sur un substrat électronique. Il faut cependant prendre garde à ne pas traduire cette multiréalislation des processus, considérés comme des abstractions, en une possibilité d'identifier les phénomènes dynamiques. Le phénomène dynamique étudié et celui reconstruit sont différents car ils ne sont pas réalisés sur le même substrat. La confusion vient comme nous l'avons déjà souligné d'une confusion entre le phénomène dynamique et le processus qui en est l'abstraction. Identifier le phénomène à son abstraction constitue une erreur de catégories. La confusion entre un phénomène dynamique et le processus abstrait est, semble-t-il, plus difficile à éviter que celle entre un objet et un phénomène apparemment stable.

La confusion entre réalisation matérielle et abstraction est particulièrement évidente dans le terme de processeur. En effet, on désigne généralement par le terme de processeur, non pas le système abstrait qui réalise des opérations abstraites, mais un système électronique matériel. Le terme de processeur est le signe de la confusion – consciente ou non – en informatique entre l'abstraction et sa matérialisation.

Dans une optique de ce type, puisque l'ordinateur permettrait la matérialisation des processus et que l'ordinateur ne ferait que réaliser des calculs, on pourrait considérer les processus comme des calculs. Les sciences cognitives n'auraient plus qu'à reconstruire les phénomènes dynamiques sur un ordinateur.

1.2.4. Les sciences cognitives doivent-elles se résumer à une méthodologie ?

Il semble alors que ce qui unifie les sciences cognitives n'est plus un objet d'étude identique, mais une méthodologie sous-jacente. Des trois termes que nous avons cités, le phénomène étudié, son abstraction puis sa rematérialisation, les deux premiers intéressent plus particulièrement le scientifique. En effet, la construction des abstractions est là pour nous permettre de mieux comprendre le phénomène qui nous intéresse. Cependant, confondre l'étude de l'abstraction avec celle de sa rematérialisation change complètement la méthodologie scientifique. On peut se demander si ce n'est pas, d'une certaine manière, ce que prônent les sciences cognitives actuellement. En effet, on se rend compte qu'il est très difficile d'abstraire les phénomènes dynamiques liés à la cognition. Quand c'est possible et qu'on peut l'exprimer sous forme mathématique, il est alors quasiment impossible d'étudier ces abstractions mathématiquement. On est alors tenté par une reprogrammation directe de lois, règles, etc..., dictées par le bon sens, et une approche souvent naïve dite "qualitative" du phénomène qui nous intéresse.

Après avoir identifié le phénomène dynamique et son abstraction, on identifie le processus et sa rematérialisation. La boucle est alors achevée. On peut identifier le phénomène et sa rematérialisation. La "reconstruction" sur ordinateur de phénomènes dynamiques devient une méthode scientifique. La validation du modèle, ou sa "falsifiabilité" sont décidés par un observateur.

Contrairement à la démarche naturelle en sciences, qui consiste à créer des abstractions avant de les confronter au phénomène étudié, on peut dans ce cas se contenter d'une démarche analogique où un phénomène dynamique est reconstruit, simulé sur un autre substrat, l'ordinateur. En confondant le processus et sa matérialisation sur un ordinateur, on oublie le terme intermédiaire de la démarche scientifique habituelle, *l'abstraction du phénomène et quand cela est possible son abstraction sous l'aspect d'un modèle formel*.

Certains scientifiques, à l'époque des conférences Macy, considéraient que la cybernétique était le signe d'une crise profonde et générale de la méthodologie scientifique, plutôt qu'un champ disciplinaire en constitution. Comme le rapporte Dupuy, Lawrence Franck considère que "l'effort [de la cybernétique est] indissociable d'une rupture épistémologique avec la science orthodoxe, et d'une mise en accusation de la "méthode analytique". [...] Franck] ne craint pas de faire en bloc le procès du mécanisme ancien, de la "causalité linéaire", du réductionnisme d'une science qui ne voit que les états et ignore les processus" [Dupuy, 1985][p. 52]. Aujourd'hui, comme à cette époque pour la cybernétique, on peut considérer les sciences cognitives comme un mouvement scientifique général qui souhaite modifier implicitement les méthodologies autant qu'il propose un objet d'étude nouveau. Les deux sont en fait liés, puisqu'étudier des processus plutôt que des objets constitue une nouvelle façon d'aborder les phénomènes et nécessite de nouveaux

outils. Von Neumann et Wiener possédaient parfaitement les outils mathématiques qu'ils utilisaient dans cette optique, puisqu'ils les avaient eux-mêmes forgés dans les années trente, comme mathématiciens ou logiciens. Ils ne pouvaient cependant pas utiliser l'ordinateur comme cela est possible aujourd'hui. Dans cette optique, les sciences cognitives, comme le fait la physique actuellement, peuvent s'en servir comme d'un outil complémentaire. L'utilisation systématique de l'ordinateur modifie en même temps profondément le mécanisme de l'argumentation scientifique, comme l'intelligence artificielle nous l'a montré.

La démarche qui consiste à reconstruire directement un phénomène dynamique sur une machine semble possible pour tous les phénomènes dynamiques. En effet les propriétés d'universalité attribuées à la machine de Turing ouvrent tous ces possibles. Le développement de l'utilisation de l'ordinateur comme outil de simulation montre que dans toutes les disciplines cette méthodologie semble applicable. Pourquoi alors ne pas étudier tous les phénomènes dynamiques de cette manière ? Le stade suivant consiste, dans cette étrange confusion des genres, à arguer de la multiplicité des réalisations possibles des modèles, c'est-à-dire des abstractions, pour *utiliser un même modèle afin de rendre compte de phénomènes qui sémantiquement n'ont rien à voir*, du moins dans la sémantique "naturelle" du sens commun. Les matérialisations d'un processus abstrait du phénomène dynamique étudié vont rendre compte également de tout phénomène dont une abstraction pourrait être un système formel dont l'expression lui ressemble tant soit peu.

Dans ce type de projets, *on trouve donc en même temps une certaine conception de la connaissance*. Wiener l'avait également bien compris. Il écrit "le modèle formel idéal couvrirait l'univers tout entier, serait en accord avec sa complexité, et aurait une correspondance avec lui terme à terme. Toute personne capable d'élaborer et de comprendre un tel modèle dans son intégralité trouverait le modèle inutile, parce qu'il pourrait alors saisir l'univers directement comme un tout. Il possèderait la troisième catégorie de connaissance décrite par Spinoza" [Wiener et Rosenblueth, 1945][p. 320]. Il conclut en précisant que ce n'est pas possible pour l'homme parce que "l'outil principal en science est l'esprit humain et que l'esprit humain est fini" [Wiener et Rosenblueth, 1945][p. 321].

Les sciences cognitives constituent un véritable renversement en considérant des processus et non plus des objets comme éléments d'étude premiers. *Si on veut aller au bout de ce renversement, les objets doivent émerger des processus et non les processus des relations entre les objets*¹¹. La spécificité principale des sciences cognitives est donc leur objet d'étude qui n'est pas un objet, mais un ensemble de processus.

¹¹Nous reviendrons sur ce point dans le chapitre 5.

2. LE CONNEXIONNISME, UN CADRE GÉNÉRAL POUR MODÉLISER LES PROCESSUS “ÉMERGENTS”

On peut considérer la **cognition** comme *l'ensemble des moyens qui permettent à un individu ou à un ensemble d'individus de rentrer en contact avec le monde extérieur.* Cependant, comme on peut le constater, ce type de définition constitue en fait une description du monde sensible dans son ensemble. Rien ne nous empêche, en effet, de considérer un nuage d'électrons au sein d'un atome comme un individu en contact avec le monde extérieur. On pourra alors étudier la cognition de l'électron. Jean Charon écrit par exemple : “la Relativité complexe démontre donc que chaque particule de matière, indépendamment de son <<corps>> physique observable, possède un Mental non directement observable, mais dont la représentation permet d'affirmer que ce Mental possède des propriétés de mémoire cumulative et de raisonnement” [Charon, 1985][p. 71].

En fait, les définitions de la cognition ne sont pas tout à fait satisfaisantes. Elles peuvent être considérées comme précises, mais ne fournissent pas de critère qui permette de discriminer ce qui est cognitif et ce qui ne l'est pas. Ces définitions ne parviennent pas à délimiter le champ d'étude des sciences cognitives en précisant au moins ce qui n'entrerait pas dans leur cadre.

On peut ainsi se demander si finalement tous les processus ne sont pas cognitifs et ne relèvent pas à ce titre des sciences cognitives. Dès que l'on essaye de définir la cognition de façon positive, on finit par tout intégrer dans les sciences cognitives. Lorsque l'on cherche à faire une présentation générale de la discipline, on obtient rapidement des théories du tout. Ceci est lié au fait que l'unification de la discipline est généralement faite du point de vue des termes, c'est-à-dire à un niveau linguistique. Or, il faut noter un manque d'analyse linguistique des termes utilisés, ce qui n'est pas sans poser problème. De plus les mots utilisés pour définir la cognition ne sont pas discriminants. Ils renvoient rapidement à des phénomènes que l'on ne considèrerait pas spontanément comme cognitifs, voire à tous les phénomènes observables. Ainsi *l'unification des sciences cognitives se fait généralement au niveau des termes employés et dans une manière commune de poser les problèmes.*

Il est à noter que cette difficulté était déjà sous-jacente au projet cybernétique. À l'époque, *l'unification avait eu lieu grâce à la mise en évidence de structures logiques*

communes plutôt que par une utilisation d'un vocabulaire commun pour décrire des phénomènes différents. Dupuy le fait remarquer, lorsqu'il insiste, reprenant les notes de von Foerster, sur le fait que la cybernétique cherche sa propre unification et ne la trouve pas au niveau des solutions, mais au niveau des problèmes. "Certaines classes de problèmes, définies par une même structure logique, traversent les disciplines les plus variées. La cybernétique s'est édifiée principalement autour de deux de ces classes : les problèmes de communication d'une part ; les problèmes posés par l'étude des mécanismes qui produisent eux-mêmes leur unité (self-integrating mechanisms) d'autre part" [Dupuy, 1985][p. 54]. Ainsi la cybernétique ne cherchait pas son unité au niveau des termes, mais dans une structure logique commune.

Aujourd'hui encore, *l'unification est souvent faite au niveau des structures logiques, c'est-à-dire des modèles et des formalismes*. Ceci ne pose pas de difficultés quand on en trace les limites, mais ce n'est pas toujours le cas. Sinon, cette unification constitue une "théorie de tout", de même nature que celle que l'on observe au niveau des termes. Un exemple d'une telle unification au niveau des formalismes est celle proposée par Farmer pour le connexionnisme.

Nous présentons ici le cadre général d'étude proposé par Farmer [Farmer, 1990]. Ce cadre permet de présenter le connexionnisme, de le définir et d'en tracer les frontières. *Il résume les différents points communs des modèles qui relèvent du connexionnisme*. Ce cadre permet aussi de comprendre pourquoi l'utilisation d'outils provenant de différentes disciplines s'avère nécessaire. En effet certains problèmes de chimie ou de physique peuvent être formulés dans les termes proposés par Farmer. Les outils habituels qui permettent de les traiter en chimie ou en physique pourront alors être eux aussi formulés dans les termes proposés par Farmer et donc dans le cadre du connexionnisme.

L'objectif de Farmer est de proposer un formalisme commun à des modèles apparemment très différents et faisant référence à des disciplines très éloignées. Farmer met en fait en évidence une structure logique particulière. *C'est cette structure logique que Farmer appelle connexionnisme et qui est sous-jacente à chacun des phénomènes étudiés*. Ainsi l'unification connexionniste proposée par Farmer n'est pas une "théorie de tout" si on accepte de la considérer comme une structure logique particulière et si on précise rapidement des exemples de phénomènes dont les modèles ne pourront pas rentrer dans ce cadre.

2.1. L'article de Farmer, un cadre général pour la formalisation connexionniste

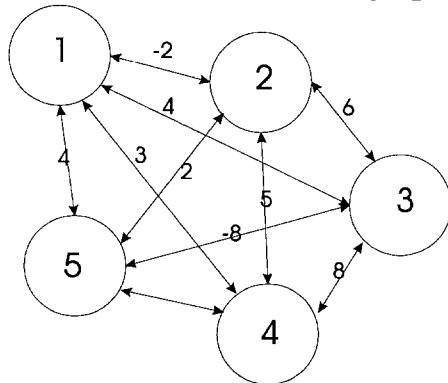
2.1.1. Le connexionnisme selon Farmer

Le connexionnisme est le terme générique utilisé par Farmer pour désigner la structure logique commune à un certain nombre de modèles de réseaux. Cette structure est constituée d'**unités élémentaires** reliées les unes aux autres pour former un **graphe**. Des symboles sont rattachés à chaque unité pour définir l'**état** de cette unité. On appelle **architecture du réseau** le graphe des relations entre les différentes unités. Les potentialités de ces modèles dépendent de façon importante de l'architecture du réseau. De plus, un symbole est attaché à chaque relation ou connexion entre unités. Ce symbole est appelé **poids de la connexion**. Il permet d'établir l'importance relative de cette connexion. L'ensemble du système peut ensuite se modifier au cours du temps. On appelle **paramètres** l'ensemble des symboles intervenant dans cette structure. Des **dynamiques**, c'est-à-dire des modifications de paramètres au cours du temps, peuvent en fait apparaître à trois niveaux. Les premières interviennent dans la modification des états de chaque unité du réseau. Les secondes permettent la modification des poids des connexions. Les troisièmes, enfin, permettent de modifier l'architecture du réseau elle-même, en faisant apparaître de nouvelles unités et de nouvelles connexions, ou éventuellement en les faisant disparaître. Si le réseau possède ce troisième type de dynamique, on dit que l'architecture du réseau est **plastique**. Farmer parle de réseau connexionniste lorsqu'au moins deux des trois types de dynamiques précédentes sont présents, généralement le premier et le second. J'appellerai **dynamique globale** du système l'ensemble des modifications de symboles de la structure au cours du temps.

Si on s'en tient à cette définition, il faut noter que les automates cellulaires, dans leur version étudiée par Wolfram [Wolfram, 1994], ne sont pas à proprement parler des réseaux connexionnistes. Dans le cas des automates cellulaires, les connexions existantes sont toutes identiques et sont donc considérées comme non pondérées. De plus la structure du graphe reste fixe au cours du temps. Ainsi les automates cellulaires comportent la première dynamique, mais ni la seconde, ni la troisième. Ils constituent ainsi un cas fortement dégénéré et beaucoup plus simple de réseaux connexionnistes. *Les automates cellulaires sont cependant un passage obligé pour bien comprendre quels outils nous possédons aujourd'hui pour étudier les systèmes connexionnistes en général.* Ces outils sont ceux utilisés pour étudier les systèmes dynamiques et les systèmes complexes, et adaptés à ce cadre. Nous en discuterons plus en détail dans le chapitre suivant.

Cette approche connexionniste se retrouve aujourd'hui très fréquemment dans de nombreux modèles, dans toutes les disciplines. On les trouve pour rendre compte du système nerveux, du réseau immunitaire, de certaines réactions chimiques auto-catalytiques ou de certains systèmes artificiels d'apprentissage comme les systèmes

Traduction d'un système connexionniste en termes de graphe



Le graphe correspond à un réseau **récurrent** où tous les sites sont reliés les uns aux autres avec un lien **symétrique**

Traduction d'un système connexionniste en termes de matrice

$$c = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & -2 & 4 & 3 & 4 \\ -2 & 0 & 6 & 5 & 2 \\ 4 & 6 & 0 & 8 & -8 \\ 3 & 5 & 8 & 0 & 7 \\ 4 & 2 & -8 & 7 & 0 \end{bmatrix}$$

Matrice des C_{ij} , c'est-à-dire des connexions dirigées de i vers j.

La matrice est **dense**. Plus précisément, la matrice est complètement interconnectée, si ce n'est que les sites ne sont pas reliés à eux-mêmes. La matrice est **symétrique**, ce qui traduit le fait que les connexions le sont également.

Traduction d'un système connexionniste en termes de liste

$1 \rightarrow 2 ; -2$	$1 \rightarrow 3 ; +4$	$1 \rightarrow 4 ; +3$	$1 \rightarrow 5 ; +4$
$2 \rightarrow 1 ; -2$	$2 \rightarrow 3 ; +6$	$2 \rightarrow 4 ; +5$	$2 \rightarrow 5 ; +2$
$3 \rightarrow 1 ; +4$	$3 \rightarrow 2 ; +6$	$3 \rightarrow 4 ; +8$	$3 \rightarrow 5 ; -8$
$4 \rightarrow 1 ; +3$	$4 \rightarrow 2 ; +5$	$4 \rightarrow 3 ; +8$	$4 \rightarrow 5 ; +7$
$5 \rightarrow 1 ; +4$	$5 \rightarrow 2 ; +2$	$5 \rightarrow 3 ; -8$	$5 \rightarrow 4 ; +7$

Le premier nombre est le numéro du premier site. La flèche indique un lien avec le site dont le numéro est le deuxième nombre. Le nombre après le point-virgule est le poids du lien. Si on veut l'interpréter en termes de règles on lit par exemple 1 entraîne 2 avec l'intensité -2. On peut passer automatiquement de la représentation sous forme de matrice à celle sous forme de graphe ou de liste.

FIGURE 3 : un système connexionniste complètement interconnecté et présenté sous ses différentes facettes : un graphe, une matrice et une liste

de classificateurs (classifier systems).

2.1.2. Un cadre mathématique général pour les modèles connexionnistes

Nous avons déjà souligné que le propre des systèmes connexionnistes est d'opérer sur un réseau reliant différents sites par l'intermédiaire de liens. Farmer parle de noeuds et de connexions, mais nous lui préférerons les termes de **sites** et de **liens**. Ces termes correspondent mieux au vocabulaire employé habituellement en physique pour décrire des modèles similaires.

Les unités : l'ensemble des sites

Les **sites** sont des unités simples. Nous entendons par **unité simple**, un élément qui peut être totalement décrit par un nombre fini de paramètres, généralement un seul, que nous noterons x_i pour l'unité i . Ce paramètre peut prendre un ensemble de valeurs appartenant à un espace mathématique défini au préalable. L'unité simple est totalement déterminée par son état, c'est-à-dire la spécification des valeurs de ses paramètres.

On peut avoir plusieurs types de sites. Chaque type est défini par le nombre et le type de ses paramètres et l'ensemble des valeurs qu'ils peuvent prendre.

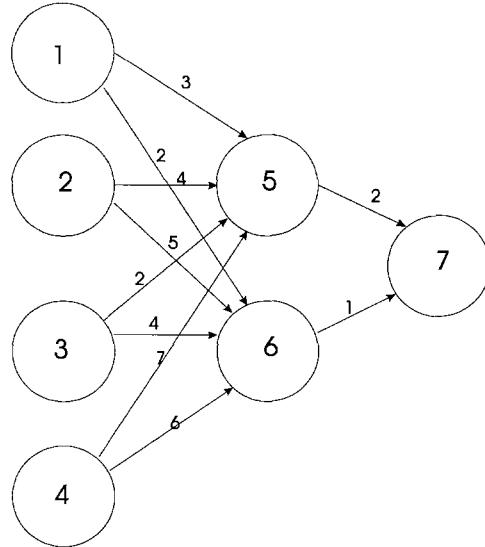
Le graphe : l'ensemble des liens

Le graphe est défini comme l'ensemble des **liens**, c'est-à-dire des connexions entre les différents sites. À chacun de ces liens on associe un symbole pris dans un ensemble quelconque et appelé **poids**. Nous noterons w_{ij} le poids du lien entre le site i et le site j . Généralement ce symbole est un nombre et l'ensemble des valeurs possibles est $[0, 1]$, l'ensemble des nombres réels compris entre 0 et 1, ou $[-1, 1]$. Mais on peut également choisir pour cet ensemble de valeurs un sous-ensemble d'un espace vectoriel de dimension supérieure strictement à un. On peut alors interpréter le poids comme un poids global associé à un lien global entre les sites i et j . Ce poids global peut être décomposé en poids élémentaires associés à des liens élémentaires. Ceci correspond à une description où on a des liens multiples entre deux mêmes sites. Ce cadre général est nécessaire pour décrire les réseaux autocatalytiques.

On peut trouver plusieurs types de graphes¹. On distingue en particulier les **graphes dirigés** de ceux qui ne le sont pas. On dit que le **graphe est non-dirigé** si les liens sont valables du site a vers le site b et réciproquement. Au contraire, si les liens ne sont valables que dans un seul sens, on parle de graphe dirigé. Dans le

¹La théorie des graphes est en soi un pan entier des mathématiques qui possède son vocabulaire et ses propres outils. Ne sont cités que quelques types de graphes qui sont les plus utilisés dans le cadre du connexionnisme.

Traduction d'un système connexionniste en termes de graphe



Le graphe est ici **feedforward**, car on peut choisir des sites comme entrées et d'autres comme sorties et les flèches nous mèneront alors toujours des entrées vers les sorties.

Traduction d'un système connexionniste en termes de matrice

$$c = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 0 & 0 & 0 & 3 & 2 & 0 \\ 2 & 0 & 0 & 0 & 0 & 4 & 5 & 0 \\ 3 & 0 & 0 & 0 & 0 & 2 & 4 & 0 \\ 4 & 0 & 0 & 0 & 0 & 7 & 6 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 6 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Matrice des C_{ij} , c'est-à-dire des connexions de i vers j.

La matrice est **creuse**. La matrice est **triangulaire supérieure** ce qui nous assure que l'on peut tracer un graphe feedforward lui correspondant. Réciproquement, si le graphe est feedforward, on peut numérotier les sites pour obtenir une matrice triangulaire supérieure.

Traduction d'un système connexionniste en termes de liste

$1 \rightarrow 5 ; +3$	$1 \rightarrow 6 ; +2$	$2 \rightarrow 5 ; +4$	$2 \rightarrow 6 ; +5$
$3 \rightarrow 5 ; +2$	$3 \rightarrow 6 ; +4$	$4 \rightarrow 5 ; +7$	$4 \rightarrow 6 ; +6$
$5 \rightarrow 7 ; +2$	$5 \rightarrow 7 ; +1$		

Le premier numéro entraîne l'apparition du second numéro avec l'intensité du troisième numéro. On peut passer automatiquement de la représentation sous forme de matrice à celle sous forme de graphe.

FIGURE 4 : un réseau à couche avec une couche cachée présenté sous ses différentes facettes : un graphe, une matrice et une liste.

cas où les liens sont dirigés, on a généralement deux liens entre deux sites différents, un dans chaque sens. C'est-à-dire que dans ce cas on a un lien de a vers b et un de b vers a , avec un poids particulier pour chacun de ces liens.

On peut aussi supposer l'existence de différents types de liens. C'est le cas en particulier quand on veut pouvoir modéliser certaines réactions chimiques mettant en jeu plusieurs espèces chimiques.

Lorsqu'il existe au plus un lien dirigé entre chaque site, ce qui est le cas habituel, on peut écrire une matrice contenant les poids des liens². Quand le graphe n'est pas dirigé, la matrice est symétrique, ce qui n'est plus le cas lorsque le graphe est dirigé. Une étude de la matrice des poids permet déjà une analyse rapide du graphe. En particulier, si la matrice est creuse³, peu de liens existent entre les différents sites et donc le graphe est peu dense⁴. On dit que le graphe est complètement connecté si aucun élément de la matrice n'est nul hors de la diagonale. On peut également avoir des liens d'une unité vers elle-même. On parle alors de **réverbération**.

Une autre représentation du graphe est possible sous la forme d'une liste qui associe un site d'arrivée à un site de départ. On peut également à partir d'une liste de ce type construire un graphe associé. Ce point de vue est particulièrement utile dans le cas des classificateurs. La dualité entre la présentation sous forme matricielle et celle sous forme de liste recouvre en fait la dualité de présentation en terme de réseau ou de règles. *Les deux points de vue, celui des réseaux et celui des règles sont complémentaires et il est possible d'effectuer une traduction automatique de l'un à l'autre.* De façon plus générale, la présentation en termes de liste est plutôt utilisée lorsque la matrice est creuse. Dans ce cas la présentation sous forme de liste est plus ramassée. Elle évite d'écrire des 0 pour les termes non utilisés.

Les dynamiques

Farmer appelle **règle de transition** la dynamique des états des sites, **règle d'apprentissage** la dynamique des liens, et **dynamique du graphe** celle qui porte directement sur l'architecture du réseau. Il appelle **métodynamiques** les modèles possédant cette troisième forme de dynamique.

Les dynamiques peuvent être de tout type, leur forme dépendant en particulier des objets qu'elles manipulent, c'est-à-dire de la représentation que l'on donne,

respectivement, des états, des liens et du graphe. On peut choisir de prendre comme représentation de ces trois types des objets construits à partir de l'arithmétique et donc discrets, aussi bien qu'un continuum de nombres réels, des symboles booléens

²Cette technique est généralisable lorsque l'on a plusieurs liens, avec des outils plus sophistiqués.

³On dit qu'une matrice est **creuse** si elle comprend peu de termes non nuls.

⁴Le graphe est **dense** si les sites sont presque tous reliés les uns aux autres.

ou toute forme symbolique. Le choix d'un ensemble discret permet de travailler de façon adaptée sur des machines discrètes et finies, comme les ordinateurs actuels. La motivation principale pour n'utiliser que des systèmes connexionnistes discrets ou symboliques consiste à essayer de construire des modèles qui, autant que faire se peut, soient directement simulables sur une machine symbolique.

Relation entre les systèmes dynamiques classiques et les systèmes connexionnistes

Le point de vue connexionniste est différent, au moins dans sa présentation, de celui des systèmes dynamiques classiques. La distinction entre les systèmes dynamiques classiques et le connexionnisme porte sur deux points. Tout d'abord les systèmes connexionnistes, contrairement aux systèmes dynamiques, manipulent des éléments qui sont de différents **types**.

Par exemple dans un système connexionniste on peut avoir des éléments qui sont discrets et valent seulement 0 ou 1, alors que d'autres sont continus et peuvent prendre n'importe quelle valeur réelle. Ceci n'est pas le cas dans ce que nous considérons comme les systèmes dynamiques classiques pour lesquels on a généralement un seul type (booléen, discret, continu), mais pas plusieurs en même temps. *Ainsi, systèmes dynamiques et systèmes connexionnistes ne font pas appel aux mêmes théories mathématiques.*

Dans le cadre des systèmes dynamiques, on commence par fixer un espace mathématique, puis on définit une ou des équations d'évolution dans cet espace. Généralement l'espace mathématique ainsi défini et les équations correspondent toutes au même cadre mathématique. On trouvera une liste non exhaustive de types de systèmes dynamiques et des espaces mathématiques qui leur sont associés sur la figure 5. L'espace sera par exemple uniquement continu comme un tore dans \mathbb{R}^3 , ou uniquement discret, comme \mathbb{N}^2 , mais on ne trouvera qu'exceptionnellement des cas mixtes comme le serait par exemple $\mathbb{R}^2 \times \mathbb{N}^2$. On peut noter la même chose pour les dynamiques qui seront toutes discrètes ou continues, mais rarement mixtes. Ainsi tous les paramètres modifiés sont de même type.

Dans les systèmes connexionnistes, tous les paramètres modifiés sont des symboles, mais certains correspondent à des états d'une unité, d'autres au poids d'une connexion et les derniers à l'existence d'unités et de connexions. Il est fréquent dans le cadre connexionniste, et c'est même le cas habituel, de trouver des états des unités qui sont discrets, alors que les poids sont continus et que l'existence d'unités et de connexions prend une valeur booléenne. *Le connexionnisme attribue des nombres à des éléments qu'il distingue sémantiquement.* Dans l'esprit du connexionnisme, une unité n'est pas un poids et on ne peut pas se contenter de dire que ce sont tous deux des éléments de \mathbb{R} ou de \mathbb{N} . *Ainsi la mixité de types observée pour les systèmes connexionnistes recouvre une mixité sémantique.* L'agrégation d'éléments de types

Type de système dynamique	Espace de départ	Temps	Espace d'arrivée
équations aux dérivées partielles EDP	continu	continu	continu
schème numérique d'une EDP	local	local	local
équations différentielles ordinaires	pas d'espace de départ autre que le temps	continu	continu
modèles sur réseau (lattice models)	local	discret ou continu	continu
automates cellulaires	local	discret	discret

FIGURE 5 : différents types de systèmes dynamiques, caractérisés par la nature de l'espace de départ lorsqu'il existe, du temps et de l'espace d'arrivée. Le terme local, dans ce contexte, signifie que l'espace de départ est discret et qu'une notion de voisinage est clairement mise en évidence (d'après Farmer).

différents dans un même espace mathématique est possible. Il est cependant difficile de l'interpréter sémantiquement.

La deuxième distinction porte sur le fait que les systèmes connexionnistes incluent plusieurs types de dynamiques et en particulier des **métodynamiques** définies comme des dynamiques sur les dynamiques.

Dans le cas des systèmes dynamiques, les dynamiques sont décrites sous la forme d'un système d'équations différentielles ou d'équations d'évolution discrètes. Les dynamiques sont spécifiées au départ, fixes et ne peuvent être modifiées par la suite que par l'intermédiaire de paramètres. Dans les modèles connexionnistes, au contraire, nous avons vu que les dynamiques peuvent apparaître à trois niveaux différents, ce qui modifie radicalement toute description en termes de systèmes d'équations intégral-différentielles. Si on souhaite décrire les systèmes connexionnistes en termes de systèmes dynamiques, on doit considérer que *la dynamique des liens modifie en même temps certains paramètres de la dynamique générale du système*. Généralement, lorsque l'on parle des réseaux connexionnistes en termes de systèmes dynamiques, on se contente de deux dynamiques, l'une qui modifie les états et l'autre qui modifie les poids.

On peut alors faire le même type de remarque sur la difficulté d'agrégation et d'interprétation sémantique des dynamiques que précédemment à propos des types de symboles manipulés. Dans le cadre des systèmes dynamiques, il y a une seule dynamique, spécifiée au départ. L'esprit du connexionnisme est de dire que plusieurs dynamiques se modifient les unes les autres. Cette approche du connexionnisme comportant de multiples dynamiques ne peut pas être réduite immédiatement aux systèmes dynamiques tels qu'ils sont généralement étudiés en mathématique.

Pour le faire, il faudrait dire que les différentes dynamiques ne font que participer à une dynamique globale. *Il faut alors savoir passer des trois dynamiques particulières à la dynamique globale et réciproquement.* Le passage des trois dynamiques à la dynamique globale se fait par agrégation comme nous l'avons vu précédemment. Et réciproquement si on étudie une dynamique globale dans le cadre des systèmes dynamiques, il faut être capable de spécifier quelles sous-dynamiques de la dynamique globale peuvent être associées respectivement aux sites, aux liens, au graphe, pour traduire l'étude de cette dynamique globale dans les termes du connexionnisme. Les réseaux connexionnistes constituent ainsi une généralisation des systèmes dynamiques à plusieurs dynamiques ou un cas particulier de système dynamique pour lequel on peut spécifier des sous-dynamiques. On ne peut donc pas affirmer directement que le connexionnisme se réduit aux systèmes dynamiques.

Une autre façon de s'en sortir consiste à considérer les métodynamiques comme des dynamiques particulières, ce qui est effectivement le cas. Cependant pour considérer les métodynamiques et les dynamiques sur le même plan, il faut changer l'espace mathématique dans lequel on travaille. *Expliciter ces changements*

constitue déjà un dépassement des conceptions classiques sur les systèmes dynamiques.

En dernier recours, on peut considérer qu'une des dynamiques est **lente**, alors que l'autre est **rapide**. Ceci permet de distinguer deux dynamiques qui ne présentent pas de différence mathématique mais qui sont de nature différente dans l'esprit du connexionnisme. Le fait que l'une soit lente et l'autre rapide est relatif et se traduit en fait par l'intermédiaire d'un ou de plusieurs coefficients qui vont contenir cette vitesse relative. Cette façon d'interpréter les modèles connexionnistes dans les termes des systèmes dynamiques n'est pourtant pas tout à fait satisfaisante. En effet les modèles connexionnistes permettent de construire des dynamiques sur les paramètres qui sont "aussi rapides" que la dynamique globale. Si on veut traduire cela en termes de système dynamique, il faut pouvoir modifier les coefficients de vitesse relative, ou qu'ils puissent être modifiés par la dynamique du système elle-même. *On quitte alors le domaine classique des systèmes dynamiques fixés initialement pour obtenir des systèmes dynamiques dont la structure est susceptible de varier au cours du temps.* La difficulté principale d'une traduction en termes de systèmes dynamiques, même dans les cas où seules deux types de dynamiques sur trois sont en jeu, consiste à exhiber la relation entre la dynamique lente et la dynamique rapide.

Mathématiquement, il n'y a cependant pas de différence de nature entre les systèmes dynamiques et les réseaux connexionnistes. En effet, tout deux manipulent des symboles. *On peut réinterpréter les systèmes connexionnistes en terme de systèmes dynamiques, mais dans un cadre plus large que celui des systèmes dynamiques classiques.* La position de Farmer à cet égard est dénuée de toute ambiguïté. Il écrit : "cet article essaye de proposer une présentation cohérente des modèles connexionnistes et de la façon dont ils diffèrent dans leur structure mathématique et dans leur philosophie des modèles conventionnels de système dynamique fixé"⁵ [Farmer, 1990][p. 153].

2.1.3. Exemples de modèles connexionnistes

Nous décrivons ici quelques modèles connexionnistes que nous décrivons dans le cadre de Farmer : les réseaux de neurones formels, les systèmes de classificateurs et certains modèles formels du système immunitaire. Nous supposons ici que le lecteur est familier avec au moins un des systèmes présentés et qu'il pourra constater la pertinence du cadre de Farmer pour établir la proximité de formalisme entre ces différents modèles. On trouvera sur la figure 6 une table de correspondance récapitulative du vocabulaire employé dans chacune des disciplines évoquées.

⁵Il faut d'ailleurs noter que dans son article Farmer n'aborde pas une comparaison systématique entre l'approche connexioniste et les systèmes dynamiques étudiés en mathématiques.

Les réseaux de neurones formels

Les **réseaux de neurones formels** ont d'abord été conçus comme des modèles des réseaux de neurones biologiques qui devaient rendre compte de fonctions cognitives comme la perception. À la suite des travaux sur le perceptron puis après la reprise de la notion de réseau de neurones par la communauté des physiciens, le modèle est devenu un outil mathématique autonome étudié pour lui-même.

La description biologique et la schématisation classique du neurone sont les suivantes. Biologiquement, le neurone est une cellule nerveuse. On distingue anatomiquement le soma, l'axone, les terminaisons axonales et les synapses, et les dendrites. On considère que l'influx nerveux est dirigé des dendrites vers le soma puis vers l'axone et les terminaisons axonales. Les influx issus des dendrites arrivent au soma. Si l'influx total est suffisant, le neurone s'active et un influx nerveux se propage dans l'axone jusqu'à la zone présynaptique. Un neurotransmetteur, c'est-à-dire une molécule, est alors sécrétée par la cellule et se déplace jusqu'à la zone postsynaptique du neurone suivant. La schématisation symbolique du neurone est la suivante. Des liens arrivent au neurone et d'autres en partent. Ces liens représentent l'ensemble des liaisons entre deux somas, c'est-à-dire à la fois l'axone, les synapses et les dendrites. Un poids est associé à ce lien. Ce poids correspond à l'importance de la liaison et de la transmission d'influx entre deux neurones. On rend compte du fonctionnement du soma en le modélisant comme un sommateur à seuil. Il fait la somme pondérée des entrées et s'active si cette somme est supérieure à une certaine valeur. Dans ce cas, une sortie correspondant à l'activation du neurone est transmise au neurone suivant.

Les *sites* correspondent aux neurones. L'état du neurone est décrit par un nombre réel, qui représente son activité. On peut lui ajouter d'autres paramètres selon la règle de transition que l'on souhaite fixer.

Les *liens* sont les relations que les neurones établissent avec d'autres neurones, par l'intermédiaire de l'axone, des synapses et des dendrites.

Différents types de graphes ont été étudiés. Les plus fréquents sont les graphes avec des rangées de sites qui ne peuvent être connectées qu'à la rangée suivante. On parle alors de **réseaux à couches**. La première couche sert pour les entrées du réseau, la dernière pour les sorties. On parle de **réseau à couches cachées**, lorsque l'on trouve des couches intermédiaires qui ne sont reliées directement ni aux entrées, ni aux sorties. On parle de **réseau feedforward**, si une couche sert d'entrée et une autre de sortie. On parle de **réseau récurrent** si au moins un neurone a une sortie qui est reliée à sa propre entrée par l'intermédiaire d'autres neurones. Les réseaux à couche cachée permettent, entre autres d'approximer des fonctions par morceaux. On dit que ces réseaux sont des approximatrices universelles [Cybenko, 1989][Sontag, 1990]. On parle également d'**apprentissage** pour désigner les propriétés de généralisation de ces réseaux. Le second type de réseau le plus

étudié est celui qui est complètement interconnecté. Tous les neurones ou presque sont alors reliés les uns aux autres. Ce type de réseau permet par exemple de résoudre des problèmes de classification.

Les *règles de transition* ou dynamiques des états des sites sont appelées **dynamiques d'activation** des neurones. On peut en trouver plusieurs types. Généralement, la modélisation du neurone consiste à dire que le soma effectue une sommation pondérée des activités des dendrites et qu'il se décharge électriquement si un seuil d'activation est atteint. En conséquence, on écrit que l'activité $x_j(t+1)$ du neurone j à l'instant $t+1$ est fonction pondérée des activités à l'instant t des neurones qui lui sont connectés et que cette activation a lieu si le seuil est franchi, soit

$$x_j(t+1) = S\left(\sum_{i=1}^N w_{ij}x_i(t) - \theta_j(t)\right)$$

La fonction d'activation S est généralement une fonction de Heaviside⁶. Le seuil $\theta_j(t)$ peut éventuellement être modifié. On parle de fonction seuil pour l'évolution de $\theta_j(t)$ en fonction du temps. On trouve également comme fonction d'activation des approximateurs réguliers de la fonction de Heaviside⁷, comme la fonction sigmoïde⁸. Dans le cas d'un réseau feedforward, la règle de transition se réduit à la fonction d'activation. Dans le cas d'un réseau récurrent, la règle de transition est plutôt celle d'un automate, ou un système d'équations différentielles ordinaires.

On trouve différentes *règles d'apprentissage*. Pour les réseaux feedforward, la

⁶La fonction de Heaviside est une fonction définie par

$$\begin{aligned} f & : \mathbb{R} \rightarrow \mathbb{R} \\ f(x) & = 0 \text{ si } x < 0 \\ f(x) & = 1 \text{ si } x > 0 \end{aligned}$$

Le seuil, qui correspond également à la discontinuité de la fonction se situe au point d'abscisse $x = 0$ et vaut 1, c'est-à-dire la hauteur du saut au point de discontinuité. Dans certains cas, on choisit comme valeurs -1 et $+1$, avec toujours une discontinuité en 0.

⁷C'est-à-dire des fonctions C^∞ , aussi proches que souhaité de la fonction de Heaviside.

⁸La sigmoïde est une approximation C^∞ de la fonction de Heaviside. Elle est définie par

$$\begin{aligned} f & : \mathbb{R} \rightarrow \mathbb{R} \\ f(x) & = \frac{1}{1 + e^{-x}} \end{aligned}$$

On peut également approximer la fonction de Heaviside par la famille de fonctions C^∞ , suivantes, avec $\beta \in \mathbb{R}^{+*}$

$$\begin{aligned} f & : \mathbb{R} \rightarrow \mathbb{R} \\ f(x) & = \frac{1}{1 + e^{-\beta x}} \end{aligned}$$

On interprète alors β comme l'inverse d'une température T .

règle la plus répandue est celle de rétropropagation de l'erreur. L'“erreur” est calculée en comparant la valeur effectivement obtenue par le réseau à une valeur de référence. Il est donc nécessaire de connaître cette bonne valeur. C'est pourquoi on dit que cette règle d'apprentissage correspond à de l'apprentissage supervisé. Cet algorithme est non-linéaire et possède généralement de nombreux attracteurs de type point fixe. Pour les règles d'apprentissage non-supervisé on trouve généralement des dynamiques de type “recuit simulé”.

Les *dynamiques du graphe* : on trouve parfois, mais rarement, une troisième dynamique dans certains modèles de réseaux de neurones. Cette dernière dynamique consiste à modifier la structure du réseau. Elle permet généralement d'accélérer l'apprentissage ou de permettre un nouvel apprentissage lorsque l'on avait atteint un minimum local. Cette dynamique est donc généralement conçue comme une forme particulière de recuit de la dynamique d'apprentissage. On la considère alors comme une dynamique “lente” par rapport à la dynamique d'apprentissage. On considère aujourd’hui que l'étude de ces dynamiques est une des voies pour les recherches à venir. Elle consiste par exemple à modifier une température $T = \frac{1}{\beta}$ qui permet de décrire une famille de fonctions d'activation, par exemple $f(x) = \frac{1}{1+e^{-\beta x}}$ comme dans la note précédente sur la fonction sigmoïde. On augmente la température T ou on la diminue et on modifie ainsi la dynamique d'activation.

Les systèmes de classificateurs

Les **systèmes de classificateurs** (classifier systems) constituent une approche de l'apprentissage des machines imaginée par un chercheur en intelligence artificielle, John Holland [Holland, 1986]. Ils permettent de modifier des systèmes de règles. Ils ne constituent pas une tentative de modélisation. Les systèmes de classificateurs, à première vue, ne ressemblent pas du tout aux réseaux connexionnistes. Ils sont présentés comme des algorithmes de manipulation de règles ou classificateurs, comportant deux parties. La première correspond aux **conditions** et la seconde aux effets ou **messages**. Il n'est pas *a priori* évident de se rendre compte que l'on peut également les représenter sous forme d'un graphe permettant la manipulation de nombres réels. On n'a généralement pas conscience de cette possibilité parce que les systèmes de règles se présentent presque toujours sous forme d'arbres plutôt que de graphes.

Ces règles sont définies comme une liste, c'est-à-dire un ensemble ordonné de symboles, de longueur fixe, comportant une première partie correspondant aux conditions d'application de ces règles ou conditions, et une seconde partie correspondant à l'effet de ces règles ou actions. Une règle peut en invoquer une autre lorsque les effets de la première sont les conditions de la seconde. Pour savoir si les effets d'une règle sont les conditions d'une autre, on essaye d'apparier (matcher) les effets d'une règle avec les conditions de toutes les autres règles. On trouve également une liste

de messages qui fonctionne comme un tableau noir sur lequel les règles inscrivent des messages les unes pour les autres. Cette liste de message est de taille fixe. Sont inscrits sur cette liste des messages externes éventuels. On trouve également sur cette liste à l'instant $t + 1$ certaines règles dont les conditions coïncident avec les effets d'un élément de la liste des règles inscrites sur la liste à l'instant t .

Les conditions, les actions et les messages sont tous des listes de longueur fixe, identique. Les messages sont des listes dont les symboles sont des 0 et des 1. Les conditions et les actions sont des listes de 0, de 1 et de $\#$, où $\#$ est un symbole quelconque parmi ceux possibles, un joker, qui en l'occurrence sera soit un 0 soit un 1, de façon indéterminée. On peut donc ne pas s'occuper des emplacements où l'on trouve le symbole $\#$, lorsque l'on essaye d'apparier deux messages différents. On appelle alors **spécificité** du classificateur le pourcentage de symboles, dans les conditions du classificateur, qui sont différents de $\#$. La spécificité permet de distinguer certaines règles considérées comme des "spécialistes", alors que d'autres seraient des "généralistes".

À chaque classificateur on associe un nombre réel qui correspond à sa **force**. Il faut en effet choisir à chaque pas de temps quels classificateurs peuvent être inscrits et envoyer des messages dans la liste des messages qui est de taille fixe. Pour cela toutes les règles, dont les conditions coïncident avec les effets d'une règle inscrite sur la liste, vont proposer une **enchère**. Les règles ayant proposé les enchères les plus élevées seront alors inscrites sur la liste à l'instant suivant. L'enchère est généralement de la forme $Constante \times force \times spécificité \times F(intensité)$. $F(intensité)$ est une fonction de l'intensité de toutes les règles qui sont candidates à être inscrites sur la liste. Différentes formes sont possibles, l'intensité la plus haute, une moyenne des intensités, etc... Ceci permet en fait d'obtenir une dynamique sur les messages.

On peut également ajouter une dynamique sur les classificateurs à l'aide d'algorithmes du type du Bucket Brigade, dont on sait qu'il est très proche des algorithmes du type Temporal Difference. On obtient alors une règle d'apprentissage.

Les sites correspondent aux messages, c'est-à-dire aux conditions et aux actions possibles. Pour un système de classificateurs de longueur N , on a 2^N messages possibles. Chaque condition d'un classificateur est alors un site, de même que chaque action. Pour un classificateur donné, sa condition comporte généralement des $\#$, ce qui permet de définir un ensemble de conditions d'application du classificateur. On peut dire la même chose des actions de ce classificateur. On peut traduire cela en disant qu'un classificateur associe un ensemble de sites à un autre ensemble de sites. L'état du site i est l'intensité x_i . L'activité du site dépend également d'un seuil global $\theta_i(t)$, qui varie au cours du temps.

Les liens sont formés à partir des classificateurs. À chaque condition du classificateur est associée un ensemble d'actions de ce classificateur. On peut alors représenter un classificateur sous la forme d'une liste dont les éléments de départ sont des sites

et les éléments d'arrivée sont également des sites. On peut également représenter cette liste comme un graphe ou en donner une version matricielle. Généralement les systèmes de classificateurs permettent d'obtenir des matrices creuses. La valeur des poids dépend de la force de la règle mais également du degré de "spécialisation" de la règle. On a alors le poids w_{ij} sous la forme $w_{ij} = \text{spécificité} \times \text{force}$.

Les règles de transition peuvent être de différentes formes. Généralement on choisit une règle dite de "soutien", linéaire par morceaux

$$x_j(t+1) = \sum_{\{i/x_i(t)w_{ij} > \theta\}} x_i(t)w_{ij}$$

avec pour θ un seuil choisi arbitrairement ou variant au cours du temps. La non-linéarité de la règle est due à ce seuil. On l'appelle règle de soutien car il y a un effet de moyennage sur l'ensemble des états des sites à l'instant précédent. Une autre règle est celle du maximum qui consiste à choisir comme règle de transition $x_j(t+1) = \max_i \{x_i(t)\}$.

La règle d'apprentissage traditionnelle pour les systèmes de classificateurs est l'algorithme de Bucket Brigade. Cet algorithme est dans son principe de même type que les règles d'apprentissage définies par Hebb. On sait également que cet algorithme de Bucket Brigade est formellement un équivalent des algorithmes d'apprentissage par renforcement comme Temporal Difference, ou comme le Q-learning [Saurel, 1992][Munos et Patinel, 1995].

Les dynamiques du graphe sont des dynamiques qui vont modifier le graphe, en l'occurrence les classificateurs. Ce sont des dynamiques qui permettent d'engendrer de nouveaux classificateurs à partir des anciens. Les algorithmes utilisés sont appelés généralement **algorithmes génétiques**. Leur principe est simple : il consiste à considérer les classificateurs comme un brin codé. On reprend alors les termes de la génétique que sont la *mutation* et le *crossover* pour décrire des opérations sur un classificateur, c'est-à-dire sur les liens et les sites. La **mutation** est la *construction d'un nouveau classificateur à partir d'un classificateur déjà existant en modifiant simplement une partie de la condition ou de l'action*, cette modification étant interprétée comme une mutation. Le **crossover** consiste, à partir de deux classificateurs déjà dans le système, à *créer deux nouveaux classificateurs dont les actions ou les conditions seront les mêmes que celles des classificateurs initiaux, si ce n'est qu'une partie de la liste de 0 et de 1 aura été échangée entre l'un et l'autre*.

Les modèles formels du système immunitaire en termes de réseau

La tâche première du système immunitaire, selon la conception de Farmer, Perelson ou Edelman, est d'être capable de distinguer le soi du non-soi. Les éléments de base du système sont des molécules, les **anticorps**. Des cellules, les **lymphocytes**, produisent les anticorps. Les lymphocytes ont des anticorps à leur surface qui servent

nom générique	réseau de neurones	système de classifieurs	réseau immunitaire
site	neurone	message	espèce d'anticorps
état	niveau d'activation	intensité	anticorps libre et concentration en antigène
lien	axone, synapses et dendrites	classifieur	réaction chimique des anticorps
paramètres	poids des connexions	force et spécificité	affinité de la réaction, concentration en lymphocyte
règle d'interaction	somme et sigmoïde	seuil linéaire et maximum	gaussienne
algorithme d'apprentissage	Hebb, backpropagation, etc...	Bucket Brigade (de type hebbien)	sélection clonale (de type hebbien)
dynamique du graphe	plasticité synaptique	algorithmes génétiques	algorithmes génétiques

FIGURE 6 : la pierre de Rosette proposée par Farmer, table des correspondances de vocabulaire entre les différents domaines (d'après Farmer).

à détecter les **antigènes**, c'est-à-dire les éléments étrangers. Les **macrophages** enfin, sont de grosses cellules qui vont détruire les éléments marqués par les anticorps.

Les observations montrent de plus que l'on trouve dans le corps moins d'anticorps que d'antigènes susceptibles d'être rencontrés par l'organisme. Pour expliquer ces observations, on a proposé que le système immunitaire ait une capacité de mémorisation et de généralisation. Comme ces propriétés sont caractéristiques d'un effet de réseau, des **modèles connexionnistes du système immunitaire** ont été proposés pour en rendre compte. L'intérêt des réseaux pour expliquer le fonctionnement du système immunitaire reste cependant très controversé dans les communautés médicales.

Les changements de concentration des lymphocytes sont le mécanisme qui permet l'apprentissage du réseau. Le mécanisme envisagé est appelé **sélection clonale**. Un lymphocyte est stimulé par un antigène particulier s'ils ont ensemble une réaction chimique. S'il est stimulé, le lymphocyte se réplique et produit d'autres lymphocytes du même type qui sécrètent également des anticorps libres qui vont marquer les antigènes. Si les lymphocytes ne sont pas stimulés par des antigènes ils ne se répliquent pas et peuvent éventuellement être supprimés du système. La sélection clonale permet d'expliquer comment le système immunitaire reconnaît et supprime les antigènes. Elle ne permet pas d'expliquer comment le système distingue le soi du non-soi et en particulier comment il distingue les antigènes des anticorps qui sont l'un comme l'autre des molécules.

La sélection clonale ne doit pas avoir lieu pour les anticorps, c'est-à-dire pour les molécules du soi. Il semble que cette capacité soit apprise par un effet de réseau plutôt que "programmée en dur" dans le système. Jerne a proposé à cet effet que ce soient les interactions des différents types d'anticorps et de lymphocytes les uns avec les autres qui confèrent au système une bonne partie de ses propriétés.

Les connexions entre nœuds sont assurées par des réactions chimiques entre anticorps, lymphocytes et antigènes. La force des connexions dépend de la force de ces réactions chimiques. Le graphe est alors déterminé en assignant à chaque anticorps un nombre entier écrit en binaire. Les liens sur le graphe sont alors obtenus en comparant le nombre binaire associé à un anticorps et celui associé à un autre anticorps. On en déduit une matrice des forces des connexions. On définit alors une dynamique portant sur la concentration en anticorps sous la forme d'une équation différentielle. Une autre équation différentielle permet de connaître l'évolution de la concentration en lymphocytes.

La correspondance proposée ici entre le système immunitaire et sa formalisation connexionniste n'est pas la seule envisageable. Elle correspond à la description qu'en propose Farmer dans son article [Farmer, 1990] et qu'il a reprise de ses travaux avec Packard et Perelson [Farmer *et al.*, 1986].

Les *sites*, dans ces modèles, sont les anticorps. Leur état est donné par la concentration correspondante en anticorps libres. Les concentrations en lymphocytes sont les paramètres des sites.

Les *liens* sont donnés par les réactions chimiques entre anticorps, lymphocytes et antigènes. La force de ces connexions dépend de la force de ces réactions chimiques. Le graphe est alors déterminé en attribuant à chaque anticorps un nombre entier écrit en binaire. *Ce nombre correspond en fait à la forme de l'anticorps.* Les liens sont obtenus en comparant le nombre binaire associé à un anticorps à celui associé à un autre anticorps. On associe un nombre à ce lien qui dépend de la complémentarité⁹ entre les deux nombres binaires. *Si ces nombres sont complémentaires, les formes des anticorps sont complémentaires.* La réaction entre ces anticorps est alors importante et on lui associe un poids élevé. Plus précisément, le poids est fonction de la longueur de complémentarité entre ces deux nombres. On construit alors une matrice de connexion avec ces poids.

Les *règles de transition* sont des fonctions gaussiennes plutôt que des sigmoïdes.

Les *règles d'apprentissage*, dans le cas du réseau immunitaire et contrairement aux réseaux de neurones, consistent à modifier les paramètres des sites et non les paramètres des liens. Concrètement ces règles consistent à modifier la concentration en lymphocytes par réPLICATION DES LYMPHOCYTES ACTIVÉS SELON LES PRINCIPES DE LA SÉLECTION CLONALE.

Les *dynamiques du graphe* proviennent de la réPLICATION DES LYMPHOCYTES QUI N'A PAS LIEU À L'IDENTIQUE MAIS PERMET D'INTRODUIRE DE NOUVELLES ESPÈCES DE LYMPHOCYTES. LA DYNAMIQUE CHOISIE CONSISTE À UTILISER DES ALGORITHMES GÉNÉTIQUES QUI PERMETTENT LA RÉPLICATION DES LYMPHOCYTES TOUT EN INTRODUISANT DE NOUVELLES ESPÈCES.

Autres exemples

D'autres disciplines offrent des modèles connexionnistes qui peuvent être décrits dans le cadre formel proposé par Farmer. On trouve des modèles connexionnistes en écologie, en génétique des populations, en économie, dans le cadre de la théorie des jeux, ou pour rendre compte de modèles d'évolution moléculaire. Dans ces disciplines, certains phénomènes étudiés se modélisent naturellement dans les termes connexionnistes. Tout en restant dans le cadre proposé par Farmer, il peut être parfois nécessaire de transcrire ces modèles sous la forme de réseaux que nous n'avons pas évoqués, comme les réseaux markoviens ou les réseaux booléens.

Farmer réinterprète également les **RÉSEAUX AUTOCATALYTIQUES** dans le cadre qu'il a proposé. On peut cependant émettre des réserves sur ce système formel dont la

⁹Deux chiffres seront complémentaires si l'un est 1 et l'autre 0. On calcule la complémentarité de deux nombres en comptant le nombre de chiffres complémentaires pour une même place dans l'écriture binaire des deux nombres. La longueur de la complémentarité sera le nombre de chiffres complémentaires lorsque l'on compare ces deux nombres.

description ne semble pas respecter les hypothèses que Farmer a lui-même posées pour définir son cadre.

2.2. Intérêts et limites du connexionnisme pour modéliser les processus

Farmer nous montre bien qu'en fait le connexionnisme est un cadre très général pour modéliser les processus. Nous n'insisterons pas sur l'étendue des phénomènes dont il peut rendre compte. Il nous reste à montrer que certains phénomènes ne peuvent pas être modélisés dans ce cadre. Ceci nous assurera que le connexionnisme n'est pas le support déguisé d'une "théorie de tout". Il reste également à vérifier que le connexionnisme permet bien de rendre compte des processus de la façon dont les sciences cognitives souhaitent les aborder.

On peut tout d'abord adresser deux réserves concernant l'approche de Farmer. La première est de ne pas offrir d'outils d'analyse théorique généraux. Farmer propose uniquement un cadre théorique général dans lequel on peut formuler des modèles apparemment très différents mais qui font, en fait, le même type d'hypothèses. Farmer ne précise pas quels outils mathématiques sont utilisés dans chaque discipline pour analyser ces modèles. Le chapitre 3 ira en partie dans ce sens en présentant les outils utilisés dans les modèles de complexité. Le pas suivant consisterait à exhiber ces différents outils dans les termes proposés par Farmer. La seconde critique est de ne pas assez insister sur les types de graphes ni les types de dynamiques susceptibles d'être intéressants. Le cadre proposé est extrêmement vaste, mais on ne sait pas quelles sont les spécifications particulières qui sont susceptibles d'être scientifiquement fructueuses.

En fait les modèles des phénomènes qui relèvent de la complexité telle qu'on la définit techniquement aujourd'hui peuvent être décrits dans le cadre de Farmer. Dans le chapitre 3 nous présenterons et étudierons plus en détail ces modèles de la complexité et nous décrirons les outils utilisés pour les étudier. Ceci nous permettra d'envisager les questions et les perspectives offertes par le cadre de Farmer, c'est-à-dire en toute généralité par les modèles actuels d'explication du fonctionnement cérébral.

2.2.1. La philosophie connexionniste de la modélisation des processus

Farmer réussit à exprimer des modèles relevant de disciplines très différentes dans un même cadre, parce que *tous ces modèles sont bâtis sur les mêmes hypothèses*. Farmer ne nous donne pas la liste de ces hypothèses. Ces hypothèses sont en fait celles qui justifient en particulier les modèles de réseaux neuronaux.

On peut discerner trois hypothèses génératives dans ces modèles : la séparation, la localité et l'immédiateté. Ces trois hypothèses sont éventuellement liées. Elles se retrouvent par exemple à différents niveaux dans le fonctionnement du neurone symbolique et sont parfois entremêlées. La **séparation** consiste à pouvoir distinguer *des unités bien déterminées et dont les états sont indépendants*. La **localité** consiste à dire que les transactions entre unités qui entraînent des modifications des états de ces unités *ont lieu seulement entre voisins*. Ceci nécessite de définir au préalable les “voisinages”. L'**immédiateté** consiste à supposer que *la transaction entre unités est instantanée*, même si elle a lieu avec un décalage temporel.

La première hypothèse est de considérer que les éléments pertinents qui constituent le phénomène global sont des unités bien localisées et séparées les unes des autres, les neurones (*hypothèse de séparation*). Ces entités s'influencent les unes les autres dans une causalité identique à celle des chocs de la mécanique classique (*hypothèses d'immédiateté et de localité*). Un élément discret est une cause pour un élément considéré comme son successeur, cause immédiatement suivie d'un effet. Le neurone est la boule ponctuelle de la mécanique classique. La quantité de mouvement est ici l'état du neurone et le poids permet de déterminer la part de quantité de mouvement transmise d'une boule à la boule suivante.

La localité ne se retrouve pas autre part dans le modèle, puisque dès que l'on a un réseau très fortement connecté tous les sites sont directement reliés à tous les autres. Ainsi cette localité de l'action d'un neurone sur un autre neurone auquel il est relié n'est pas en contradiction avec le fait que tous les neurones influencent tous les neurones auxquels ils sont reliés en un pas de temps, c'est-à-dire tout le réseau si le graphe est complètement interconnecté et la dynamique synchrone.

On peut noter que, dans le cadre du connexionnisme, modéliser les processus consiste à considérer un système formel dynamique que l'on identifie à ce processus. Ce système formel est dynamique comme modification d'unités plus petites qui constituent le système dans son ensemble. Les unités elles-mêmes, formelles, sont identifiées aux neurones réels ou artificiels. Dans cette conception la dynamique globale du système n'est pas première, elle est “émergente” et constituée par les modifications d'unités simples. Le fonctionnement du système nerveux, modélisé par les réseaux de neurones, est considéré comme un phénomène dynamique émergent. Le connexionnisme en rend compte par les modifications des états des neurones artificiels. Ces neurones eux-mêmes sont considérés comme des objets, ou comme des symboles décrits par quelques paramètres. Ainsi les modèles connexionnistes, s'ils permettent bien de rendre compte de processus, le font cependant en s'appuyant sur des objets qui sont premiers et permettent l'apparition de ces processus.

Modéliser un processus revient alors à modéliser les variations de paramètres définissant des objets. Ce type de modèle est donc relativement peu satisfaisant si l'objectif affiché des sciences cognitives est bien de considérer les processus comme premiers et non les objets. Il faut alors trouver d'autres types de modèles construits

sur d'autres hypothèses et qui permettent bien de considérer les processus comme premiers. Nous verrons comment cela est envisageable dans le chapitre 5.

2.2.2. Ce que le connexionnisme ne peut pas faire

Ces trois hypothèses de localité, d'immédiateté et de séparation constituent en fait le choix d'une certaine causalité considérée comme pertinente à une échelle d'observation fixée. Ces hypothèses que l'on retrouve dans le connexionnisme sont pertinentes pour modéliser les phénomènes présentés précédemment dans le cadre de Farmer parce que dans ce cadre on considère que cette causalité est pertinente. D'autres causalités ne pourraient pas être rendues directement par ce type de modèles.

Certains phénomènes sont modélisés habituellement à partir d'autres hypothèses. On ne pourrait pas en rendre compte avec les hypothèses de localité, d'immédiateté et de séparation. On peut citer comme premier exemple les phénomènes quantiques pour lesquels les hypothèses de séparation des objets ne sont pas présentes.

Lorsque l'on modélise actuellement les phénomènes quantiques, on décrit mathématiquement les éléments en jeu avec des fonctions d'onde. Ceci signifie en particulier que ces éléments ne sont pas considérés comme discrets. L'hypothèse de séparation ne se retrouve donc pas dans les modèles de phénomènes quantiques et le connexionnisme dans le cadre de Farmer ne pourrait donc pas en rendre compte. Le cadre de Farmer ne peut donc pas être considéré comme une "théorie de tout".

On peut donner un deuxième exemple de phénomène pour lequel l'absence de localité et d'immédiateté est liée à *l'échelle d'observation* qui a été fixée. Lorsque l'on décrit les phénomènes neuronaux, on peut considérer qu'ils sont immédiats et locaux pour une échelle d'observation bien choisie. Cependant il est possible de fixer une échelle d'observation et qu'à cette échelle on puisse interpréter l'activité de certaines cellules nerveuses (par exemple les neurones pyramidaux) selon les hypothèses de Farmer, et qu'on ne puisse accepter cette interprétation pour l'activité d'autres cellules nerveuses (comme les neurones moteurs qui agissent "à longue distance").

Si on interprète certains phénomènes neuronaux (neurones moteurs à une échelle bien choisie) à longue distance comme des phénomènes de diffusion, d'éloignement et qui se déroulent dans la durée, les modèles connexionnistes ne peuvent pas en rendre compte. Ceci ne signifie pas pour autant qu'on ne peut pas faire de modèles connexionnistes des neurones moteurs. *Tout dépend de la façon dont on les observe et ce dont on veut rendre compte à leur propos.* Il est possible de faire un modèle connexionniste d'un phénomène comportant des neurones moteurs, si on observe ce phénomène à d'autres échelles, pour lesquelles ce qu'on devait interpréter comme de la durée peut être considéré comme immédiat, ce qui semblait de l'éloignement peut être considéré comme local et ce qui semblait être de la diffusion peut être considéré d'un bloc, comme discret et séparé.

En ce sens les modèles connexionnistes sont une façon particulière de rendre compte des phénomènes. Ce choix implique des hypothèses sur le phénomène dont on veut rendre compte. Ces hypothèses ne sont pas nécessairement compatibles avec l'interprétation que l'on fait habituellement à une échelle donnée. À une autre échelle ces phénomènes pourront par contre être interprétés avec les hypothèses connexionnistes.

2.2.3. Connexionnisme et processus cognitifs

Il faut noter également que même si le connexionnisme ne constitue pas un cadre pour développer une “théorie de tout”, il ne permet pas pour autant de préciser la différence entre des processus cognitifs et des processus qui ne le seraient pas. Le connexionnisme permet de considérer sous le même format des modèles qui relèvent de phénomènes dont la sémantique nous semble très différente. Mais cette sémantique elle-même ne se retrouve pas dans les modèles considérés.

Si la cognition est une propriété de structure logique, et si on peut rendre compte de cette structure dans les termes du connexionnisme, il faut préciser quels modèles connexionnistes sont à même de le faire et lesquels ne le peuvent pas.

Si on peut rendre compte dans le même cadre formel du comportement d'un électron et d'un homme, par exemple, on considérera tout de même que seul le phénomène qui met en jeu l'homme est cognitif. *Dans ce cas la cognition est une affaire de contexte et pas seulement de structure logique particulière.* La notion de modèle connexionniste serait alors insuffisante pour rendre compte de la spécificité de la cognition, parce qu'elle ne tient pas compte du fait que *des sémantiques très différentes peuvent cohabiter dans un même phénomène et doivent alors être prises en compte dans le même cadre formel.* On trouvera des éléments qui vont dans ce sens dans le chapitre 5.

3. EN GUISE DE CONCLUSION SUR LA FAÇON D'ÉTUDIER LES PROCESSUS COGNITIFS

Comme nous l'avons constaté, le cadre de Farmer et la modélisation des processus n'ont pas suffi à distinguer les processus cognitifs de ceux qui ne le sont pas. La cognition n'est pas alors simple question de modèles formels, mais aussi de sémantique. On peut légitimement espérer que la spécification des processus cognitifs sera possible en tenant mieux compte des phénomènes qui sont en jeu. Dans ce cas *la mesure, parce qu'elle est en prise directe avec les phénomènes, est susceptible de nous préciser quels sont les phénomènes dynamiques qui sont cognitifs.* Pour réussir cela il faut expliciter comment on peut mesurer un phénomène dynamique¹.

Pour définir le champ des sciences cognitives, il faut tout d'abord affirmer que le propos est d'analyser, de comprendre et d'expliquer des processus. La mesure permet de proposer quelques critères supplémentaires qui permettent de restreindre le champ d'étude des sciences cognitives tout en conservant les spécificités déjà entrevues et liées à l'étude de processus. L'objectif ici est de montrer comment on peut limiter le champ de la cognition plus que cela n'a été fait jusqu'à présent. Les définitions que nous proposons sont en partie subjectives et tâtonnantes. Elles devraient permettre cependant d'aborder de façon constructive une difficulté essentielle passée bien souvent sous silence, sauf par les détracteurs de la discipline.

Lorsque nous disons qu'un processus est cognitif nous faisons en fait référence à une conception de la cognition où l'homme prend une place centrale. La cognition humaine est la seule à laquelle nous puissions effectivement accéder, soit par l'intermédiaire d'une forme de langage s'il s'agit de la cognition des autres, soit par l'intermédiaire de "l'introspection" lorsque nous faisons référence à des expériences personnelles. Il est alors naturel de délimiter les mesures sur les processus cognitifs en faisant référence aux mesures sur les processus cognitifs humains. La description macroscopique, c'est-à-dire à l'échelle de l'individu dans son ensemble, doit donc toujours être présente lorsque l'on parle de processus cognitif. Pour cela il faut d'abord définir des paramètres caractéristiques de la cognition. On précise ensuite des échelles de grandeur pour encadrer les domaines de ces paramètres susceptibles

¹Si l'on est rigoureux, ce sont les phénomènes dynamiques qui sont mesurés. Malgré cela, nous parlerons plutôt de mesure des processus, par abus de langage. De même, on dit que l'on mesure la masse de l'électron, c'est-à-dire la valeur d'un paramètre caractérisant un objet, alors qu'en fait on mesure cette valeur sur des phénomènes.

de relever des sciences cognitives.

Avant de préciser ces paramètres caractéristiques et les échelles de grandeur qui délimitent en partie la cognition, il faut d'abord choisir un niveau d'explication. On peut au moins distinguer trois niveaux d'explications possibles des phénomènes cognitifs. Ces différents niveaux d'explication sont complémentaires, mais *le choix d'un de ces niveaux entraîne des modifications très importantes dans le choix des paramètres pertinents et des échelles de grandeur en jeu*. Le premier niveau d'explication est celui de la psychologie et des Sciences Humaines. Ce premier niveau possible s'appuie sur des explications des différents phénomènes cognitifs qui relèvent du comportement individuel ou collectif. Les échelles en jeu sont celles de l'individu dans son ensemble. Le second niveau d'explication possible est celui de la biologie et des Sciences de la Vie. Il consiste à chercher à comprendre les différents phénomènes cognitifs à partir d'explications relevant avant tout de descriptions biochimiques du cerveau et donc des neurosciences. Un troisième niveau d'explication est possible à partir d'un niveau différent et encore plus petit, celui de la physique quantique et des Sciences de la Nature. Il est alors possible de donner des explications qui s'appuient sur des descriptions quantiques de la matière pour comprendre les phénomènes cognitifs. Ce dernier niveau d'explication quantique fait actuellement sourire les biologistes. Il est pourtant tout autant justifié que peut l'être le niveau biologique pour certains spécialistes du langage qui ne voient pas bien comment la biologie aujourd'hui pourrait apporter des éléments essentiels à leur discipline.

Ces trois niveaux de description sont en fait tout aussi possibles à l'heure actuelle. Il convient de savoir quels sont ceux que les sciences cognitives cherchent à développer. Il est tout à fait clair que le choix d'un de ces niveaux d'explication plutôt qu'un autre modifiera profondément les paramètres à mesurer et les plages de valeurs dans lesquelles on pourra considérer que le processus est cognitif. Il nous semble cependant d'après les éléments que nous avons apportés jusqu'à présent que *le niveau macroscopique qui tient compte de l'individu dans son ensemble doit toujours être présent*. En effet c'est le seul pour lequel le terme de cognition peut prendre son sens, puisqu'il fait référence directement à la cognition humaine, seule cognition à laquelle nous avons effectivement accès. Si on choisit alors le niveau d'explication biologique ou le niveau d'expliquer quantique, **une théorie de l'émergence est nécessaire pour faire le lien entre ce niveau d'explication et l'individu dans son ensemble**. Seule une explication purement psychologique² ne nécessite pas une telle théorie de l'émergence. Elle ne serait sans doute pas dans l'esprit des sciences cognitives car elle ne se situerait que dans le champ des Sciences de l'Homme. L'esprit des sciences cognitives est d'établir un lien entre les Sciences de la Vie ou celles de la Nature et les Sciences de l'Homme.

Le premier critère pour déterminer quels types de processus relèvent des sciences

²Une théorie qui fonderait des explications du comportement social en fonction de la cognition individuelle et de la psychologie en particulier nécessiterait également une théorie de l'émergence.

cognitives est un critère lié à l'échelle du processus. La conception sous-jacente fondamentale est alors qu'un processus à une échelle donnée admet des particularités que l'on ne retrouve pas à d'autres échelles. Cette conception est dans l'esprit très proche de celles sur l'observabilité en physique qui notent qu'un instrument d'observation doit avoir à peu près les mêmes dimensions que l'objet qu'il cherche à observer. En effet, si ce n'est pas le cas, l'instrument d'observation ne peut pas constater lui-même les modifications phénoménales en jeu lors de l'expérience. Dans le cas des sciences cognitives, il est tout à fait clair que les échelles en jeu sont celles qui correspondent aux dimensions humaines. En effet, ces dimensions sont les seules pour lesquelles le terme de cognition prend son sens.

On peut ensuite être plus précis sur ce critère d'échelle du processus. En effet, pour les objets habituels, on peut toujours dire que la **dimension** d'un objet est la valeur de la mesure de cet objet dans une des unités du Système International (SI). L'échelle sera alors définie par un ordre de grandeur, c'est-à-dire en fait un encadrement acceptable de la valeur de la mesure. Dans le cas des processus, ce critère de mesure ne peut pas être utilisé directement. En effet un processus étant l'abstraction d'un phénomène dynamique, on ne mesure pas directement ce phénomène dynamique, mais plutôt les modifications du phénomène statique qui le permet et qui en constitue l'observable. On peut donner un critère équivalent pour les processus à celui que l'on donne pour les objets, en considérant que la dimension d'un processus est mesurée sur les observables liés à ce processus. Plus précisément, on spécifie ces processus en affirmant qu'ils doivent être susceptibles d'entraîner des modifications observables et mesurables au niveau macroscopique. La dimension ne porte plus directement sur l'objet en jeu, mais sur les dimensions des objets qui portent les observables.

Cette détermination de la dimension d'un processus a une grande importance. Elle nous oblige à considérer comme partie intégrante des sciences cognitives des processus qui, en fait, mettent nécessairement en jeu différentes échelles. Nous revenons ainsi sur le fait que dans l'esprit des sciences cognitives un unique niveau d'explication est insuffisant et qu'il est nécessaire de le relier, quoi qu'il arrive, à l'individu dans son ensemble. Ainsi des études sur des processus moléculaires en jeu dans le cerveau seront considérées comme intéressantes, non pas parce qu'ils entraînent des modifications d'une certaine quantité de molécules, mais parce que certaines observables, par exemple comportementales, sont liées à ce processus moléculaire. Au contraire un processus moléculaire qui ne permettrait pas de rendre compte d'observables à l'échelle humaine ne serait pas considéré comme relevant des sciences cognitives. Ces études dans ce cas relèveraient seulement des neurosciences et non des sciences cognitives. *Les sciences cognitives nécessitent une conception de l'émergence parce que dans leur esprit le processus dynamique "émerge" du substrat statique.*

Nous devons alors définir plus précisément ce que nous appelons l'échelle humaine

d'observabilité. Nous appellerons cette échelle **échelle cognitive**. Cette échelle s'exprime dans diverses unités de mesure selon l'aspect dont on veut rendre compte. Certains phénomènes visuels, auditifs ou olfactifs sont à dimension humaine. Pour un phénomène visuel, l'échelle cognitive sera définie entre autres par l'ensemble des longueurs d'ondes perceptibles par un œil humain. Il est tout à fait évident que cette définition semblera restrictive à ceux qui ne veulent pas s'en tenir à la cognition humaine. Elle exclut par exemple les ultra-violets à 400 nanomètres (nm) et les infra-rouges à 780 nm, alors que certains infra-rouges sont perceptibles par les chats. Cela nécessite également de définir des conditions standard de l'environnement pour effectuer cette mesure, la vision nocturne étant différente de la vision diurne. De plus, en proposant cette définition nous faisons implicitement une hypothèse sur l'universalité des capacités perceptives de l'œil humain, les cas de daltonisme nous rappelant que cette hypothèse est restrictive. Ces capacités seraient à peu près les mêmes pour tous les hommes. Même avec cette hypothèse, il n'est pas clair qu'au cours du temps la notion d'homme ne va pas changer et intégrer par exemple des éléments artificiels créés par l'homme et qui prolongeront ses capacités perceptives. Il est possible, par exemple avec des lunettes spéciales, de prolonger les capacités visuelles humaines habituelles. Un homme peut alors percevoir certains infra-rouges.

La notion de capacité perceptive humaine, enfin, peut changer et a certainement changé au cours du temps, du fait de l'évolution de l'espèce. Nous définirons cependant une échelle visuelle qui comprendra les longueurs d'ondes entre 410 et 700 nm. Les maxima d'absorption pour un homme "normal" sont de 415 – 420 nm, 530 – 535 nm et de 560 – 565 nm pour les cônes et de 495 – 500 nm pour les bâtonnets. En ce qui concerne la vision, cependant, on ne peut pas considérer que l'unique paramètre en jeu soit la longueur d'onde. L'échelle cognitive, dans le cadre de la vision, devrait tenir compte d'un paramètre pour la perception des textures, etc... On pourrait et devrait certainement faire la même chose pour les différentes capacités cognitives humaines. Tous ces intervalles de valeurs admissibles pour chaque paramètre qui définissent l'échelle cognitive sont en partie arbitraires. Ces paramètres et ces intervalles sont cependant nécessaires si on veut essayer de cerner les capacités cognitives que nous cherchons à comprendre. Sinon, rien ne nous empêche de considérer comme une perception visuelle cognitive la réception sur une antenne du rayonnement fossile à 2,7 Kelvin du fond de l'univers, puisqu'elle traduit également une longueur d'onde, de l'ordre du millimètre.

De même, pour ces processus, des intervalles de temps caractéristiques doivent être précisés pour chaque phénomène dynamique considéré comme partie prenante de l'échelle cognitive. Plusieurs intervalles de temps doivent être définis. Les biologistes et les psychologues connaissent déjà des temps caractéristiques pour certains phénomènes. Il faut alors définir des encadrements de temps acceptables pour considérer les processus comme cognitifs.

Ce type de travail est minimal. On peut ensuite envisager de catégoriser différents

types de processus cognitifs en fonction de leur temps caractéristique. En sciences cognitives, on étudie des processus comme modification du substrat au cours du temps. Ceci explique pourquoi le temps est la seule grandeur caractéristique mesurée qui se retrouve dans la même unité international pour tous les processus cognitifs. Le temps est donc le seul paramètre qui permette de comparer des processus cognitifs qui se développent selon des grandeurs mesurées différentes. Ainsi, on peut comparer le temps de réaction olfactive ou visuelle à l'apparition d'un nouvel élément dans le champ réceptif. Par contre on ne pourra pas comparer une grandeur olfactive et une grandeur visuelle. Ainsi le temps est un paramètre particulier en sciences cognitives, car il est le seul qui permette de proposer une comparaison entre phénomènes cognitifs du point de vue de la mesure.

Il faut préciser que la difficulté pour définir des échelles cognitives n'est pas seulement liée à la définition un peu arbitraire des valeurs d'encadrement entre lesquelles des paramètres comme la longueur d'onde doivent se situer. En effet, même si l'on a une observable qui est perçue par un homme, encore faut-il être capable de la mesurer si l'on veut pouvoir donner une échelle cognitive caractéristique pour cette grandeur. Il semble tout à fait clair qu'une des difficultés par exemple pour la texture est de réussir à définir une bonne mesure qui parvienne à rendre compte de notre perception de la texture. Lorsque cette mesure est définie, on peut alors chercher à obtenir des encadrements caractéristiques pour une échelle cognitive de cette grandeur. Nous appelons **mesure** *une application, lorsqu'elle est constructible, qui va d'un ensemble de phénomènes dans l'ensemble des nombres réels.*

Comme l'ensemble des nombres réels est ordonné, cela entraîne une contrainte supplémentaire sur les phénomènes qui peuvent être effectivement mesurés. Une structure d'ordre des mesures est nécessaire pour délimiter des échelles cognitives caractéristiques. Il est donc nécessaire qu'une certaine notion d'ordre sur les phénomènes se retrouve de façon sous-jacente, pour que la mesure puisse en rendre compte. On appelle **phénomène mesurable** *un ensemble de phénomènes pour lesquels une application compatible avec la structure d'ordre est constructible.* La mesure constitue alors un point de vue particulier sur le phénomène considéré. La possibilité de mesurer le phénomène permet ensuite d'être plus exigeant lorsque l'on veut comparer le phénomène au comportement d'un modèle formel.

En résumé, deux éléments essentiels peuvent être dégagés des processus cognitifs. Le premier est la nécessité de définir une échelle cognitive caractéristique pour chaque grandeur mesurable. Les valeurs caractéristiques sont définies par rapport à des capacités humaines moyennes, même si la notion de capacité moyenne est évidemment relative. Une solution consiste à préciser cette valeur, même si elle ne correspond pas nécessairement à celle que l'on mesurera sur un échantillon "représentatif". Le second élément est la particularité de la notion de temps qui est la seule grandeur mesurable pouvant unifier les différentes capacités cognitives. En

effet, tous les processus ont des temps caractéristiques qui font référence à un même paramètre, le temps considéré comme absolu “de la physique”.

Cette réflexion sur la mesure des processus cognitifs nous amène à conclure que les mêmes hypothèses reviennent au niveau théorique et expérimental. *Dans ces deux cadres, les processus sont des abstractions de phénomènes dynamiques considérés comme les modifications de phénomènes statiques.* Trois notions viennent alors naturellement à l'esprit : la notion de **connexionnisme** apparaît tout d'abord, comme relation entre les éléments qui constituent ces phénomènes statiques ; la notion d'**émergence** arrive ensuite, qui doit permettre de justifier ou d'expliquer le passage du substrat au processus ; on peut s'interroger enfin sur la relation de nécessité entre les phénomènes statiques et leur abstraction comme objet d'une part, les phénomènes dynamiques et leur abstraction comme processus d'autre part.

La notion d'émergence et celle de connexionnisme nécessitent pour être pertinentes un nombre important d'éléments en interaction. Elles nous amènent inéluctablement à concevoir les sciences cognitives dans le cadre de la complexité, d'autant plus qu'on laisse alors de côté l'aspect “cognitif” des processus que l'on veut étudier. On se concentre ainsi sur certaines particularités dont rendent bien compte les modèles de la complexité, mais qui ne sont pas encore assez contraignantes de notre point de vue et ne déterminent pas assez la cognition.

RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 2

- [CNR, 1992] Sciences cognitives. *Le courrier du CNRS. Dossiers scientifiques*. 1992, vol. 79.
- [Andler, 1992] ANDLER D. Calcul et représentation : les sources. Dans *Introduction aux sciences cognitives*, Andler D., éditeur, p. 9–46. Gallimard, Paris, 1992. Livre correspondant à un colloque de Cerisy sur les sciences cognitives.
- [Charon, 1985] CHARON J. Imaginaire et réalité en relativité complexe. Dans *L'esprit et la science 2 : imaginaire et réalité*, p. 65–82. Colloque de Washington, Albin Michel, 1985.
- [Cybenko, 1989] CYBENKO G. Approximation by superposition of a sigmoidal function. *Math. of control, signals and systems 2*. 1989, p. 303–314.
- [Dupuy, 1985] DUPUY J.-P. L'essor de la première cybernétique. *Cahiers du CRÉA*. 1985, no 7, p. 9–139.
- [Farmer *et al.*, 1986] FARMER J., PACKARD N. ET PERELSON A. The immune system, adaptation and machine learning. *Physica D*. 1986, vol. 22, p. 187–204.
- [Farmer, 1990] FARMER J. D. A rosetta stone for connectionism. *Physica D*. 1990, vol. 42, p. 153–187.
- [Holland, 1986] HOLLAND J. Escaping brittleness: the possibilities of general purpose machine learning algorithms applied to parallel rule-based systems. Dans *Machine Learning II*, Michalski, Carbonell et Mitchell, éditeurs. Kaufmann, Los Altos, 1986.
- [Le Bellac, 1988] LE BELLAC M. *Des phénomènes critiques aux champs de jauge : une introduction aux méthodes et aux applications de la théorie quantique des champs*. InterÉditions/Éditions du CNRS, Paris, 1988. 639 p. Collection Savoirs actuels. ISBN 2-7296-0197-X 2-222-04026-4.
- [Le Ny, 1990] LE NY J.-F. Article Apprentissage. *Encyclopædia Universalis, Corpus*. 1990, vol. 2.

- [Marr, 1982] MARR D. *Vision: a computational investigation into the human representation and processing of visual information*. Freeman, New York, 1982. 397 p. ISBN 0-7167-1567-8.
- [Munos et Patinel, 1995] MUNOS R. ET PATINEL J. Bucket brigade et Q-learning. *Communication personnelle*. 1995.
- [Newell, 1990] NEWELL A. *Unified theories of cognition - the William James lectures, 1987*. Harvard University Press, Cambridge (Mass.), 1990. ISBN 0-674-92099-6.
- [Popper, 1978] POPPER K. *La logique de la découverte scientifique*. Reproduction de la première édition française parue en 1973. Payot, Paris, 1978. 484 p. Collection Bibliothèque scientifique. Traduit de The logic of scientific discovery. Première édition anglaise Hutchinson, 1959. Première édition allemande 1934. ISBN 2-228-11391-3.
- [Reeke et Edelman, 1987] REEKE G. ET EDELMAN G. Selective neural networks and their implications for recognition automata. *The international journal of supercomputer applications*. 1987, vol. 1, no 1, p. 44–69.
- [Saurel, 1992] SAUREL P. *Q-learning et apprentissage par renforcement*. DEA de sciences cognitives, EHESS, Paris, 1992.
- [Searle, 1985] SEARLE J. *Du cerveau au savoir : conférences Reith 1984 de la BBC*. Hermann, Paris, 1985. 144 p. Collection Savoir. ISBN 2-7056-6014-3.
- [Sontag, 1990] SONTAG E. Feedback stabilisation using two-hidden-layer nets. *Report Sycon, Rutger Center for System and Control*. 1990.
- [Varela, 1989] VARELA F. *Autonomie et connaissance : essai sur le vivant*. Première édition. Seuil, Paris, 1989. 256 p. Collection La couleur des idées. Traduction de Principles of biological autonomy. Première édition Elsevier, 1980, pour les chapitres 2, 3, 4, 6, 7 et 8. ISBN 2-02-010030-4.
- [Vauclair, 1992] VAUCLAIR J. *L'intelligence de l'animal*. Seuil, Paris, 1992. 256 p. Collection science ouverte. ISBN 2-02-013291-5.
- [von Neumann, 1992] VON NEUMANN J. *L'ordinateur et le cerveau*. Première édition française. La découverte, Paris, 1992. 130 p. Collection Textes à l'appui/série sciences cognitives. Traduction de The computer and the brain - the Silliman lectures. Première édition américaine, Yale University Press, New Haven, 1958. ISBN 2-7071-2164-9.
- [Wiener et Rosenblueth, 1945] WIENER N. ET ROSENBLUETH A. The role of models in science. *Philosophy of science*. 1945, vol. 12, p. 316–322.

[Wolfram, 1994] WOLFRAM S. *Cellular automata and complexity*. Première édition. Addison-Wesley, Reading (Mass.), 1994. 600 p. Collected papers de Stephen Wolfram. ISBN 0-201-62664-0.

CHAPITRE 3

**La complexité :
des théories de la quantité ?**

1. LES MODÈLES DE LA COMPLEXITÉ/QUANTITÉ, CAS PARTICULIERS DU CADRE DE FARMER

Dans ce chapitre, il pourrait sembler au lecteur que nous sommes bien loin des sciences cognitives et de l'étude du cerveau. Il n'en est rien. Bien au contraire nous sommes au cœur du sujet. En effet, dans l'optique que nous avons suivie au chapitre 2, les sciences cognitives, dans leur composante connexionniste¹, s'appuient sur une certaine conception de *l'émergence* qui lie le substrat considéré comme un objet au "processus émergent". Cette conception elle-même est justifiée par les modèles de la complexité qui sont les principaux modèles explicatifs de l'émergence à l'heure actuelle. C'est sur ces modèles et leur contexte que nous allons nous pencher dans ce chapitre. Il s'agit pour nous de cerner la complexité/quantité, d'étudier les outils qu'elle apporte et les méthodologies qu'elle peut utiliser. Les modèles de la complexité/quantité peuvent être décrits dans le cadre de Farmer². Les outils et les méthodologies employés pour les étudier pourront donc être également utilisés dans ce cadre.

Pour cela, nous envisagerons dans une première section, quelques *modèles paradigmatisques* de certaines théories de la complexité. Nous montrerons que les hypothèses qu'ils partagent permettent de parler de complexité/quantité. Dans une seconde section, nous étudierons *les outils proposés* actuellement pour analyser ces modèles. La complexité/quantité peut s'exprimer soit sous des formes complètement continues, soit sous des formes complètement discrètes. Cette dernière possibilité permet alors de se placer dans le cadre des automates cellulaires. Nous avons choisi dans cette section de nous concentrer sur les outils, discrets, utilisés par Wolfram pour analyser le comportement de certains automates. Nous montrerons que ces outils et les questions qu'ils suscitent sont imprégnés d'une conception de la complexité comme quantité. Dans une troisième section enfin, nous regarderons certaines méthodologies associées à la complexité/quantité. Pour ce faire nous présenterons trois modèles particuliers, utilisés dans des domaines extrêmement

¹Je laisse donc volontairement de côté tout ce qui pourrait relever d'une conception purement symbolique des Sciences Cognitives.

²Nous renvoyons à la section 2 du chapitre 2, ou au n°9627 des Rapports et Documents du CREA [Saurel, 1996], pour une présentation du cadre de Farmer.

différents, et nous montrerons que les méthodologies associées à chacun de ces modèles sont radicalement différentes.

Dans cette première section, nous allons examiner deux classes de modèles considérés comme représentants des modèles de la complexité et essayer de dégager les hypothèses qu'ils partagent et qui définissent les sciences actuelles de la complexité. Dans cette section nous allons donc être confrontés à deux formes de la complexité/quantité. La première version est continue et nous la rencontrons dans les modèles de type Ising, alors que la deuxième version est discrète et se retrouve pour les automates. Dans un premier temps nous allons examiner *les modèles de type Ising* et les outils qui en sont issus (sous-section 1). Nous présenterons ensuite un deuxième type de modèles, *les automates cellulaires* (sous-section 2). Ceci nous permettra de dégager les grandes caractéristiques des modèles de la complexité et de comprendre pourquoi ils servent de justification à la conception du cerveau comme complexe (sous-section 3).

Nous avons déjà dit rapidement dans le chapitre 2 qu'un lien étroit apparaissait entre la notion de complexité et les sciences cognitives. Une fois de plus le lien entre domaines apparemment très différents s'établit d'abord par l'intermédiaire des modèles utilisés. Ce lien est double. Il consiste tout d'abord en la possibilité de mettre en parallèle les hypothèses faites sur l'objet d'étude. Ce sont ces hypothèses semblables qui vont permettre de proposer des modèles identiques. De plus, on peut établir une relation étroite entre les modèles étudiés dans le cadre des sciences cognitives et ceux étudiés dans le cadre de la complexité. Ces deux domaines sont liés à l'étude de phénomènes dynamiques. L'abstraction de phénomènes dynamiques constitue ce que nous avons appelé processus. Pour l'instant nous avons insisté sur le fait que, pour les sciences cognitives, ces processus sont objets premiers d'étude. Les phénomènes dynamiques dont ils sont abstraits sont les seuls qui nous intéressent dans le cadre de la complexité comme dans celui des sciences cognitives. On considère également que, dans les deux cas, on décide de délaisser le substrat et de considérer que son étude ne nous apprendrait rien sur le phénomène dynamique.

Le cadre formel proposé par Farmer constitue en fait une généralisation qui englobe les modèles que nous allons présenter ici et qui sont issus de la physique. Ces modèles sont eux-mêmes les exemples paradigmatisques de la notion technique de complexité. On peut attribuer trois sens au terme de "complexité". *Le premier sens est celui qui oppose le complexe au simple, constitué d'un seul tenant.* Le complexe est alors ce qui est composé de plusieurs éléments, ce qui ne constitue pas un bloc indivis. *Le deuxième sens du mot complexité est synonyme de compliqué.* On utilise alors le terme de manière assez floue pour désigner tout ce que l'on ne sait pas encore expliquer de façon précise et scientifique. *Le troisième sens du terme complexité est technique.* Il provient essentiellement de l'étude de certains systèmes physiques. La complexité est alors "ce qui fait intervenir un grand nombre d'éléments en interactions mutuelles" [Lestienne, 1990][p. 18]. Cette définition technique est

généralement celle à laquelle on fait appel lorsque l'on utilise aujourd'hui le terme de complexité. La complexité fait alors référence à une quantité : le nombre d'éléments en interactions (et implicitement le nombre de relations possibles entre ces différents éléments). Nous utiliserons le terme de **complexité/quantité** pour faire référence à ce sens technique du mot complexité. Nous définissons ainsi de manière restrictive la complexité/quantité comme *ce qui fait intervenir un grand nombre d'éléments en interactions mutuelles*.

1.1. Les modèles de type Ising, exemple de complexité/quantité

Dans cette sous-section³ nous présentons les grandes lignes du modèle de référence des sciences de la complexité, le modèle d'Ising. Ce modèle permet de justifier l'étude des transitions de phase et se trouve également à l'origine des modèles de verres de spins. Dans le premier paragraphe nous aborderons tout d'abord le phénomène étudié, c'est-à-dire la transition ferromagnétique. Dans ce même paragraphe nous présenterons les notions de lois puissances, d'exposant critique et de loi d'échelle qui sont des signes caractéristiques des transitions de phase. Ces notions sont très utilisées pour étudier les modèles de complexité. Nous expliciterons ensuite, dans le deuxième paragraphe, les modèles de type Ising qui ont été proposés pour expliquer ces phénomènes. Dans ce même paragraphe nous aborderons également la méthode de résolution exacte développée en dimension un et qui a pu être généralisée en dimension deux. Cette méthode permet d'obtenir des résultats substantiellement différents en dimensions un et deux. Dans le troisième paragraphe, nous expliquerons deux méthodes d'étude approchée, l'approximation en champ moyen et la renormalisation. Nous montrerons comment l'approximation en champ moyen permet de simplifier les équations caractéristiques de ces modèles et d'envisager des résolutions formelles ou numériques. Nous donnerons ensuite les idées directrices des techniques de renormalisation et en particulier celle des blocs de Kadanoff qui permettent d'explorer la variation du comportement des gros systèmes lorsque l'on regroupe les éléments pour changer l'échelle d'étude. Dans un quatrième paragraphe, nous produirons quelques éléments permettant d'aborder l'étude de la dynamique des modèles de type Ising. Enfin, dans un cinquième paragraphe, nous commencerons à commenter les modèles des phénomènes critiques.

³Cette sous-section doit beaucoup aux manuels de référence que sont les livres de Diu [Diu *et al.*, 1989] et de Le Bellac [Le Bellac, 1988].

1.1.1. Le phénomène étudié et les observations associées

La transition ferromagnétique

Ising a introduit son modèle en 1925 [Ising, 1925] pour expliquer les propriétés de transition de phase de certains matériaux ferromagnétiques, comme le fer, le cobalt ou le nickel. Ces matériaux peuvent présenter une aimantation permanente, même en l'absence de champ magnétique extérieur. On peut être plus précis en comparant le comportement d'un morceau de fer, à température ambiante et à une température supérieure à 770°C .

À température ambiante, un morceau de fer est attiré par un aimant. On dit alors qu'il est dans un *état* ou une *phase ferromagnétique*. Cette aimantation spontanée du fer à température ambiante est expliquée par des considérations microscopiques. Les électrons d'une couche incomplète⁴ ont leurs spins⁵ pratiquement alignés. Les moments magnétiques de chacun de ces spins s'additionnent vectoriellement : comme les spins sont pratiquement alignés, les moments magnétiques s'ajoutent et permettent d'observer une aimantation macroscopique.

Au-dessus de 770°C , le morceau de fer n'est plus attiré par l'aimant. Il est alors

⁴Très grossièrement, on peut dire que le fer est un élément neutre et stable. Comme tous les éléments (cf. table de Mendeleïev), il est composé d'une part d'un noyau de protons et neutrons et d'autre part d'électrons positionnés sur des couches successives. On donne souvent comme image grossière celle d'une planète (le noyau) dont les satellites seraient les électrons situés à différentes orbites. Chaque orbite comporte un nombre maximum d'électrons. Grossièrement, lorsqu'une orbite est remplie on peut passer à l'orbite suivante. Chaque élément, et le fer en particulier, comme élément stable, est électriquement neutre. Il comporte donc un nombre fixe d'électrons. Ce nombre d'électrons ne permet pas de remplir complètement la dernière couche électronique. La dernière couche est alors incomplète.

⁵Le **spin** ou spin 1/2 est défini en mécanique quantique. “Une quantité d’observations expérimentales et d’arguments théoriques ont mené à la découverte de l’existence d’un *degré de liberté interne* pour l’électron : son *moment cinétique propre*” [Basdevant, 1987][Tome 1, p. 189]. Ce moment cinétique est une grandeur purement quantique, de valeur demi-entière. “La propriété fondamentale d’une *particule de spin* 1/2 est que lors de la mesure de la projection de son moment cinétique propre, *dorénavant appelé spin*, suivant un axe *quelconque*, les *seules* modalités observées sont les *deux* valeurs $+\hbar/2$ et $-\hbar/2$ ” [Basdevant, 1987][Tome 1, p. 191]. Mathématiquement, “l’observable S_z associée à la composante du spin le long de l’axe O_z n’a que deux valeurs propres, $+\hbar/2$ et $-\hbar/2$, qui ne sont pas dégénérées si la position ou l’impulsion de la particule est fixée” [Diu *et al.*, 1989][p. 975]. “Il en résulte que tout état de spin est superposition linéaire de deux états de base, dont le choix est arbitraire, et que, par conséquent, le *degré de liberté de spin* se décrit dans un espace de Hilbert à *deux dimensions*” [Basdevant, 1987][Tome 1, p. 191]. Mathématiquement, “l’espace des états d’une particule de spin 1/2 s’obtient par produit tensoriel de l’espace des états d’une particule sans spin (espace isomorphe à l’espace L^2 des fonctions de carré sommable) et de l’espace des états de spin, qui est de dimension 2. Il faut ainsi *deux fonctions d’onde* de carré sommable pour caractériser l’état d’une particule de spin 1/2” [Diu *et al.*, 1989][p. 975]. Il nous faut parler également du moment magnétique créé par les particules de spin 1/2, puisque les modèles de type Ising s’appuient exclusivement sur l’observable moment magnétique. “Au moment cinétique intrinsèque d’une particule est associé un moment magnétique qui lui est proportionnel” [Basdevant, 1987][Tome 1, p. 196].

passé dans une *phase paramagnétique*. La température T est le paramètre dont la modification permet le passage d'une phase à l'autre, on parle de **paramètre de contrôle**. La transition de la phase ferromagnétique à la phase paramagnétique a lieu à la température de 770°C qui est la **valeur critique** T_c du paramètre de contrôle. Cette température est également appelée "température de Curie". Lorsque le fer est à cette température seuil, on dit qu'il est à l'**état critique**. L'explication microscopique consiste à dire que l'aimantation spontanée à basse température est liée à une interaction entre spins qui favorise leur alignement. Au contraire, l'agitation thermique à haute température s'oppose à cet alignement.

Les phénomènes magnétiques ne sont pas les seuls pour lesquels on observe des transitions de phase. Par exemple l'eau présente trois phases, la phase solide (glace), la phase liquide (eau liquide) et la phase gazeuse (vapeur d'eau). Des transitions d'une phase à l'autre sont possibles. De plus on observe un point triple, défini par une pression et une température, pour lequel les trois phases de l'eau peuvent coexister. Ce point de séparation entre phases de l'eau correspond à un état critique. Les transitions de phase de l'eau sont cependant phénoménologiquement très différentes des transitions de phase magnétiques. En effet, pour l'eau, *on peut observer au cours de la transition une coexistence entre les différentes phases*. On dit alors que la transition est du premier ordre. Au contraire les transitions pour lesquelles on n'observe pas de coexistence entre les différentes phases sont dites du second ordre. La transition de l'état ferromagnétique à l'état paramagnétique est une transition du deuxième ordre.

Les observations macroscopiques sont les suivantes pour le morceau de fer, dont la structure est celle d'un cristal. Un cristal magnétique de volume V est en équilibre avec un thermostat à la température T et est soumis à un champ magnétique extérieur \vec{B}_0 , supposé uniforme. V , T et \vec{B}_0 constituent l'ensemble des paramètres d'état

extérieurs⁶. Ils doivent permettre à eux seuls d'expliquer les propriétés magnétiques du cristal. L'objectif est d'expliquer les propriétés phénoménologiques des transitions de phase à partir de trois grandeurs définies mathématiquement et dont les valeurs peuvent être obtenues expérimentalement : il s'agit de l'énergie magnétique moyenne, de la capacité calorifique et de l'aimantation. L'**aimantation** \vec{M} et sa norme M sont définies comme le moment magnétique moyen par unité de volume. L'aimantation ne dépend alors pas de V . Le but est de réussir à déterminer \vec{M} sous la forme d'une fonction dépendant des paramètres d'état extérieurs. La relation recherchée est appelée "équation d'état magnétique" du cristal, $\vec{M} = \vec{M}(T, \vec{B}_0)$. Prenons le cas d'un cristal paramagnétique parfait, c'est-à-dire sans interactions, et dont les éléments constitutifs sont des spins $1/2$ portant

⁶Les paramètres d'état ne sont pas tous des paramètres de contrôle. Les paramètres de contrôle sont des paramètres d'état sur lesquels on agit extérieurement pour voir apparaître les transitions.

chacun un moment magnétique $\mu = \frac{g\mu_b}{2}$, avec μ_b le magnéton de Bohr⁷ et g le facteur de Landé⁸. On obtient⁹ alors comme équation d'état magnétique du cristal $M = N \frac{g\mu_B}{2V} \text{th}\left(\frac{g\mu_B B_0}{2kT}\right)$.

Les caractéristiques des transitions : lois puissances, exposants critiques et équations universelles

Une **loi puissance** donne un ordre de grandeur et un type de comportement. On dit que $f(x)$ suit une loi puissance si on peut trouver deux nombres α et x_0 appartenant tous deux à \mathbb{R} et tels que¹⁰ $f(x) \sim (x - x_0)^\alpha$. On a

$$(f(x) \sim (x - x_0)^\alpha) \iff \left(\lim_{x \rightarrow x_0} \left(\frac{f(x)}{(x - x_0)^\alpha} \right) \text{ existe et est finie non nulle} \right)$$

Ceci signifie que $f(x)$ a le même comportement asymptotique que $(x - x_0)^\alpha$ en $x = x_0$. Concrètement, $f(x)$ admet un développement limité généralisé en $x = x_0$ et ce développement est de la forme $a \cdot (x - x_0)^\alpha$. Ce développement limité est généralisé dans le sens où les puissances du développement ne sont pas nécessairement entières, puisque $\alpha \in \mathbb{R}$. Comme tout “développement limité”, ces lois puissances sont seulement valables localement en $x = x_0$. Lorsque l'on obtient une telle loi puissance autour du point critique, on appelle α l'**exposant critique**. Dans le cas des phénomènes magnétiques, le rôle de x_0 est joué par la température critique T_c .

Dans les problèmes où interviennent des états critiques (phénoménologiques), on observe que certaines grandeurs mesurées obéissent à des lois puissances. Par exemple l'aimantation en champ nul $M_0(T)$ vérifie $M_0(T) \sim (T_c - T)^\beta$ pour $0 < (T_c - T)/T_c \ll 1$. Plus précisément on peut alors définir l'exposant critique¹¹ $\beta = \lim_{T \rightarrow T_c} \frac{\ln M_0(T)}{\ln(T_c - T)}$. Pour la capacité calorifique, on a $C_0(T) \sim (T - T_c)^{-\alpha}$ pour $0 < (T - T_c)/T_c \ll 1$. De même la susceptibilité¹² magnétique vérifie $\chi(T) \sim (T - T_c)^{-\gamma}$ pour $0 < (T - T_c)/T_c \ll 1$.

On observe expérimentalement que les valeurs obtenues pour ces exposants critiques sont pratiquement indépendantes du matériau considéré. Par exemple la valeur de β est toujours comprise entre 0,34 et 0,38, celle de α est de l'ordre de 0,1 et enfin γ vaut entre 1,3 et 1,4. On parle d'universalité de ces exposants critiques,

⁷Le magnéton de Bohr est défini comme $\mu_b = \frac{e\hbar}{2m_e}$. Avec $-e$ la charge de l'électron et m_e sa masse. Le magnéton de Bohr est homogène à un moment magnétique et sa valeur numérique est d'environ $9,27 \cdot 10^{-24} \text{ A.m}^2$.

⁸Ce facteur est proche de l'unité.

⁹On trouvera dans [Diu *et al.*, 1989][p. 313] le calcul qui permet d'obtenir ce résultat classique.

¹⁰Dans la littérature anglo-saxonne l'habitude est de noter l'équivalence \propto et non \sim .

¹¹L'exposant critique β est donc obtenu comme la pente caractéristique d'une droite sur un diagramme $\log - \log$.

¹²La susceptibilité magnétique est définie comme $\chi(T) = \lim_{B_0 \rightarrow 0} \frac{\partial M(B_0, T)}{\partial B_0}$.

car ils sont indépendants du matériau particulier considéré, du moment qu'une transition ferromagnétique est observée. "Cette universalité des exposants critiques, *a priori* surprenante, indique qu'ils ne dépendent que des caractéristiques générales de la transition de phase et non de la structure microscopique propre à chaque matériau. On comprend ainsi leur importance et pourquoi les physiciens se sont attachés à les déterminer le plus précisément possible, tant expérimentalement que théoriquement" [Diu *et al.*, 1989][p. 449].

En fait, lorsque quelques exposants critiques sont déterminés, on peut immédiatement obtenir la valeur des autres exposants critiques. On appelle **équation universelle** l'équation qui relie les différents exposants critiques. On obtient par exemple dans le cas du ferromagnétisme l'équation universelle¹³ $\alpha + 2\beta + \gamma = 2$. En fait on parle même de classe d'universalité, bien au-delà du ferromagnétisme, pour chaque phénomène présentant une transition de phase du second ordre. Pour un tel phénomène, on peut trouver une équation universelle entre les exposants critiques. Chaque équation universelle permet alors, en retour, d'obtenir toute une classe de phénomènes de transitions de phase du second ordre. Cette classe s'appelle **classe d'universalité**. Tous les phénomènes ferromagnétiques, par exemple, sont dans la même classe d'universalité. On peut alors, également au niveau des modèles, essayer de dépasser chaque modèle spécifique pour accéder à des classes de modèles qui correspondront à des classes d'universalité.

1.1.2. Les modèles de type Ising et les méthodes d'étude exactes

Introduction aux modèles de type Ising

Pour modéliser ce phénomène, l'hypothèse principale consiste à supposer que le ferromagnétisme s'explique essentiellement par l'interaction entre les spins, qui tend à les aligner. La modélisation doit donc permettre de rendre compte du comportement global et des observations macroscopiques faites sur un cristal ferromagnétique en utilisant un modèle microscopique. Les modèles vont donc considérer un nombre important de spins en interaction, qui vont donner lieu à un champ magnétique global. Les exemples les plus importants sont les modèles d'Ising et de Heisenberg. Nous appellerons parfois modèles de type Ising l'ensemble des modèles que l'on peut considérer comme des généralisations du modèle d'Ising [Ising, 1925]. Ces modèles ne nécessitent pas obligatoirement une phénoménologie de type spin, mais une modélisation des éléments en interaction qui peut prendre une forme identique à celle des spins. Parmi ces modèles de type Ising on trouve en particulier les modèles appelés **verres de spins**. La distinction entre modèle d'Ising et modèles de verres de spins porte sur l'ensemble des valeurs possibles des spins \vec{S}_i du modèle. Dans le

¹³On peut obtenir cette équation universelle grâce à la méthode de renormalisation. On pourra observer sur la figure 3 la correspondance entre l'équation universelle et les différentes exposants critiques obtenus.

cas du modèle d'Ising, les \vec{S}_i ne peuvent prendre que deux valeurs selon une seule direction, alors qu'ils peuvent prendre une infinité de valeurs dans le cadre de verres de spins¹⁴. Pour le modèle de Potts [Potts, 1952], les spins peuvent prendre un nombre fini de r valeurs, de 1 à r . En tant que modèles, c'est-à-dire de systèmes formels, les modèles de type Ising pourront éventuellement s'appliquer à des, ou être issus de, phénomènes totalement différents des phénomènes magnétiques.

Les modèles de type Ising, cas particuliers du cadre de Farmer

Tous ces modèles de type Ising constituent un cas très particulier du cadre défini par Farmer¹⁵. En effet, dans ces modèles un graphe est défini, qui est généralement un réseau euclidien. Il s'agit donc d'un graphe particulièrement régulier pour lequel on peut même définir une dimension. Plus précisément, nous dirons que le graphe est **homogène pour les voisinages** lorsque, comme c'est le cas ici, les voisinages sont les mêmes en tout site du graphe. On parle de modèles en dimension un, deux ou trois, qui correspondent à des descriptions spatiales, liées à la disposition des sites les uns par rapport aux autres.

Dans ces modèles chaque site est simple et peut prendre un ensemble de valeurs, soit un ensemble binaire de valeurs (modèle d'Ising), soit un ensemble binaire de valeurs en trois dimensions (modèle de Heisenberg), soit un nombre fini de valeurs en dimension un (modèles de Potts), soit un nombre infini de valeurs en dimension un (modèles de verres de spins). En ce qui concerne les sites, on se trouve donc bien dans différents cas particuliers du cadre de Farmer.

Le troisième élément constitutif du cadre de Farmer est relatif aux dynamiques. Pour les modèles d'Ising, les dynamiques sont en général absentes, et nous verrons un peu plus loin pour quelles raisons. Lorsqu'il n'y a pas de dynamiques, on se trouve bien dans un cas très dégénéré du cadre de Farmer. On peut cependant trouver quelquefois des dynamiques pour les modèles de type Ising (annexe 1). Elles seront alors définies localement comme identiques en tout site. Nous dirons alors que le graphe est **homogène pour les dynamiques**, lorsque les dynamiques sont les mêmes en tout site du graphe. Ces dynamiques porteront généralement sur les valeurs des sites, elles consisteront à modifier les spins, et on parlera de "retournement de spins". Dans ces modèles on ne trouve donc pas, habituellement, de dynamique sur les liens. Il faut noter cependant que les modèles plus récents de verres de spins possèdent également des dynamiques sur les J_{ij} , c'est-à-dire sur les couplages entre spins. Ces dynamiques constituent bien des dynamiques d'apprentissage

¹⁴Cette modification entraîne des difficultés techniques substantielles. Une partie de ces difficultés a été résolue grâce à la *méthode dite des répliques*. Il est hors de question pour nous de présenter ici cette technique, que personnellement nous ne maîtrisons pas. Un recueil d'articles techniques constitue la référence actuelle sur le sujet [Mézard *et al.*, 1987].

¹⁵Voir le chapitre 2, section 2, pour plus de précisions sur le cadre de Farmer.

au sens de Farmer. À notre connaissance, on ne trouve cependant pas, dans la communauté travaillant sur ces modèles, de travaux ayant introduit des dynamiques du graphe. Dans tous les cas, comme nous l'avons montré, les dynamiques présentes dans ces modèles peuvent bien être décrites directement dans le cadre proposé par Farmer.

Finalement on retrouve bien dans ces modèles les trois éléments spécifiés par Farmer, et on peut même les décrire directement à partir du vocabulaire qu'il a proposé. De plus, ces modèles ont permis de développer des outils extrêmement puissants, que Farmer ne présente pas du tout.

La construction du modèle de Heisenberg

Pour construire un tel modèle local, il faut parvenir à représenter les interactions entre les différents spins pour tous les sites du cristal. Comme nous l'avons écrit, c'est en fait de l'interaction entre les différents moments magnétiques issus des spins que nous allons rendre compte. Pour exprimer le facteur énergétique entre les sites, on cherche la forme la plus simple possible. Une contrainte s'impose sur ce facteur dans le cadre des modèles issus du magnétisme¹⁶, la nécessité de tenir compte de la symétrie supposée des interactions entre les sites i et j . On définit ce facteur comme $H_{ij} = -J_{ij} \vec{S}_i \cdot \vec{S}_j$ avec \vec{S}_i et \vec{S}_j les spins aux sites i et j et J_{ij} le coefficient¹⁷ d'interaction entre les sites i et j . On définit alors l'**hamiltonien de Heisenberg** comme l'hamiltonien ou le facteur énergétique qui s'écrit, en appelant R le réseau dans son ensemble :

$$H_H = \sum_{i \in R} \left(-g \mu_b \vec{B}_0 \cdot \vec{S}_i - \sum_{j \in v(i)} J_{ij} \vec{S}_i \cdot \vec{S}_j \right)$$

ou

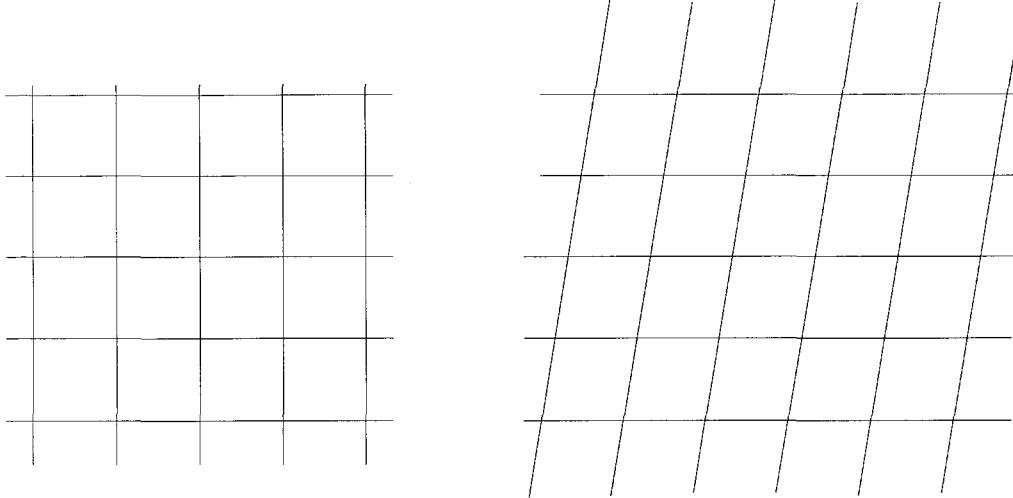
$$H_H = \sum_{i \in R} \vec{S}_i \cdot \left(-g \mu_b \vec{B}_0 - \sum_{j \in v(i)} J_{ij} \vec{S}_j \right)$$

À chaque spin on associe alors le moment magnétique $\vec{\mu}_i = g \mu_b \vec{S}_i$ avec μ_b le magnéton de Bohr et g le facteur de Landé. On obtient l'hamiltonien de Heisenberg dans un cas très simple lorsque le terme d'interaction est celui défini précédemment.

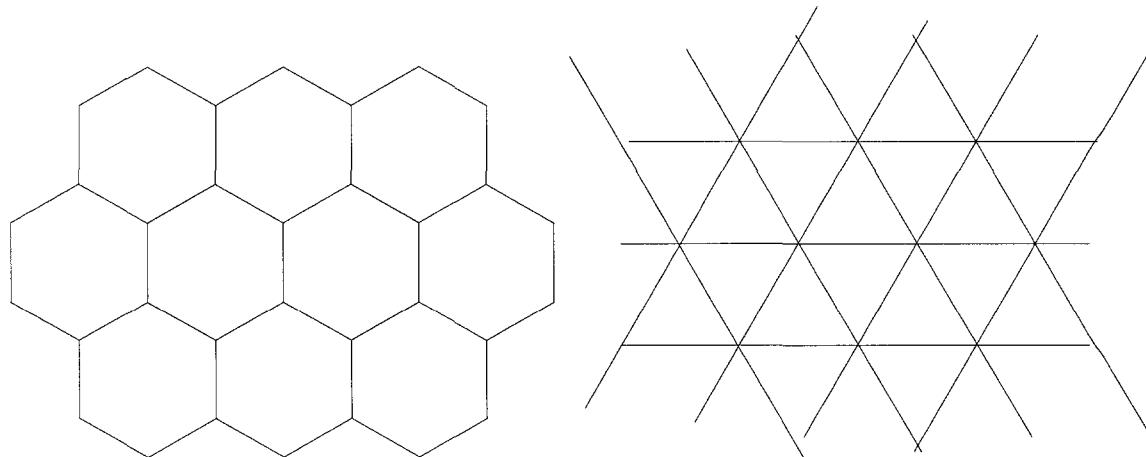
¹⁶Cette contrainte a été levée dans les modèles plus récents appelés à proprement parler modèles de verres de spins et qui servent de modèle pour les réseaux de neurones. On trouvera une présentation sur ce sujet dans [Mézard et Toulouse, 1991].

¹⁷Dans le cadre des modèles issus du magnétisme, ce facteur est positif et symétrique. Il est appelé **coefficient de couplage**. Ce coefficient n'est plus nécessairement symétrique dans les modèles de verres de spins utilisés pour les réseaux de neurones [Mézard et Toulouse, 1991][p. 623].

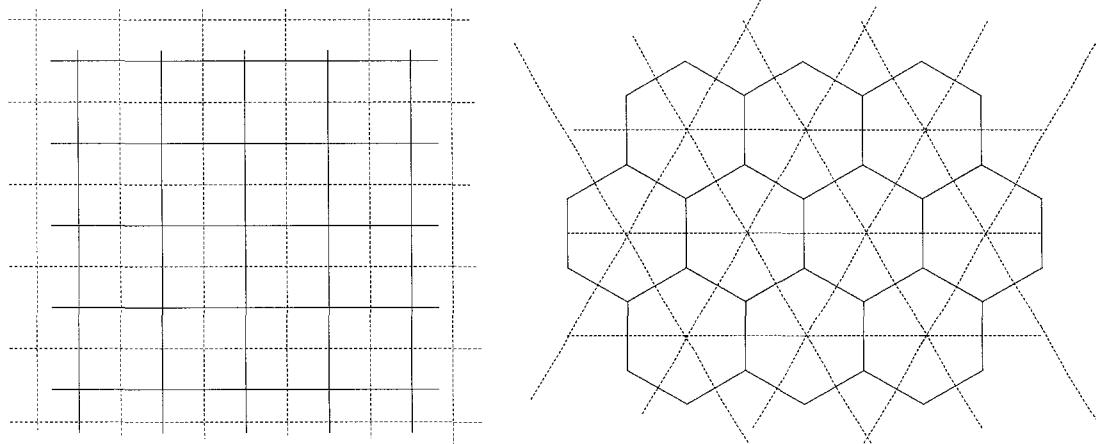
Il faut noter également que ce facteur J_{ij} est nul si les sites i et j ne sont pas voisins. On notera $j \in v(i)$ le fait que le site j appartient à un voisinage du site i .



Un réseau carré (ou euclidien) en dimension deux présenté sous différents aspects. On parle de réseau carré du fait de la répartition des voisins et non de la forme particulière de la maille élémentaire.



Un réseau hexagonal et un réseau triangulaire



Un réseau carré avec son réseau dual, carré, associé et un réseau hexagonal avec son réseau dual, triangulaire, associé. Les liens du réseau dual coupent chacun un lien du réseau initial. Les sites du réseau dual sont donc situés au milieu des cellules du réseau initial et réciproquement. La notion de réseau dual permet d'obtenir certains résultats théoriques exacts.

FIGURE 1 : Différents réseaux utilisés pour les modèles de type Ising.

Chapitre 3 -- Figure 1

Il suffit de prendre N atomes situés sur les sites d'un réseau cristallin avec un champ uniforme \vec{B}_0 et de tenir compte uniquement des interactions entre les plus proches voisins (figure 1). L'hamiltonien de Heisenberg est la superposition de deux termes, $H_H = H_H^a + H_H^b$; le premier terme H_H^a provient du champ magnétique extérieur :

$$H_H^a = -g \mu_b \vec{B}_0 \cdot \left(\sum_{i \in R} \vec{S}_i \right)$$

alors que le second terme H_H^b est issu des interactions entre plus proches voisins¹⁸ :

$$H_H^b = - \sum_{\langle i,j \rangle} J_{ij} \vec{S}_i \cdot \vec{S}_j$$

L'hamiltonien du système H ou H_H est une grandeur qui permet de mesurer l'état énergétique des différentes configurations possibles du système. L'hamiltonien permet de calculer la fonction de partition Z qui est définie comme un nombre sans dimension et qui vaut $Z = \sum_l e^{-E_l/kT}$, où la somme porte sur toutes les valeurs E_l de l'énergie que le système peut présenter¹⁹ et k la constante de Boltzmann. L'énergie libre est définie comme $F(B_0, T) = -kT \ln Z$. La connaissance de la fonction de partition et de l'énergie libre permet d'obtenir les valeurs des différentes grandeurs thermodynamiques.

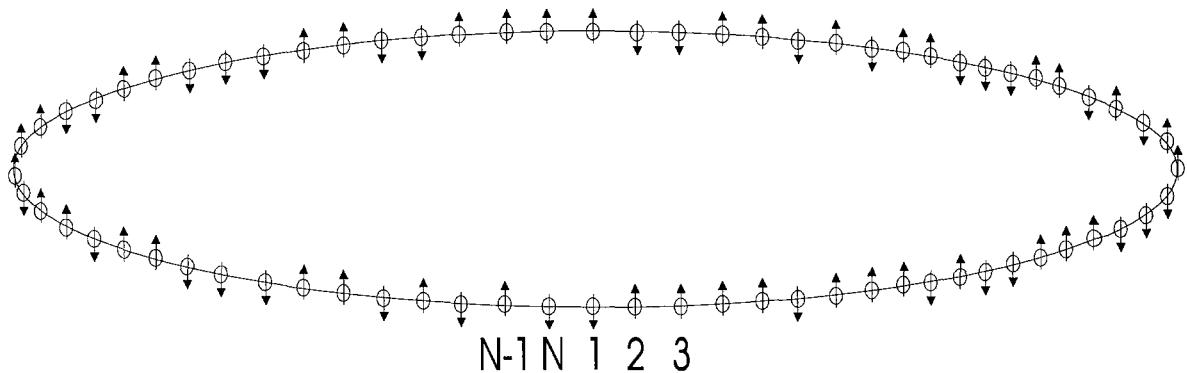
La valeur de $M(B_0, T)$ à l'équilibre est celle qui rend F minimal ; elle est donnée par $\frac{\partial F}{\partial M} = 0$. Les états d'équilibre sont donc déterminés par les *minima* de l'énergie libre F par rapport à M . On peut alors également déterminer la stabilité du système en fonction de la forme et du signe des dérivées partielles successives de l'énergie libre en chaque point d'équilibre. Ceci est un simple prolongement des méthodes développées en mécanique du point avec la notion d'énergie mécanique.

Le modèle d'Ising et ses solutions exactes

Le modèle d'Ising est un cas très particulier dont l'hamiltonien est une simplification de celui de Heisenberg. La difficulté principale avec l'hamiltonien de Heisenberg est liée au fait que les trois composantes d'un opérateur de spin ne commutent pas. Le modèle d'Ising est une approximation du modèle de Heisenberg qui ne tient compte que d'une seule direction pour les spins, par exemple \vec{O}_z . Au lieu de travailler sur des spins vectoriels \vec{S}_i , dans le cas du modèle d'Ising on utilise uniquement des spins scalaires S_i^z . On peut également interpréter le modèle d'Ising comme le modèle de

¹⁸La notation $\langle i, j \rangle$ indique que la somme porte sur tous les couples de plus proches voisins.

¹⁹Ces valeurs de l'énergie E_l sont généralement l'ensemble des valeurs que peut prendre l'hamiltonien H_H .



Le modèle d'Ising en dimension un : une chaîne linéaire finie de N spins prenant deux valeurs (représentation vers le haut ou vers le bas). On choisit de représenter la chaîne sous la forme d'un tore pour compenser d'éventuels effets de bord. La périodicité imposée grâce au tore permet avec un nombre fini d'éléments de faire comme si on avait un nombre infini d'éléments. On peut ainsi étudier sur le tore les propriétés de l'ensemble. Ces propriétés sont en principe valables uniquement à la limite thermodynamique, c'est-à-dire sous la contrainte d'avoir en particulier un nombre infini d'éléments. La mise sous forme de tore permet de contourner cette contrainte. En dimension un le tore est en fait un anneau, ou encore un cercle.

FIGURE 2 : Le modèle d'Ising en dimension un, représenté sous la forme d'un cercle.

Heisenberg que l'on essaierait d'étudier en champ moyen²⁰, sauf pour la composante selon \vec{O}_z .

Après avoir défini l'hamiltonien dans le cas du modèle d'Ising, le second point consiste à préciser la topologie du modèle et les états accessibles de chaque site (figure 1). Les sites sont définis sur un réseau carré. En chaque site, on trouve un spin qui peut avoir deux valeurs différentes, +1 ou -1. On suppose également que les interactions sont constantes sur tout le réseau, c'est-à-dire que $J_{ij} = J$. Le but est alors de savoir si on peut identifier une transition de phase, c'est-à-dire dans le cas du modèle à deux valeurs de spin, s'il existe des états d'équilibre comportant à la fois des systèmes dont les spins sont globalement dans le même état et qui constituent des systèmes ordonnés et d'autres états d'équilibre pour des systèmes dont les sous blocs de spins sont dans des états mélangés, c'est-à-dire des systèmes désordonnés. Or le signe distinctif de ces transitions de phase est l'existence de singularités pour certaines fonctions caractéristiques du système comme l'énergie libre ou l'aimantation.

Pour un cristal en dimension un, le modèle d'Ising est une chaîne de N sites. Chaque site interagit avec ses deux voisins les plus proches. On peut noter cependant que si la chaîne est finie, en bout de chaîne chaque site n'a plus qu'un seul voisin. Pour conserver des propriétés de symétrie et que chaque site puisse tout de même jouer le même rôle, le modèle d'Ising en dimension un et pour un nombre fini de sites relie le dernier site au premier, comme sur un anneau (figure 2). Pour un tel modèle, il est possible de calculer les solutions exactes, c'est-à-dire d'obtenir une expression analytique à la fois de la fonction de partition, de l'énergie libre et de l'aimantation.

Dans ce cas, on peut alors écrire l'hamiltonien sous la forme :

$$H_I = - \sum_{i=1}^N (g \mu_b B_0 S_i^z + JS_i^z S_{i+1}^z)$$

où le site $N + 1$ et le site 1 coïncident. Le système peut alors présenter 2^N états différents et la fonction de partition a alors pour expression :

$$\begin{aligned} Z &= \sum_{S_1^z=\pm 1} \sum_{S_2^z=\pm 1} \dots \sum_{S_N^z=\pm 1} \exp \left[\frac{1}{kT} \sum_{i=1}^N [g \mu_b B_0 S_i^z + JS_i^z S_{i+1}^z] \right] \\ \text{soit } Z &= \sum_{S_1^z=\pm 1} \sum_{S_2^z=\pm 1} \dots \sum_{S_N^z=\pm 1} \prod_{i=1}^N \exp \left[\frac{1}{kT} (g \mu_b B_0 S_i^z) \right] \exp \left[\frac{1}{kT} (JS_i^z S_{i+1}^z) \right] \end{aligned}$$

On peut alors poser $T(S_i^z, S_j^z) = \exp \left[\frac{1}{kT} \left[\frac{g \mu_b B_0}{2} (S_i^z + S_j^z) + JS_i^z S_j^z \right] \right]$ et écrire Z en fonction de $T(S_i^z, S_j^z)$. L'objectif est d'obtenir une écriture où S_i^z et S_j^z (en

²⁰Pour la notion de champ moyen, voir le paragraphe suivant.

l'occurrence S_i^z et S_{i+1}^z) jouent un rôle symétrique. On obtient alors :

$$Z = \sum_{S_1^z=\pm 1} \sum_{S_2^z=\pm 1} \dots \sum_{S_N^z=\pm 1} T(S_1^z, S_2^z) T(S_2^z, S_3^z) \dots T(S_N^z, S_1^z)$$

Si l'on pose alors P la matrice 2×2 dont les coefficients sont les $T(S_i^z, S_j^z)$, on obtient :

$$P = \begin{pmatrix} T(+1, +1) & T(+1, -1) \\ T(-1, +1) & T(-1, -1) \end{pmatrix}$$

et on note $(P)_{S_i^z S_j^z}$ le coefficient de P qui correspond à la première ligne si $S_i^z = +1$, et à la deuxième ligne si $S_i^z = -1$, à la première colonne si $S_j^z = +1$, et à la deuxième colonne si $S_j^z = -1$. On observe alors que

$$\begin{aligned} & \sum_{S_j^z=\pm 1} T(S_i^z, S_j^z) T(S_j^z, S_k^z) \\ &= T(S_i^z, +1) T(+1, S_k^z) + T(S_i^z, -1) T(-1, S_k^z) \\ &= (P^2)_{S_i^z S_k^z} \end{aligned}$$

Cette propriété nous permet d'écrire Z plus simplement :

$$\begin{aligned} Z &= \sum_{S_1^z=\pm 1} (P^N)_{S_1^z S_1^z} \\ \text{soit } Z &= \text{Trace } (P^N) \end{aligned}$$

La matrice P étant réelle et symétrique, elle est diagonalisable dans une base de vecteurs propres et ses valeurs propres λ_1 et λ_2 sont réelles. On peut alors écrire : $Z = \lambda_1^N + \lambda_2^N$. En prenant $|\lambda_1| < |\lambda_2|$, on peut écrire : $Z = \lambda_1^N \left[1 + \left(\frac{\lambda_2}{\lambda_1} \right)^N \right]$. On fait alors l'hypothèse : $N \rightarrow +\infty$. On obtient alors : $Z \sim \lambda_1^N$. Il est possible de calculer les valeurs propres²¹, et en particulier λ_1 . On peut alors calculer la valeur²² de l'énergie libre, qui est donnée par : $F(B_0, T) = -kT \ln Z$. L'aimantation est définie comme la valeur moyenne $M(B_0, T) = \frac{g\mu_b \sum_{i=1}^N S_i^z}{L}$, avec L pour la longueur de la chaîne. On peut alors remarquer que $M(B_0, T) = \frac{kT}{L} \frac{\partial \ln Z}{\partial B_0} = -\frac{\partial F(B_0, T)}{L \partial B_0}$. On obtient donc pour l'énergie une fonction de B_0 et T continue, à dérivées continues. *Le modèle d'Ising unidimensionnel ne présente donc pas de transition de phase.*

Pour un cristal en dimension deux, le modèle d'Ising a la structure d'un réseau carré, c'est-à-dire d'un réseau euclidien (figure 1). Chaque site est alors relié à

²¹On obtient : $\lambda_1 = \exp\left(\frac{J}{4kT}\right) \times \left[\operatorname{ch}\frac{g\mu_b B_0}{2kT} + \sqrt{\operatorname{sh}^2\left(\frac{g\mu_b B_0}{2kT}\right) + \exp\left(\frac{-J}{kT}\right)} \right]$.

²²On obtient : $F(B_0, T) = -NkT \left[\frac{J}{4kT} + \ln \left[\operatorname{ch}\frac{g\mu_b B_0}{2kT} + \sqrt{\operatorname{sh}^2\left(\frac{g\mu_b B_0}{2kT}\right) + \exp\left(\frac{-J}{kT}\right)} \right] \right]$

On obtient alors en l'absence de champ extérieur : $F(B_0 = 0, T) = -NkT \ln\left(2\operatorname{ch}\frac{J}{4kT}\right)$.

ses quatre voisins les plus proches si l'on ne considère que les directions orthogonales. Dans le cas où le champ magnétique extérieur est nul²³ ($\vec{B}_0 = \vec{0}$), Onsager [Onsager, 1944] a obtenu l'expression analytique exacte de la fonction de partition par une méthode qui généralise²⁴ en dimension deux la méthode développée ci-dessus en dimension un. On peut ensuite obtenir directement l'énergie libre et on observe que l'aimantation en champ nul admet une singularité. Ce résultat est particulièrement important car c'est lui et lui seul qui justifie de parler de transition de phase pour tous les modèles de ce type. “Insistons encore sur l'importance théorique considérable de la solution du modèle d'Ising à deux dimensions : *un calcul exact* de la fonction de partition d'un système de particules en interaction mutuelle *prévoit l'existence d'une transition de phase* sans la postuler *a priori* (comme on le fait dans l’“approximation” de champ moyen). Les méthodes de la mécanique statistique sont donc aptes à décrire de tels phénomènes, et c'est seulement pour des raisons de complications mathématiques que l'on arrive rarement à un résultat sans devoir faire appel à des approximations (pas toujours bien contrôlées)” [Diu *et al.*, 1989][p. 474]²⁵.

Onsager a également obtenu [Onsager, 1944][p. 140] pour le modèle d'Ising en dimension deux un développement asymptotique de la capacité calorifique au voisinage du point critique qui est de type logarithmique. On ne peut alors pas dire qu'on obtient une loi puissance. Par abus de langage on dit alors que l'exposant critique associé à la capacité calorifique est nul.

Il faut noter également que ces résultats mettent en évidence une différence qualitative essentielle entre le modèle d'Ising à une dimension et celui à deux dimensions. La dimension de l'espace peut donc avoir une importance cruciale dans ce type de modèles.

Rapide historique des méthodes exactes de résolution du modèle d'Ising

Pour donner une idée des difficultés associées à ces modèles, on peut donner la chronologie suivante, en précisant que les meilleurs scientifiques de l'époque s'étaient attaqués à la question. Les explications causales du ferromagnétisme données initialement par Weiss datent de 1907 – 1908. En 1920 Lenz énonce ses suggestions sur le lien entre l'orientation des champs magnétiques locaux et la valeur globale de l'énergie du système. Ising a défini son modèle original en 1924 [Ising, 1925]. La même année il proposait une solution pour le cristal en dimension un. En 1928, Heisenberg proposait son modèle. C'est seulement en 1943 que Onsager réussit à calculer l'énergie libre pour le modèle d'Ising en dimension deux [Onsager, 1944].

²³“Il n'a pas été possible, jusqu'à présent, de résoudre exactement le problème en présence de champ extérieur” [Diu *et al.*, 1989][p. 470].

²⁴La méthode employée peut s'appliquer avec des liens J_{ij} qui ont une valeur constante J suivant un axe et une valeur constante J' suivant la direction orthogonale.

²⁵Les italiques sont de l'auteur.

En 1949, Onsager annonçait le résultat pour l'aimantation en dimension deux. Mais c'est seulement en 1952 que Yang publiait le calcul qui permet d'obtenir l'aimantation pour le modèle d'Ising en dimension deux.

1.1.3. Les méthodes d'étude approchées

La méthode d'approximation du champ moyen et son application au modèle de Heisenberg

Pour étudier la structure d'un tel cristal, il faut d'abord obtenir la fonction de partition qui, comme nous l'avons vu, peut être obtenue à partir de l'hamiltonien. Il n'a pas été possible de calculer formellement l'expression analytique en fonction des variables du système de la fonction de partition pour un système décrit par l'hamiltonien de Heisenberg. Comme souvent quand on essaye d'étudier un système de particules en interactions, on ne peut obtenir une solution analytique exacte. La première technique, la plus simple, est alors d'étudier une approximation de champ moyen du système initial. On espère que cette approximation²⁶ offrira une solution correspondant qualitativement au système initial.

L'idée générale de ce principe d'approximation a été développée par Weiss en 1907. Dans le cas du modèle de Heisenberg, l'hamiltonien s'écrit :

$$H_H = \sum_{i \in R} \vec{S}_i \cdot \left(-g\mu_b \vec{B}_0 - \sum_{j \in v(i)} J_{ij} \vec{S}_j \right)$$

On peut donc écrire :

$$H_H = - \sum_{i \in R} \vec{S}_i \cdot g\mu_b \vec{B}_i \text{ avec } \vec{B}_i = \vec{B}_0 + \frac{1}{g\mu_b} \sum_{j \in v(i)} J_{ij} \vec{S}_j$$

Le terme \vec{B}_i est un champ magnétique qui comporte à la fois \vec{B}_0 et un terme supplémentaire appelé "champ moléculaire" qui comprend les interactions des autres sites sur le site i . L'approximation de champ moyen, dans ce cas, consiste à considérer le terme $\frac{1}{g\mu_b} \sum_{j \in v(i)} J_{ij} \vec{S}_j$, c'est-à-dire le champ moléculaire en i , comme constant en tout point du réseau. On pose alors p le nombre d'éléments dans le voisinage d'un site, c'est-à-dire le cardinal de l'ensemble $v(i)$, ce nombre étant constant sur tout le réseau et donc étant indépendant de i . La valeur moyenne du champ moléculaire est alors $\frac{pJ}{g\mu_b} \langle \vec{S}_i \rangle$. On peut considérer ce terme comme un champ local ajouté au champ global \vec{B}_0 . On pose alors : $\vec{B}_i = \vec{B}_{eff}$. Cette approximation consiste à négliger les fluctuations locales du champ. L'approximation

²⁶D'autres techniques d'approximations peuvent être utilisées. On note les techniques d'approximation numérique, les techniques de perturbation, mais aussi les techniques de renormalisation que nous aborderons un peu plus loin.

de champ moyen néglige les fluctuations du champ moléculaire qui est remplacée par sa valeur moyenne sur l'ensemble du cristal²⁷. Ceci nous donne : $\overrightarrow{B_{eff}} = \overrightarrow{B_0} + \frac{J}{Ng\mu_b} \langle \overrightarrow{S_i} \rangle$. Si de plus chaque spin a la même valeur moyenne, on obtient alors pour l'aimantation : $\overrightarrow{M} = \frac{N}{V} g\mu_b \langle \overrightarrow{S_i} \rangle$. On obtient donc : $\overrightarrow{B_{eff}} = \overrightarrow{B_0} + \lambda \overrightarrow{M}$; avec $\lambda = \frac{VpJ}{N(g\mu_b)^2}$.

Dans l'approximation du champ moyen, l'hamiltonien de Heisenberg est égal à l'hamiltonien de N spins indépendants soumis à un champ magnétique extérieur uniforme. Le nouveau problème consiste maintenant à étudier N spins indépendants soumis à un champ magnétique uniforme. Pour la magnétisation, on retrouve un problème identique à celui du cristal paramagnétique parfait, avec cette fois un champ magnétique qui est égal à B_{eff} et non à B_0 . Dans ce cas le terme constant joue un rôle complètement différent, puisqu'il dépend de l'aimantation M . Le principe de l'approximation en champ moyen est alors de calculer toutes les propriétés du système comme si M était donné, même si "M est en réalité une *variable interne qui doit s'ajuster pour que l'énergie libre soit minimum*" [Diu *et al.*, 1989][p. 455].

L'autocohérence de l'approximation de champ moyen

On sait déjà que si le cristal est paramagnétique parfait dans un champ B_0 , alors $M = N \frac{g\mu_b}{2V} \text{th}\left(\frac{g\mu_b B_0}{2kT}\right)$. On obtient alors, dans un champ magnétique B_{eff} , une valeur de la magnétisation $M = N \frac{g\mu_b}{2V} \text{th}\left(\frac{g\mu_b B_{eff}}{2kT}\right)$. Dans l'approximation du champ moyen, on a vu également que la valeur du champ magnétique dépend de la magnétisation et vaut $\overrightarrow{B_{eff}} = \overrightarrow{B_0} + \lambda \overrightarrow{M}$. En rassemblant ces deux équations, on obtient une équation implicite de M . Cette équation est une condition d'autocohérence de l'approximation en champ moyen. On peut aussi l'écrire :

$$\text{th}^{-1} \left(\frac{2MV}{Ng\mu_b} \right) = \frac{g\mu_b}{2kT} B_0 + \frac{pJV}{2NkTg\mu_b} M$$

À partir de l'équation d'autocohérence écrite sous cette forme, on peut effectuer un développement limité de th^{-1} et ainsi obtenir le comportement local de l'aimantation M , c'est-à-dire éventuellement les lois puissances associées²⁸.

Par exemple, si on se place en champ nul, l'équation d'autocohérence s'écrit $\text{th}^{-1} \left(\frac{2M_0 V}{Ng\mu_b} \right) = \frac{pJV}{2NkTg\mu_b} M_0$. On pose alors $M_\infty = \frac{Ng\mu_b}{2V}$. L'équation d'autocohérence

²⁷Dans d'autres types de modèles, comme les modèles de percolation, l'approximation de champ moyen consiste à remplacer la probabilité effective d'existence de chaque lien par une probabilité moyenne, calculée sur l'ensemble du réseau et donc constante sur tout le réseau.

²⁸On voit bien ici que si les équivalents obtenus pour l'approximation en champ moyen sont des lois puissances c'est uniquement parce que l'on fait un développement limité de th^{-1} , c'est-à-dire de l'équation d'autocohérence. En fait les équivalents exacts pourraient très bien être logarithmiques par exemple, comme Onsager l'a montré pour la capacité calorifique dans le modèle d'Ising en dimension deux.

s'écrit alors $\text{th}^{-1}\left(\frac{M_0}{M_\infty}\right) = \frac{pJ}{4kT} \frac{M_0}{M_\infty}$. On pose $T_c = \frac{pJ}{4k}$. Pour $\frac{M_0}{M_\infty} \ll 1$, c'est-à-dire au voisinage de T_c , on obtient alors $\text{th}^{-1}\left(\frac{M_0}{M_\infty}\right) = \frac{M_0}{M_\infty} + \frac{1}{3} \left(\frac{M_0}{M_\infty}\right)^3 + o\left(\left(\frac{M_0}{M_\infty}\right)^5\right)$. Cette équation s'écrit aussi $\frac{T}{T_c} \left[\frac{M_0}{M_\infty} + \frac{1}{3} \left(\frac{M_0}{M_\infty}\right)^3 + o\left(\left(\frac{M_0}{M_\infty}\right)^5\right) \right] = \frac{M_0}{M_\infty}$. On obtient alors $\frac{M_0}{M_\infty} \sim \sqrt{3 \left(1 - \frac{T}{T_c}\right)}$ soit $M_0(T) \sim \frac{\sqrt{3} M_\infty}{T_c} (T_c - T)^{\frac{1}{2}}$ pour $0 < (T_c - T)/T_c \ll 1$. Finalement, dans l'approximation du champ moyen, on peut donc calculer la valeur des exposants critiques et en particulier de β , comme nous venons de le faire en obtenant $\beta = \frac{1}{2}$. On a donc bien une transition de phase dans l'approximation en champ moyen d'un système en dimension un décrit par l'hamiltonien de Heisenberg.

Comparaison entre l'approximation en champ moyen et la solution exacte en dimension un

Comme nous l'avons déjà vu, la solution exacte du modèle d'Ising en dimension un ne présente pas de transition de phase. Ce résultat diffère de celui obtenu avec une approximation de champ moyen pour l'hamiltonien de Heisenberg. Ceci signifie en fait qu'en dimension un on ne peut pas négliger les fluctuations autour du champ moyen. En fait, en dimension un, ces fluctuations sont déterminantes pour expliquer le comportement global du système, qui ne présente effectivement pas de transition de phase. Sans ces fluctuations, une transition de phase apparaîtrait. *On a donc ici une situation où la solution exacte et la solution dans l'approximation du champ moyen diffèrent qualitativement.* On n'est donc jamais assuré que la solution obtenue dans l'approximation du champ moyen est bien la même que la solution exacte, et ce même si l'équation d'autocohérence de l'approximation est vérifiée. On ne peut même pas affirmer que la solution exacte et la solution approchée sont qualitativement du même type. On trouvera sur la figure 3 un récapitulatif des valeurs obtenues pour différents exposants critiques et pour la température critique à partir des différentes méthodes.

La méthode d'approximation du “groupe de renormalisation”

Nous n'abordons que le principe de la renormalisation, les détails techniques nous semblant inutiles à évoquer dans le cadre qui nous intéresse ici. L'outil est en lui-même très sophistiqué. Il peut par contre être une source d'inspiration heuristique.

Comme dans l'approximation de champ moyen, le principe de cette approximation réside dans la substitution d'un nouveau modèle au modèle exact. Dans le cadre de la renormalisation, le principe consiste à substituer au modèle exact un modèle comportant “moins” d'éléments²⁹ à partir d'un regroupement adapté de

²⁹Le moins est entre guillemets, car si le modèle initial comporte un nombre infini d'éléments, le modèle renormalisé comportera encore un nombre infini d'éléments.

certains éléments du modèle initial. En l'occurrence on va substituer à des blocs d'éléments un seul élément qui sera sémantiquement “plus grand”. Pour cela il faut en même temps remplacer les relations qui existaient dans le modèle exact entre les éléments par une nouvelle relation entre les blocs. Ceci signifie qu'il faut en même temps modifier l'hamiltonien du système. Cette technique était appelée initialement “technique des blocs de Kadanoff” [Droz, 1993].

La méthode du groupe de renormalisation “consiste à calculer la fonction de partition par étapes successives ; à chaque étape, une resommation partielle (par “blocs de spins”) remplace le système initial par un système constitué d'un nombre de “particules” plus faible décrit par un hamiltonien “renormalisé”. Si l'on est capable de caractériser correctement le passage d'une étape à la suivante, on peut en principe obtenir le comportement exact du système”[Diu *et al.*, 1989][p. 475].

On peut réitérer ce processus de transformation, de renormalisation, qui permet d'obtenir un nouveau système plus petit. On obtient alors un nouveau système, etc. Si on suppose que cette méthode est réitérable un nombre infini de fois, le problème consiste à étudier les configurations fixes dans une transformation du système. Comme dans le cas de l'approximation de champ moyen, on obtient une condition d'autocohérence de la méthode. Ici l'autocohérence consistera à vérifier que l'on peut bien réitérer le procédé de renormalisation un nombre infini de fois, et que l'on reste à l'état critique. Il faut noter également qu’“une conséquence intéressante (et inattendue) de la théorie est que les exposants critiques calculés dans l'approximation de champ moyen deviennent exacts³⁰ si le nombre de dimensions de l'espace est supérieur à 4”[Diu *et al.*, 1989][p. 475].

1.1.4. Dynamiques des modèles de type Ising

En ce qui concerne les dynamiques des modèles de type Ising, il faut distinguer deux cas différents qui expliquent pourquoi les études sur les dynamiques qui mènent à l'équilibre n'ont pas été effectuées dès l'abord mais ne proviennent que d'études ultérieures rendues d'ailleurs possibles par l'ordinateur. On appelle paysage d'énergie la courbe qui à toute valeur de l'aimantation M associe la valeur de l'énergie libre. Dans le cas du modèle de Heisenberg dans l'approximation du champ moyen, le paysage d'énergie est une fonction. Cette fonction est paire et comporte deux minima. Partant de n'importe quel point de la courbe, toute dynamique qui permettra de descendre sur la courbe amènera à un minimum³¹ qui sera aussi le minimum

³⁰C'est-à-dire identiques à ceux obtenus par la méthode de renormalisation. Dans tous les cas, ces exposants sont des valeurs approchées car obtenues à partir d'approximations. Ce qui est possible en revanche dans ce cadre est de vérifier les équations d'autocohérence. Mais ceci ne nous assure pas pour autant que l'approximation est bonne. Pour cela il faudrait non pas confronter l'approximation avec elle-même ou avec une autre approximation, mais avec le modèle exact.

³¹Soit $D : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un difféomorphisme, $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est une **fonction de Lyapounov** pour D centrée en x_p ssi

Modèle	Dimension du modèle	Nombre de voisins	$T_c / T_c^{\text{Champ Moyen}}$
Ising réseau carré	d=2	4	0,567 (solution exacte)
Ising réseau triangulaire	d=2	6	0,607 (solution exacte)
Ising réseau cubique simple	d=3	6	0,752
Ising réseau cubique à faces centrées	d=3	12	0,816
Heisenberg réseau cubique simple	d=3	6	0,562

Comparaison entre la valeur de la température critique T_c et la valeur $T_c^{\text{Champ Moyen}}$ obtenue dans l'approximation du champ moyen, pour différents modèles en dimensions deux et trois.

Définition	Champ moyen	Ising (d=2) (solutions exactes)	Ising (d=3)	Heisenberg (d=3)
$M_0 \sim (T_c - T)^\beta$	$\beta = 1/2$	$\beta = 1/8$	$\beta = 0,327$	$\beta = 0,365$
$C_0 \sim (T - T_c)^{-\alpha}$	$\alpha = 0$ (continuité)	$\alpha = 0$ (divergence logarithmique)	$\alpha = 0,107$	$\alpha = -0,115$
$\chi_0 \sim (T - T_c)^{-\gamma}$	$\gamma = 1$	$\gamma = 7/4$	$\gamma = 1,239$	$\gamma = 1,386$

Comparaison des valeurs des principaux exposants critiques obtenues pour différents modèles en dimensions deux et trois.

FIGURE 3 : Comparaison des valeurs des exposants et paramètres critiques obtenues avec des modèles différents (figure reprise de Diu).

absolu de la fonction. Dans ce cas la notion de minimum local se confond avec celle de minimum global.

L'énergie libre est de par sa définition un facteur essentiellement statique. La connaissance de ses *minima* ne nous dit pas comment on fait pour les atteindre en partant d'une valeur quelconque de l'énergie. L'énergie libre ne donne aucune indication sur les dynamiques qui permettent³² d'accéder à ses *minima*. Ceci est particulièrement important lorsque l'énergie libre présente plusieurs *minima* locaux. En effet lorsque l'on cherche à atteindre ces différents *minima* à partir d'une valeur quelconque de l'énergie, des techniques comme une descente de gradient³³ vont amener le système vers un *minimum* local et non vers le *minimum* global recherché³⁴. Seules des techniques plus sophistiquées de type **recuit simulé** (simulated annealing) permettront d'obtenir un *minimum* global³⁵. Dans les études qui s'appuient sur les simulations, l'approche est un peu différente. Des dynamiques sont proposées et programmées par le scientifique et permettent l'étude des *minima* de l'énergie libre. Cependant, d'un point de vue théorique, on doit d'abord statuer sur la possibilité

-
- 1) $F(x) > 0$ pour $x \neq x_p$
 - 2) $F(x_p) = 0$
 - 3) $F \circ D(x) \leq F(x)$ avec égalité ssi $x = x_p$

Sémantiquement, on appelle **fonction de Lyapounov pour une dynamique**, toute fonction (par exemple en dimension deux) telle que la trajectoire de la dynamique (c'est-à-dire la courbe régulière portée par cette fonction) soit monotone.

Nous définissons aussi la notion de **fonction de Lyapounov au sens faible pour une dynamique**.

Soit $D : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un difféomorphisme, $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est une **fonction faiblement Lyapounov** pour D centrée en x_p et de valeur a_p ssi

- 1) $F(x_p) = a_p$
- 2) $F \circ D(x) \leq F(x)$ avec égalité ssi $x = x_p$

Cette définition des fonctions faiblement Lyapounov a l'avantage d'être purement locale, la condition 1) $F(x) > 0$ pour $x \neq x_p$ pour les fonctions de Lyapounov nécessitant au contraire une connaissance globale de F .

³²Le problème peut être formulé en disant qu'étant donnée une fonction F , en l'occurrence la fonction énergie libre, on cherche à savoir pour quelles dynamiques D cette fonction F est Lyapounov ou faiblement Lyapounov pour D .

³³La technique de descente de gradient permet de manière sûre de construire une dynamique D telle que la fonction F soit une fonction faiblement Lyapounov pour D . Par contre cette fonction ne sera pas nécessairement Lyapounov pour D s'il existe un minimum local de F .

³⁴On a bien réussi à construire une dynamique D telle que F soit une fonction faiblement Lyapounov pour D . Cependant il peut arriver que cette fonction soit faiblement Lyapounov pour D en x_p et que ce x_p ne soit pas le "bon" x_p . En effet une fonction faiblement Lyapounov peut avoir plusieurs minima a_p associés à des valeurs x_p . Il est possible que la valeur a_p associée à x_p ne soit pas le plus petit minimum a_i de la fonction F .

³⁵Avec ce type de dynamiques, on obtient une dynamique D telle que F soit une fonction faiblement Lyapounov "par morceaux" pour D . La fonction est faiblement Lyapounov pour la dynamique D jusqu'à l'obtention d'un minimum local a_p de F en x_p . Puis lors du recuit, on passe à une nouvelle valeur de F et on poursuit la descente, c'est-à-dire que la fonction est de nouveau faiblement Lyapounov jusqu'au prochain minimum local a_q en x_q .

d'obtenir ou non une transition de phase, c'est-à-dire sur l'existence et la nature de ces *minima*, avant de s'interroger sur les dynamiques adéquates qui permettront de les atteindre. Ceci explique pourquoi jusqu'à présent les dynamiques des modèles de type Ising ont rarement constitué un objet d'étude prioritaire.

Nous présentons dans l'annexe 1 une liste de quelques dynamiques³⁶ [Prum, 1986] qui ont été plus particulièrement étudiées et qui sont valables sans restriction si l'on travaille sur un réseau fini, ce qui est toujours le cas lorsque l'on fait des simulations. Il faut noter que ces dynamiques peuvent ne dépendre que du site i , de ses paramètres et de ceux de ses voisins, les sites $j \in v(i)$. C'est ce que nous définissons comme une **dynamique locale**. Les dynamiques locales sont des **dynamiques individuelles**, dans le sens où *les dynamiques se définissent pour chaque individu*. Nous appellerons **dynamiques par paquets** des dynamiques qui ont lieu par "paquets", par exemple si on prend tout un sous-ensemble B de sites de R , pas nécessairement voisins, qui sont susceptibles de se modifier en même temps. Cette distinction entre dynamiques individuelles et dynamiques par paquets est particulièrement importante si les dynamiques sont asynchrones, c'est-à-dire si les différents sites ne sont pas tous modifiés en même temps. Nous venons d'opposer des dynamiques individuelles et des dynamiques par paquets. Il faut également distinguer ce que nous appellerons des **dynamiques sitistes** qui pour chaque individu i ne dépendent que des paramètres auxquels le site i peut immédiatement accéder, c'est-à-dire les paramètres du site i et des sites de son voisinage $j \in v(i)$. Nous appellerons **dynamiques ubiquistes** les dynamiques qui ne sont pas sitistes et qui pour chaque individu i dépendent non seulement des paramètres auxquels le site i peut immédiatement accéder, mais également des paramètres auxquels il ne peut pas accéder, c'est-à-dire des paramètres qui nécessitent de parcourir ou de "connaître" un ensemble de sites du système qui sont hors de son voisinage³⁷. Les dynamiques sitistes se distinguent des dynamiques ubiquistes, de même que les dynamiques individuelles se distinguent des dynamiques par paquets. Par contre les dynamiques peuvent être par paquets et ubiquistes, ou individuelles et sitistes, etc. Nous avons défini les **dynamiques locales** comme des dynamiques individuelles et sitistes. On peut distinguer encore des **dynamiques énergétiques**, qui sont des dynamiques ubiquistes particulières qui dépendent de l'énergie libre du système, de la fonction de partition ou de l'hamiltonien³⁸.

³⁶Ces dynamiques nous ont influencé. Elles ont permis de proposer une dynamique appropriée dans le modèle que nous avons construit sur le mouvement révolutionnaire en RDA en 1989. On en trouvera une présentation dans le chapitre 4.

³⁷Le terme d'ubiquiste laisse entendre qu'en fait, dans ces cas-là, la dynamique nécessite l'accès à des variables qui dépendent de tous les sites du système.

³⁸Les dynamiques de Glauber et de Métropolis sont des dynamiques énergétiques. On peut noter pour la dynamique de Métropolis qu'elle n'est pas d'utilisation récente [von Neumann *et al.*, 1959]. Les dynamiques de descente de gradient et de recuit simulé font partie de ces dynamiques énergétiques qui sont donc ubiquistes.

1.1.5. Premiers commentaires sur les modèles de phénomènes critiques

Quelques difficultés et arbitraires dans la constitution des modèles de phénomènes critiques

Lorsque l'on construit des modèles de phénomènes critiques, de nombreux choix doivent être effectués par le modélisateur. Les hypothèses générales d'explication sont souvent les mêmes, mais certains éléments qui peuvent sembler relativement accessoires sont modifiés d'un modèle à l'autre. Nous avons essayé de montrer dans les paragraphes précédents que les modèles de phénomènes critiques sont en fait très sensibles à ce type de variations, malgré les incontestables éléments d'"universalité" qu'ils proposent. Seuls les modèles d'Ising en dimension un et en dimension deux sans champ extérieur et avec des couplages J_{ij} identiques en tout point du réseau ont pu être étudiés de manière exacte.

Des méthodes approchées, comme celle du champ moyen ou du groupe de renormalisation, permettent d'obtenir de nouveaux résultats extrêmement puissants et généraux. Cependant, comme le montre la comparaison entre le résultat exact et l'approximation en champ moyen, en dimension un, un résultat approché peut être qualitativement différent du résultat exact. Il faut noter de plus que la comparaison avec une solution exacte est rarement possible. La puissance et la généralité des résultats approchés obtenus est donc souvent le fait des méthodes utilisées elles-mêmes.

Dans ce cadre, un certain nombre d'éléments constitutifs restent arbitraires, quoique contraints, et on ne peut pas affirmer à l'heure actuelle qu'ils ne jouent pas un rôle déterminant sur le mode de fonctionnement des modèles considérés.

Le premier élément à choisir est l'hamiltonien lui-même, comme mesure de l'état global du système dépendant des interactions entre les sites. L'hamiltonien de Heisenberg a été choisi pour sa "simplicité", mais d'autres formes d'interactions auraient pu être proposées, éventuellement moins simples en apparence, mais d'étude analytique plus aisée. En fait, l'hamiltonien est défini en fonction des états microscopiques de manière relativement libre et en cela on peut le qualifier d'arbitraire. Ce choix est essentiel car l'hamiltonien doit permettre de retrouver un certain nombre de caractéristiques expérimentalement observées au niveau macroscopique. Dans le cas du modèle d'Ising, les contraintes sont donc de type expérimental. Pour les modèles de type Ising, nous n'avons pas à définir, par contre, une mesure *a priori*. Ceci pourrait surprendre les physiciens. Nous souhaitons insister par là sur le fait que la mesure doit être choisie en analysant le phénomène lui-même et non *a priori*. On ne peut donc pas prédéterminer la mesure appliquée à un modèle sans tenir compte du phénomène lui-même, exactement comme en physique. Or lorsque nous parlons de modèles de type Ising, nous ne préjugeons pas du phénomène auquel nous allons appliquer ce modèle.

Outre l'hamiltonien, d'autres éléments sont encore à choisir dans le modèle et en particulier la forme des voisinages. Dans tous les modèles, l'ensemble des sites est discret et fini. Ce choix a été réalisé au départ pour expliquer des propriétés macroscopiques d'un cristal à partir de ses propriétés microscopiques. Ce choix paraît d'autant plus judicieux aujourd'hui qu'il met en jeu des éléments et des paramètres discrets. Or pour ce type de modèles, les difficultés de résolution analytique poussent à utiliser l'ordinateur pour réaliser des simulations et obtenir des solutions approchées. Mais il faut alors spécifier les voisinages. Le choix du modèle d'Ising est celui d'un réseau carré, mais on aurait tout aussi bien pu choisir un réseau hexagonal ou un graphe aléatoire par exemple. Dans l'approximation de champ moyen, ces différents modèles présentent des comportements identiques quand la dimension est suffisamment grande. Mais qu'en est-il pour le modèle exact ? La forme des voisinages est-elle également un facteur accessoire dans ce cas ?

Le modélisateur doit effectuer un autre choix lorsqu'il veut spécifier l'ensemble des valeurs possibles pour chaque spin en chaque site. On suppose pour tous ces modèles que cet ensemble est le même pour tout site du réseau. Dans le cas du modèle d'Ising, cet ensemble contient deux valeurs, qui sont choisies comme étant +1 et -1. Cet ensemble pourrait aussi être un sous-ensemble fini des entiers, par exemple de 1 à n . Cet ensemble pourrait aussi être un espace de Hilbert quelconque, mais de nouveau pour des raisons d'adéquation entre le modèle et la simulation, nous ne considérerons que des ensembles discrets et finis. Il faut bien noter que de grosses différences qualitatives apparaissent entre les différents modèles issus de ces différents choix.

Quelques difficultés dans l'étude des modèles de phénomènes critiques

Étant données les difficultés inhérentes à l'analyse mathématique et théorique de ces modèles, il est extrêmement important et utile d'obtenir des solutions approchées de ces différents modèles grâce à des simulations informatiques. Cependant cette stratégie amène de nouveaux problèmes.

Lorsque l'on fait des simulations ou des expériences, on observe un **ralentissement critique**. Ce phénomène n'est jamais mentionné dans les analyses purement statiques des phénomènes critiques qui abordent essentiellement la question de l'existence d'une transition. Pourtant les expérimentateurs et les simulateurs y sont confrontés [Droz, 1993]. Ce phénomène apparaît lorsque le paramètre de contrôle est déplacé dans la direction du point critique. Plus le paramètre de contrôle se rapproche du point critique et plus l'évolution globale du système vers la transition de phase s'opère lentement. Autant que nous le sachions, ce phénomène qui a été observé n'a pas encore été expliqué théoriquement.

Rôle de l'homogénéité et de la quantité dans les modèles de phénomènes critiques

Les modèles considérés, comme nous l'avons déjà mentionné, sont généralement homogènes pour les voisinages et homogènes pour les dynamiques. Cet aspect des modèles considérés est crucial et délimite le cadre dans lequel l'analyse des phénomènes critiques peut opérer actuellement. “La thermodynamique s'intéresse le plus souvent à des systèmes *homogènes* (au moins localement) à une échelle macroscopique. Les dimensions de tels systèmes peuvent être considérées comme infinies devant les distances entre constituants élémentaires microscopiques, ainsi que le nombre de ceux-ci, et les effets de surface deviennent négligeables en valeur relative. C'est ce qu'on appelle la *limite thermodynamique*” [Balian, 1986][tome 1, IV-17].

Il faut signaler enfin, et ce sera le point sur lequel nous insisterons par la suite, que la criticalité nécessite la quantité. Sans une certaine forme d'infini, on ne peut pas parler formellement de criticalité. L'infini apparaît dans ces modèles sous la forme de ce que l'on appelle la **limite thermodynamique**. “La fonction de partition donne accès à toutes les fonctions thermodynamiques, et permet en particulier de décider de l'existence éventuelle d'une transition de phase. Il faut cependant faire la remarque suivante : pour un système *fini*, la fonction de partition est une somme finie de fonctions analytiques de la température (pour $T \neq 0$). Une transition de phase correspond à une singularité des fonctions thermodynamiques, et avant de se prononcer sur l'existence d'une transition, il faut passer à la limite thermodynamique $N \rightarrow \infty$; en effet, d'après l'argument précédent, il est impossible de voir mathématiquement une transition dans un système fini” [Le Bellac, 1988][p. 31-32].

“Les systèmes thermodynamiques sont tous caractérisés par des interactions persistantes et doivent être décrits par des distributions délocalisées. Pour ce faire, on recourt le plus souvent à la limite thermodynamique. Elle exprime le fait que le nombre N de particules et le volume V augmentent alors que le rapport N/V reste constant. Formellement, on considère les limites $N \rightarrow \infty$, $V \rightarrow \infty$. Bien sûr, il n'y a pas de système, même l'univers, dans lequel le nombre de particules serait infini. Le passage à la limite signifie simplement que les effets de bords, qui sont décrits par des termes en $1/N$ ou en $1/V$, peuvent être négligés. Cette limite thermodynamique joue un rôle central dans toute la physique macroscopique. *Sans elle on ne pourrait pas introduire des états de la matière bien définis, tels les états gazeux, liquide et solide, ni les transitions de phase entre ces états*³⁹. On ne pourrait pas non plus faire la différence entre les régimes proches et loin de l'équilibre dont nous avons rappelé l'importance” [Prigogine, 1996][p. 135-136].

On peut déduire de nombreuses choses de cette citation. Elle permet d'affirmer que les théories actuelles des transitions de phase nécessitent absolument la quantité. *La quantité est un élément constitutif sans lequel aucune théorie actuelle ne pourrait*

³⁹ C'est nous qui soulignons.

expliquer les transitions de phase. La limite thermodynamique est la traduction, dans le cadre des transitions de phase, de l'hypothèse nécessaire de quantité pour rendre compte de ces transitions. L'infini est ici représenté par la quantité de matière, c'est-à-dire le nombre d'éléments. La spatialité est elle-même définie par la façon dont les éléments de matière sont reliés les uns aux autres : on peut dire que cette hypothèse d'infini de la quantité de matière est coextensive d'une infinité en espace. On peut d'ailleurs être plus précis puisque N et V tendent vers l'infini en même temps avec un rapport constant. Le volume V étant une quantité d'espace, c'est bien également l'espace qui est supposé infini à la limite thermodynamique.

Cette citation nous oblige également à aller plus loin. Ce qu'affirme Prigogine en même temps, c'est que finalement la notion de phase et d'état stable de la matière n'est pas absolue et intrinsèque au monde. Cette notion nécessite une abstraction de la part d'un observateur "capable de faire tendre N vers l'infini". Si l'observateur est nécessaire pour définir les phases macroscopiques, ceci signifie que les explications que fournissent les théories actuelles de l'émergence en physique sont insuffisantes. Elles ne permettent pas de rendre compte de la notion de phénomène émergent macroscopique sans l'intervention de la perception par un observateur humain.

Pourtant l'expérience commune nous laisse penser que la notion d'état gazeux, solide ou liquide, ou plus généralement d'état macroscopique et stable de la matière, n'est pas seulement un jeu de l'esprit et une abstraction de nature purement mathématique qui consiste à imaginer que l'"on fait tendre N vers l'infini". En ce qui concerne l'observation des phénomènes, les humains partagent la perception commune de certains états stables de la matière. Cette perception est-elle déjà une abstraction du type décrit par Prigogine ? Cette abstraction est-elle uniquement "une vue de l'esprit" partagée par les humains ou correspond-elle aussi à certains états particuliers de la matière ? Prigogine, en nous affirmant que la notion de phase est purement abstraite, nous invite à croire que les théories actuelles de l'émergence sont insuffisantes car elles ne fournissent aucun élément permettant de penser que les phénomènes macroscopiques émergents le sont de manière intrinsèque. Prigogine nous dit en fait que les théories actuelles de l'émergence ne sont pas intrinsèques.

L'expérience commune peut à la rigueur accepter que la notion de gaz ou de liquide soit extrinsèque. Elle aurait par contre du mal à accepter que le sujet se définisse de façon extrinsèque. Je me définis comme sujet parce que je me maintiens et que je persiste comme moi-même sujet⁴⁰. Je peux accepter d'être décrit comme sujet et comme phénomène émergent, mais cette émergence doit être intrinsèque et ne pas nécessiter la présence d'un autre observateur. Je me tolère comme observateur qui observe les autres et je définis les autres par cette observation. Mais l'observateur qui m'observe moi-même et me définit, c'est moi, et moi seul, comme sujet. Les théories actuelles de l'émergence sont incapables de rendre compte de ce type de phénomènes car l'émergence dont elles parlent n'est jamais intrinsèque.

⁴⁰On pourrait voir à ce propos le rapport de Patrice Prez [Prez, 1996][chapitre 2].

Pour nous résumer, on peut répéter ici le jugement péremptoire de Jean-Louis Rolles : “La complexité est le dernier avatar des théories de la quantité. La complexité c'est la thermodynamique à l'époque de l'ordinateur” [Rolles, 1993][p. 155].

1.2. Les automates cellulaires, modèles de dynamiques et exemples de machines

1.2.1. Les origines des travaux sur les automates cellulaires

Von Neumann fait partie de ceux qui ont conçu les machines séquentielles [Goldstine, 1993]. Dans ses travaux, de 1948 à 1957 [von Neumann, 1951] [von Neumann, 1956][von Neumann, 1966]⁴¹[von Neumann, 1992], il montre un intérêt égal pour des automates qui constitueraient un exemple paradigmique de machines massivement parallèles. Dans ce travail, sous l'impulsion d'Ulam, il a plus particulièrement étudié un automate cellulaire à 29 états [von Neumann, 1966]. *Il considère ce système comme une structure idéale pour modéliser le processus biologique de l'auto-reproduction.* Dès l'origine, l'automate formel conçu par von Neumann est ambivalent. Il est à la fois une copie et l'élément que l'on doit copier. L'automate est considéré comme un modèle de certaines propriétés biologiques. Mais il est aussi conçu comme l'exemple paradigmique de ce que peut être une machine parallèle. La copie du biologique est déjà l'exemple que l'on s'évertuera à copier.

La question générale que pose alors von Neumann est de savoir quelle organisation logique est suffisante pour qu'un automate se contrôle lui-même et soit capable de se reproduire. Von Neumann a défini un automate cellulaire en deux dimensions, avec 29 états pour chaque site et un voisinage comportant cinq éléments, les voisins directs du réseau carré et soi-même. Il a réussi à montrer que cet automate est capable de **computation universelle**. Ainsi cet automate possède les mêmes capacités logiques qu'une machine de Turing. Cet automate est alors capable de simuler une machine de Turing.

L'automate de von Neumann est également capable de **construction universelle**. En effet, si on lui donne la description d'un autre automate, il est capable de construire ce deuxième automate dans une région vide de l'espace cellulaire. L'auto-reproduction apparaît alors comme un cas particulier, pour lequel l'automate décrit est le constructeur universel lui-même. Initialement l'automate de von Neumann est donc un modèle formel particulier susceptible de rendre compte de deux capacités logiques observées, la computation universelle et la construction universelle. *Ce modèle est abstrait et éloigné des réalités biologiques, mais pour von Neumann sa vertu principale est de rendre compte d'une propriété minimale du vivant, la capacité*

⁴¹Ces conférences ont été publiées par Burks en 1966, c'est-à-dire neuf ans après la mort de von Neumann en 1957.

de reproduction. En ce sens, ce système formel est bien un modèle. De plus il reste relativement simple eu égard à l'enchevêtrement inextricable que semble être la reproduction des êtres vivants.

Nous venons d'insister sur le fait qu'initialement l'automate cellulaire de von Neumann était considéré comme un modèle. Il faut également insister sur le fait qu'il est très différent, dans son état d'esprit, des modèles de type Ising. En effet, dès l'abord et *dans leur conception même, les automates sont conçus pour être étudiés dans leur comportement dynamique.* La dynamique dans l'automate de von Neumann joue un rôle fondamental. Elle est conçue comme étant la propriété même dont le modèle doit rendre compte. Nous avons vu au contraire pour les modèles de type Ising que la dynamique ne venait qu'en second plan.

1.2.2. Les automates cellulaires dans le cadre de Farmer

Un automate cellulaire est défini comme un objet mathématique, censé idéaliser un système physique dans lequel l'espace et le temps sont considérés comme des paramètres discrets. Chaque point de l'espace est constitué de cellules que nous appellerons sites pour conserver le vocabulaire de Farmer. À chaque point de l'espace on associe un symbole, généralement un nombre que l'on prend dans un ensemble fini. Les sites sont reliés entre eux par des liens. Pour des raisons d'ordre pratique liées à sa programmation, un automate cellulaire est généralement pris comme un réseau homogène pour les voisinages⁴², et de dimension N entière. On retrouve alors la description de Farmer avec comme graphe un réseau carré. Généralement, les études se font en dimension deux, ce qui correspond au plan. Cependant, une bonne partie des études réalisées par Wolfram au début des années 80 ont été effectuées en dimension un. Contrairement aux graphes de Farmer qui peuvent avoir leur propre dynamique, la structure du réseau de l'automate reste fixe au cours du temps.

Quand il est étudié d'un point de vue théorique, ce réseau est supposé d'extension infinie dans l'espace, c'est-à-dire comportant un nombre infini de sites. Lorsque l'on cherche à programmer l'automate, on se contente généralement d'un réseau fini périodique, avec une structure de "tore" qui permet de ne prendre en considération qu'un nombre fini de sites. Chaque site peut prendre une valeur discrète, et finie, généralement dans un sous-ensemble fini des entiers : cette valeur est appelée l'état du site. Lorsque cette valeur ne peut être que 0 ou 1, on parle d'automates booléens. L'état de l'automate cellulaire est défini comme la liste des états de chacun des sites du réseau. Ainsi la connaissance de tous les états des sites à un instant donné permet de connaître, et constitue, **l'état de l'automate** tout entier.

Un automate possède des règles d'évolution de l'état de chaque site. Ces règles s'appliquent à chaque pas de temps. Ce temps est lui-même considéré comme un

⁴²Un réseau est homogène pour les voisinages, comme nous l'avons défini, si le graphe est tel que les voisinages sont identiques en tout site.

paramètre discret et uniforme. Les règles d'évolution de l'état d'un site ne dépendent que de l'état des sites voisins. Il est donc nécessaire de fixer un voisinage *a priori* et pour chaque site. De plus les règles d'évolution des états des sites sont les mêmes pour tous les sites du réseau. C'est ce que nous avons défini comme un graphe homogène pour les dynamiques. On parle ainsi des règles de l'automate, sans préciser en quel site de l'automate on se trouve. Ces règles sont en général déterministes ; elles peuvent être fournies par une table. On parle d'automates stochastiques, lorsque ces règles sont stochastiques.

Il est important de noter également que tous les états sont remis à jour simultanément grâce aux règles d'évolution, ce qui évite tout problème de choix séquentiel dans l'ordre des sites à remettre à jour. On parle alors de **dynamique synchrone**. Certains automates peuvent être asynchrones, lorsque les remises à jour ne sont pas simultanées. Il faut être très vigilant lorsque l'on fait le choix entre automates synchrones ou asynchrones. On peut parfois montrer théoriquement que les comportements sont similaires. Cependant, dans d'autres cas, les études de simulations mettent en évidence des différences de comportement très importantes [Bersini et Detours, 1994][Lumer et Nicolis, 1994]. Le choix séquentiel des sites à réactualiser peut changer radicalement le comportement de l'automate. On ne peut donc pas traiter de manière asynchrone un problème qui se modélise naturellement en termes synchrones sans modifier profondément ce que l'on étudie effectivement. Lorsque l'on construit un automate, il faut donc le choisir synchrone ou asynchrone. De plus, il existe autant de façons *a priori* différentes d'être asynchrone pour un automate que de combinaisons du choix séquentiel des sites à réactualiser, c'est-à-dire un nombre infini si l'automate est infini.

Pour résumer, on peut dire que les automates cellulaires sont un cas particulier du formalisme de Farmer, pour lequel le graphe est homogène pour les voisinages. Les états des sites sont généralement discrets et finis. On trouve un seul type de dynamique, celle des règles de transition. Elles sont identiques en tout point du réseau, et ne varient pas au cours du temps, c'est-à-dire que le graphe est homogène pour la dynamique des règles de transitions. Les automates sont cependant un cas très particulier du cadre de Farmer, car on n'y trouve qu'une seule dynamique parmi les trois possibles. En particulier, pour les automates, le graphe reste complètement fixe au cours du temps.

1.2.3. Problématiques pour les automates cellulaires

Progressivement, une discipline à part entière s'est développée, dont l'objectif est d'étudier le comportement des automates cellulaires. On trouve des applications très nombreuses et diverses qui motivent ces travaux. Ainsi l'automate de von Neumann, qui était d'abord un modèle particulier, va devenir un exemple de machine universelle. L'automate de von Neumann était au départ l'image du phénomène

d'autoreproduction. *Il va devenir désormais l'exemple que l'on copie. L'automate, modèle et image du réel, se transforme en exemple de machine. Il est copié à son tour et l'on construit des modèles de cet automate.*

On peut globalement résumer les questions sur les automates en deux types de problèmes : le problème direct et le problème inverse. On parle de **problème direct** lorsque l'on se donne une règle de l'automate et que l'on cherche à en déduire ses propriétés. Le problème inverse consiste à trouver un ensemble de règles possédant une propriété donnée. On peut pratiquement lire tous les articles sur les automates cellulaires avec cette ligne directrice qui permet de considérer que l'on cherche à résoudre un problème direct ou un problème inverse.

Pour étudier le problème direct, il faut s'être donné des règles dont on sait *a priori* qu'elles possèdent des propriétés intéressantes. Le problème direct se retrouve par exemple lorsque l'on cherche à décomposer des règles comme combinaisons d'opérateurs. Il apparaît également dans les études de tables de règles. On peut alors considérer des règles réversibles, invariantes, critiques, ou qui permettent des propriétés de généralisation.

Le problème inverse apparaît lorsque l'on cherche quelles règles permettent d'obtenir un comportement souhaité. On trouve un exemple dans le travail de von Neumann, qui cherchait quelles règles il fallait fournir pour obtenir un automate capable d'auto-reproduction. Le "jeu de la vie" de Conway⁴³ se situe également dans cette optique, puisqu'il cherche à trouver quelles règles permettent d'obtenir des comportements considérés comme complexes. Pour résoudre ce problème, on est souvent obligé d'attribuer à un paramètre un certain nombre de règles, comme a pu le faire Langton [Langton, 1990]⁴⁴. On peut alors parcourir cet ensemble de règles pour en analyser une large panoplie et en extraire celles qui semblent fournir des comportements intéressants. La classification proposée par Gutowitz [Gutowitz, 1990] en termes de chaînes de Markov (annexe 3) se situe également dans cette optique, puisqu'elle cherche à obtenir une classification des règles en fonction de leur comportement. Les recherches sur l'apprentissage au moyen d'automates se situent également dans cette optique. D'autres automates sont aussi utilisés pour étudier la croissance dendritique d'une figure, ce qui n'est pas sans rappeler les résultats auxquels nous étions parvenu dans notre étude sur les figures obtenues au cours d'un apprentissage utilisant le Q-learning [Saurel, 1992]. De plus Toffoli et Margolus ont fait remarquer qu'il est souvent plus facile de créer des règles possédant un invariant donné, c'est-à-dire sous forme d'un problème inverse, plutôt que de chercher les invariants d'une règle donnée, sous la forme d'un problème direct, ces invariants n'existant peut-être pas toujours.

Le problème inverse permet de bien comprendre l'état d'esprit et la motivation

⁴³On trouvera quelques détails et des références sur ce sujet dans la section suivante.

⁴⁴Des explications et des références sur ce travail de Langton se trouvent dans la section suivante.

des différentes questions sur les automates. C'est dans ce sens qu'il convient de remplacer l'historique suivant. En 1968, E.F. Codd [Codd, 1968] a proposé un automate cellulaire à 8 états, capable de computation et de construction universelle. D'autres simplifications ont été introduites par la suite, par Smith en 1971, Conway en 1982, et Fredkin et Toffoli en 1982. Cependant, l'automate de von Neumann n'a jamais été simulé actuellement. Lorsque l'on cherche à simuler cet automate, on constate qu'il possède 295 règles locales différentes possibles, ce qui correspond à près de 20 millions de configurations possibles. Il est donc délicat de mettre en mémoire dans une table l'ensemble de toutes ces configurations, pour ensuite aller les chercher lorsque l'on doit les utiliser. Ce travail de von Neumann sur des machines parallèles autoreproductrices n'a pas été directement poursuivi. Avec l'augmentation de la puissance des machines et leur plus large diffusion, de nombreux scientifiques ont exploré d'autres automates cellulaires, relativement proches de celui de von Neumann. *Ces travaux ont en partie été guidés par une recherche de simplification et par la description d'un automate de taille minimale capable de computation et de construction universelle.*

La motivation des travaux sur les automates porte donc sur la générnicité. On va chercher la structure de taille minimale capable de réaliser une tâche particulière. La complexité d'un ensemble de tâches sera donc liée à la taille minimale de l'automate nécessaire pour le réaliser. La quantité prend ici une nouvelle forme. Il ne s'agit plus d'une quantité spatiale, mais de la constitution d'une potentialité fonctionnelle résumée. Cette fonction se développe dans le cadre des automates sous une forme temporelle et non spatiale, c'est-à-dire au cours de la dynamique, contrairement à ce que l'on observe dans les modèles de phénomènes critiques. L'infini en jeu ici est temporelle. On cherche à la résumer grâce aux différentes formes que présente la **redondance**.

Dans le cadre des modèles de phénomènes critiques, l'essentiel des propriétés était concentré à la limite thermodynamique. Dans le cadre des automates, l'objectif est de toujours résumer et concentrer davantage les différents comportements dynamiques. Le concept unificateur est alors celui de redondance.

1.2.4. Applications des automates

Les automates cellulaires ont été utilisés pour étudier des systèmes chimiques non-linéaires comportant des réactions couplées avec de la diffusion spatiale [Greenberg *et al.*, 1978]. Un modèle controversé d'évolution des galaxies spirales utilise ces idées [Gerola et Seiden, 1978][Schewe, 1981]. On peut décrire la croissance des cristaux comportant des dendrites [Langer, 1980] comme l'agrégation de paquets en nombre discret, avec des règles locales adéquates. La structure spatiale des fluides turbulents peut être modélisée avec des automates cellulaires, comme nous le verrons dans notre partie concernant les gaz sur réseau. De nombreux

systèmes biologiques ont été modélisés avec des automates cellulaires [Rosen, 1981]. Le développement de structures et de motifs dans la croissance des organismes apparaît souvent comme un phénomène lié à des règles locales extrêmement simples [Stevens, 1974]. Ces quelques références sont évocatrices de la diversité des modèles qui prennent le formalisme des automates cellulaires. On peut noter cependant que ces références ne sont pas toutes récentes. En effet, des modèles actuels de phénomènes identiques tiendraient compte de la dynamique du graphe et devraient être considérés dans le cadre connexioniste de Farmer et non dans celui des automates cellulaires *stricto sensu*.

1.3. Les hypothèses justifiant l'utilisation des modèles de la complexité/quantité pour rendre compte du fonctionnement cérébral

1.3.1. Précisions sur les relations entre la complexité/quantité et le chaos

Nous savons depuis les études sur les phénomènes chaotiques que certains systèmes comportant un nombre faible d'éléments, par exemple trois, peuvent avoir un comportement très irrégulier [Bergé *et al.*, 1988][Gleick, 1989]. Ces comportements sont par exemple très fortement dépendants des conditions initiales. On conserve la possibilité de décrire le système en termes d'équations différentielles et on peut ainsi garder la notion de déterminisme ancienne [Enriques, 1941][deb, 1990]. On observe cependant, pour ces phénomènes, une divergence exponentielle des trajectoires. Ainsi, deux positions initiales très proches et éventuellement indistinguables, obéissant au même système d'équations différentielles, vont progressivement, mais exponentiellement diverger l'une de l'autre. Comme nous l'avons dit, cette forte dépendance vis-à-vis des conditions initiales est possible, même avec un nombre très faible d'éléments. Certains considèrent que la notion ancienne de trajectoire n'est plus pertinente dans ce contexte. Ils proposent de la remplacer par une notion de probabilité de présence [Prigogine, 1996].

Dans notre discussion, nous distinguons volontairement les phénomènes chaotiques de la notion de complexité. Nous avons mentionné ces phénomènes uniquement parce qu'ils sont souvent associés, sans doute de manière abusive, à la notion de complexité⁴⁵. En fait l'association entre chaos et complexité a lieu à partir d'une complexité comprise comme "ce que je ne sais pas encore expliquer" et non à partir du sens technique de la complexité/quantité. Selon nous il faut bien distinguer chaos et complexité/quantité. **Le chaos** au sens technique doit être entendu comme *la forte dépendance vis-à-vis des conditions initiales*. Cette condition peut avoir lieu tout aussi bien pour de petits systèmes que pour de grands systèmes qui relèvent de

⁴⁵Voir par exemple le colloque de Blois, *Chaos et Complexité*.

la complexité/quantité. Les deux concepts peuvent se retrouver ensemble dans un unique phénomène. Mais il est rare que les deux concepts soient mis sérieusement en évidence pour le même phénomène. On observe cependant que les phénomènes critiques qui relèvent bien de la complexité donnent lieu à des corrélations à grande distance ou sans distance caractéristique. Ceci correspond généralement à des cas de forte dépendance aux perturbations et aux fluctuations. On peut alors observer un phénomène chaotique dans le cadre de la complexité/quantité.

1.3.2. Cinq caractéristiques communes aux modèles de la complexité/quantité et aux conceptions des neurobiologistes sur le fonctionnement cérébral

On peut présenter cinq caractéristiques généralement partagées par les descriptions du fonctionnement cérébral et qui sont communes aux modèles que nous avons décrits comme relevant de la complexité/quantité, c'est-à-dire aux modèles de type Ising et aux automates. Parmi ces caractéristiques communes, il faut insister sur les deux premières, qui justifient plus particulièrement une conception émergentielle du fonctionnement cérébral.

La première caractéristique est le nombre très important d'éléments.

Cette caractéristique est au cœur même des considérations et des méthodes issues de la mécanique statistique et cherchant à expliquer la thermodynamique. “On pourrait conclure naïvement [...] que seule importe la physique microscopique, puisque sa compréhension entraînerait aussitôt celle de ses manifestations macroscopiques. [...] Mais une telle conclusion serait fondamentalement incorrecte. Il existe en effet de nombreux phénomènes macroscopiques qui échappent à une explication directe, même seulement qualitative, à partir des lois microscopiques. [...] Le passage du microscopique au macroscopique se présente comme hautement non trivial, et comme devant nécessiter une étude approfondie et particulière. Ce problème étant posé, comment peut-on espérer le résoudre ? La méthode qui va être utilisée tire parti du fait que ce passage du microscopique au macroscopique met en jeu des nombres extrêmement grands” [Diu *et al.*, 1989][p. 6].

Cette conception de la complexité comme quantité n'est pas propre aux sciences physiques et on la retrouve en particulier dans les conceptions actuelles de l'organisation cérébrale. Dans une conférence intitulée *La chimie des communications cérébrales*, prononcée le 27 janvier 1994 au Collège de France, Jean-Pierre Changeux s'exprime à propos du cerveau en ces termes : “Notre cerveau se compose de neurones et le nombre de cellules n'est pas encore parfaitement bien établi. Ça peut paraître surprenant à une époque où l'anatomie est quand même très perfectionnée. Mais le nombre absolu de ces neurones est extrêmement élevé puisqu'on en compte entre 10 et 20 milliards dans le cortex cérébral de l'homme et peut-être de

l'ordre de 100 milliards dans l'ensemble de notre encéphale. Donc vous voyez que ce nombre est un nombre réellement astronomique qu'il faut garder en mémoire car il y a toujours des débats en ce qui concerne la mise en relation entre l'organisation du cerveau et ses fonctions et on oublie en général que cette organisation se constitue à partir d'un nombre de cellules qui est extrêmement élevé et que si on associe ces cellules les unes aux autres, on arrive à un nombre de combinaisons qui est plus qu'astronomique puisque pour certains auteurs il serait égal au nombre de particules existant dans l'univers (d'après Gerald Edelman). En tout état de cause nous avons affaire à une organisation qui est d'emblée d'*une extraordinaire complexité*⁴⁶. Les propos de J.-P. Changeux correspondent tout à fait à la conception actuelle du fonctionnement du cerveau. Lorsque l'on parle du cerveau et de son organisation, on leur associe immédiatement le terme de complexité. On justifie alors ce terme, comme J.-P. Changeux, par le nombre des éléments, en l'occurrence des cellules, qui sont en jeu. De plus, comme on parle d'organisation, la complexité n'est pas liée uniquement aux éléments qui sont en jeu, mais également aux relations qu'ils entretiennent les uns avec les autres. Ainsi la complexité dont parle Changeux est liée non seulement au nombre d'éléments en jeu, mais surtout au nombre de combinaisons possibles entre ces éléments. La complexité dont il est ici question se résume en deux nombres, le nombre d'éléments en jeu et le nombre de connexions possibles entre ces éléments. La complexité prend la forme d'une quantité, c'est-à-dire d'un nombre. Le terme de complexité/quantité est donc parfaitement justifié.

Cette conception n'est pas propre aux biochimistes et on la retrouve par exemple chez Searle. "Selon les estimations actuelles, le cerveau humain comprendrait plus de 100 milliards de neurones, le nombre de liaisons synaptiques de chaque neurone allant de quelques centaines à plusieurs dizaines de milliers. Cette structure biologique d'une extrême complexité est tout entière regroupée dans un volume plus petit que celui d'un ballon de football" [Searle, 1996][p. 62]. Searle, lui aussi, décrit ici la complexité uniquement dans les termes de la quantité.

Le propre de cette conception est de considérer que la complexité/quantité permet les processus émergents. Selon la théorie de la complexité/quantité, les processus émergents apparaissent à la "limite thermodynamique", c'est-à-dire lorsque le nombre d'éléments en interaction est très important voire infini.

La deuxième caractéristique commune est une hypothèse générale d'homogénéité.

Nous avons mis en évidence cette hypothèse générale d'homogénéité dans le cas des modèles de la complexité/quantité. Elle prenait alors à la fois la forme de l'homogénéité des voisinages, et celle de l'homogénéité des dynamiques, que ce soit dans le cadre des modèles de type Ising ou des automates. En ce qui concerne les observations sur le fonctionnement cérébral, l'homogénéité fondamentale observée

⁴⁶ C'est nous qui soulignons. Mais nous reproduisons, en fait, l'insistance de l'orateur.

est celle du néocortex⁴⁷. Les éléments cellulaires qui constituent le cortex sont en proportion grossièrement identique dans tout le néocortex. On sait de plus que la structuration en six couches corticales est la même pour l'ensemble du néocortex.

La troisième caractéristique commune porte sur les aspects temporels et en particulier la redondance et la générnicité.

Nous avons trouvé ces caractéristiques essentiellement pour les automates parce que, comme nous l'avons dit, ils sont conçus dès l'abord pour étudier des comportements dynamiques. Cette caractéristique se retrouve également dans notre conception du fonctionnement cérébral. On parle de redondance dans les signaux perçus, et on considère par exemple que le propre d'un système perceptif est de réduire la redondance pour obtenir les éléments significatifs, c'est-à-dire pertinents, du signal. On considère également que le cerveau est capable de résoudre un grand nombre de tâches. En ce sens, comme pour les automates, on insiste sur la générnicité du cerveau, entité "spatiale" qui constitue un résumé "spatial" capable de générer un comportement temporel.

La quatrième caractéristique commune est l'hypothèse de localité.

Selon cette hypothèse, chaque élément du système obéit à une dynamique locale. Celle-ci peut s'exprimer en termes d'équations différentielles ou de règles de comportements. Ces dynamiques ne dépendent généralement que de l'état du système à proximité et c'est dans ce sens qu'elles sont locales. Cela n'empêche pas que de proche en proche tous les éléments peuvent devenir partie prenante de la dynamique des autres éléments.

Cette caractéristique commune reste cependant trop générale. Il faut la préciser dans chaque modèle étudié en expliquant comment sont définis les voisinages. Dans le cas des cellules nerveuses, par exemple, les voisinages peuvent soit être définis par la proximité dans l'espace, soit par l'existence d'une liaison synaptique. Ainsi plusieurs "localités" peuvent jouer en même temps.

L'hypothèse de localité est en même temps préalable aux hypothèses générales d'homogénéité.

La cinquième caractéristique commune affirme que le comportement global du système dépend du comportement local de chaque élément.

Dans cette optique le comportement global du système est une mesure définie à partir du comportement de chacune des individualités distinguées. La mesure est généralement imposée de l'extérieur, de manière relativement arbitraire et du seul point de vue de l'observateur.

⁴⁷On pourra consulter [Mumford, 1991] pour une présentation de cet aspect par un modélisateur.

Selon nous les caractéristiques communes portant sur la quantité et l'homogénéité prennent généralement sur toutes les autres. Ces caractéristiques peuvent nécessiter implicitement une autre caractéristique, comme la localité vis-à-vis de l'homogénéité. Globalement, on peut dire que ces caractéristiques sont rarement explicitées, même si elles fondent implicitement les modèles du fonctionnement cérébral.

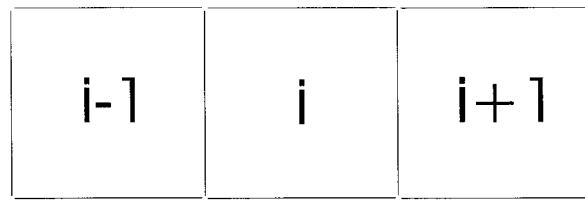
On peut dégager deux propriétés générales des modèles générés à partir des caractéristiques précédentes. Tout d'abord, il est le plus souvent impossible de résoudre analytiquement les équations obtenues pour le comportement global du système. Ce point n'est évidemment pas une nécessité, mais il recouvre la plupart des situations effectivement étudiées actuellement. Ceci justifie alors l'utilisation massive de l'ordinateur pour obtenir des solutions approchées. Il faut mentionner de plus la fréquente instabilité des solutions obtenues. Cette instabilité peut être liée aux conditions initiales et on retrouve alors éventuellement certains points évoqués pour les systèmes chaotiques. Ainsi la notion de trajectoire chaotique est une propriété des systèmes complexes, souvent observée, même si elle n'est pas nécessaire.

2. QUESTIONS ET OUTILS POUR ABORDER LA COMPLEXITÉ/QUANTITÉ

Depuis une quinzaine d'années, les études concernant les phénomènes et les modèles considérés comme liés à la complexité/quantité se sont multipliées. Elles sont indissociables d'une utilisation nouvelle, systématique et massive des possibilités offertes par les ordinateurs. À cet effet, de nouveaux outils ont été élaborés. On peut les considérer comme mathématiques, dans la mesure où ils constituent des outils symboliques. Ils ne sont pourtant que rarement complètement formalisés¹. On peut penser que ceci est le fait de tous les débuts et des tâtonnements qui leur sont associés. On peut défendre au contraire que ces outils nouveaux ne pourront pas être parfaitement formalisés parce qu'ils ont été construits dans une interaction étroite avec l'utilisation de l'ordinateur. Cette situation justifie de les présenter non pas en insistant sur les formalismes, *mais rapidement, afin de mettre en évidence leur diversité*. Leur caractéristique principale est de faire appel à tous les types de formalismes abstraits déjà élaborés. On peut donner de ces outils soit une approche continue soit une approche discrète et finie, ces deux présentations étant généralement parallèles. Nous choisissons ici de privilégier l'approche discrète.

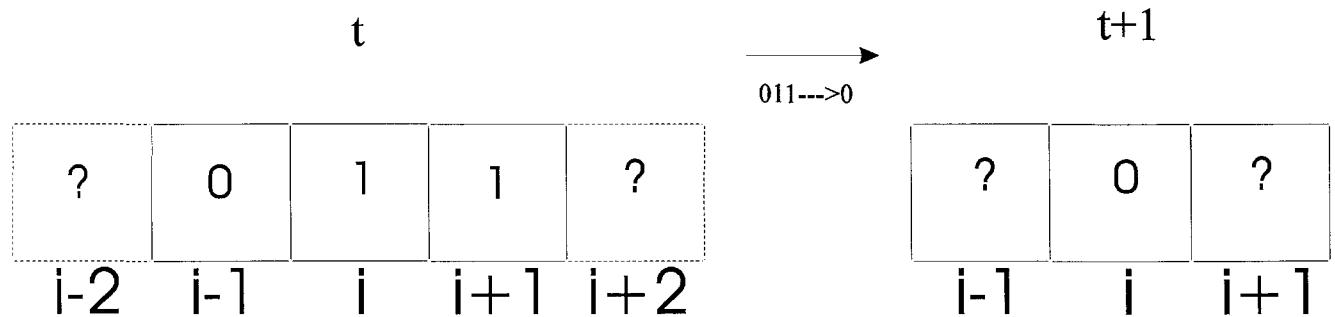
En fait nous insisterons essentiellement *sur les outils élaborés ou utilisés pour étudier les automates cellulaires*. Notre regard privilégiera les descriptions formalisées discrètes et finies. Ceci signifie implicitement que nous privilégions l'utilisation de l'ordinateur et les outils de calcul sur les descriptions analytiques continues. On peut s'interroger sur les raisons qui expliquent le rôle de l'ordinateur. Comme nous l'avons écrit dans la section précédente, le rôle accordé aux dynamiques permet de distinguer les automates cellulaires des modèles de type Ising. Dans le cadre des automates cellulaires, il est généralement nécessaire d'appliquer les règles et les dynamiques en jeu pour pouvoir les étudier. Au contraire, dans les modèles de type Ising, la première question posée est de savoir si différentes phases peuvent apparaître et non pas quelles dynamiques permettent de les faire apparaître. Aussi nous avons choisi d'étudier les outils développés dans le cadre des automates cellulaires, parce qu'ils nous permettront de savoir comment on peut envisager à l'heure actuelle d'étudier le comportement d'une dynamique, à partir de son fonctionnement.

¹Comme c'est d'ailleurs généralement le cas en mathématiques.



1	1	1
1	1	0
1	0	1
1	0	0
0	1	1
0	1	0
0	0	1
0	0	0

Les valeurs possibles du voisinage du site i qui se trouve au milieu de ses deux voisins. Au-dessous, on trouve les valeurs possibles du voisinage, pour un automate à deux états par site, 0 et 1. L'ensemble des valeurs possibles du voisinage est donné par l'ensemble des valeurs possibles du triplet $(i-1, i, i+1)$. Les valeurs possibles du voisinage sont données dans l'ordre utilisé par Wolfram.



Exemple du passage de l'instant t à l'instant $t+1$ par application de la partie de la règle $011 \rightarrow 0$ puisque le site i a pour état 1, que son voisin à gauche ($i-1$) a pour état 0 et que son voisin à droite ($i+1$) a pour état 1. On ne connaît pas les états à l'instant $t+1$ du voisin de gauche ($i-1$) et du voisin de droite ($i+1$) car on ne connaît pas l'état de leurs propres voisins respectivement ($i-2$) et ($i+2$) à l'instant t .

FIGURE 4 : Quelques explications sur la numérotation des valeurs des voisinages utilisée par Wolfram.

Chapitre 3 -- Figure 4

2.1. L'approche de Wolfram des automates cellulaires

Dans ce premier paragraphe consacré aux études de Wolfram sur les automates cellulaires, nous établissons un premier panorama pour fixer le cadre dans lequel il a travaillé² [Wolfram, 1983]. Il s'agit d'études des automates en une dimension, avec comme états possibles 0 ou 1 pour chaque site. De plus, comme il est impossible de simuler des automates de taille infinie, Wolfram choisit d'étudier des automates finis et périodiques. Cette hypothèse de périodicité se traduit par une construction en tore de l'automate, qui nous rappelle celle du modèle d'Ising en dimension une. Le voisinage d'un site à l'extrême gauche est constitué des sites à l'extrême droite de l'automate.

Les propriétés théoriques des automates cellulaires sont valables avec un nombre infini d'éléments. Pour passer de la théorie à la mise en œuvre pratique, on ajoute cette hypothèse de périodicité, qui restreint l'espace des automates effectivement étudiés, mais permet de concilier résultats théoriques et expériences numériques. On retrouve ici une situation rencontrée dans la théorie de traitement du signal pour la définition de la transformation de Fourier. Dans ce cadre, on souhaite travailler sur des fonctions dont l'ensemble de définition est un ensemble compact pour des raisons de convergence des intégrales définies à partir de ces fonctions. Il est en même temps souhaitable d'avoir des fonctions définies sur \mathbb{R} tout entier, parce que c'est l'ensemble de définition *a priori* des fonctions associées aux signaux rencontrés. L'hypothèse de périodicité est alors nécessaire et permet de concilier ces deux contraintes, en considérant en fait des fonctions qui sont définies sur un tore [Marco, 1996].

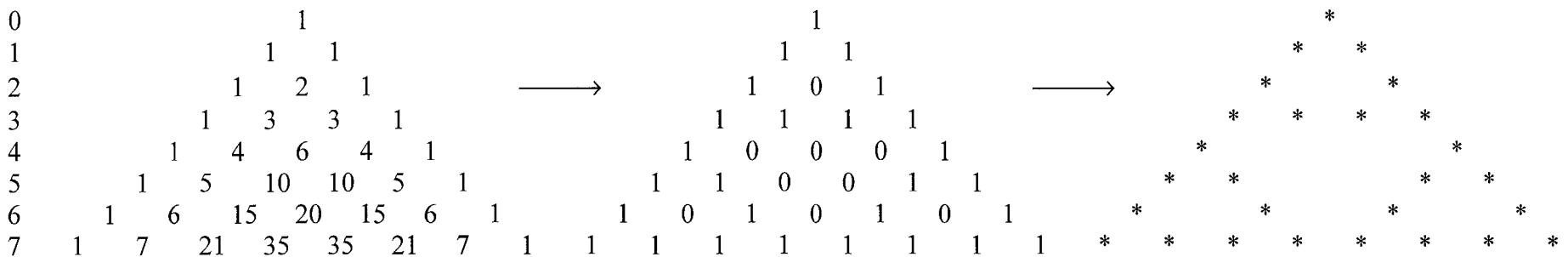
2.1.1. Introduction aux automates

Le premier travail consiste à définir l'ensemble des règles que l'on souhaite étudier. Wolfram réduit son étude aux voisinages comportant trois éléments, le voisin de gauche, celui de droite, **ainsi que soi-même**. On a alors 2^3 , soit 8 valeurs³ possibles pour les voisinages (figure 4). On obtient alors 2^8 règles⁴ possibles soit 256. Chaque règle est explicitement codée dans un nombre binaire à huit bits. Chaque bit correspond à un voisinage possible. Par exemple, 90 s'écrit $2^6 + 2^4 + 2^3 + 2^1$, soit 01011010 en binaire. Les voisinages sont toujours codés dans le même ordre, respectivement 111, 110, 101, 100, 011, 010, 001 et 000 (figure 4). Le premier bit

²Les articles de Wolfram publiés entre 1983 et 1988 ont été rassemblés dans [Wolfram, 1994].

³On obtient 2^3 valeurs possibles, parce que l'on va d'un ensemble à trois éléments (les trois voisins) vers un ensemble à deux éléments (les deux valeurs 0 et 1 en chaque site).

⁴Comme précédemment, on obtient 2^8 valeurs possibles, parce que l'on va d'un ensemble de 8 éléments (les 8 voisinages possibles) vers un ensemble à deux éléments (les valeurs 0 et 1 du site étudié après application de la règle).



La première colonne correspond aux «instants» successifs, c'est-à-dire à la valeur de n . On reconnaît le premier triangle comme étant le triangle de Pascal, qui nous donne les coefficients du binôme, c'est-à-dire que le $k^{\text{ième}}$ coefficient à la ligne n est le facteur de x^k dans le développement de $(1+x)^n$.

Le passage du premier triangle au deuxième triangle est obtenu par calcul de la valeur correspondante dans le triangle précédent modulo 2. On a donc le $k^{\text{ième}}$ coefficient à la ligne n qui est le facteur de x^k dans le développement de $(1+x)^n$.

Le passage du deuxième triangle au troisième triangle est obtenu par le remplacement de la valeur correspondante dans le triangle précédent par une étoile si cette valeur est 1 et par un espace vide si la valeur est 0.

Le dernier triangle correspond également à une représentation du déroulement de la règle 90 de Wolfram en partant d'un unique site égal à 1, à l'instant n . Chaque ligne correspond alors à l'état de l'automate à l'instant n .

FIGURE 5 : Construction algébrique d'un automate cellulaire suivant la règle 90 selon l'appellation de Wolfram.

Chapitre 3 -- Figure 5

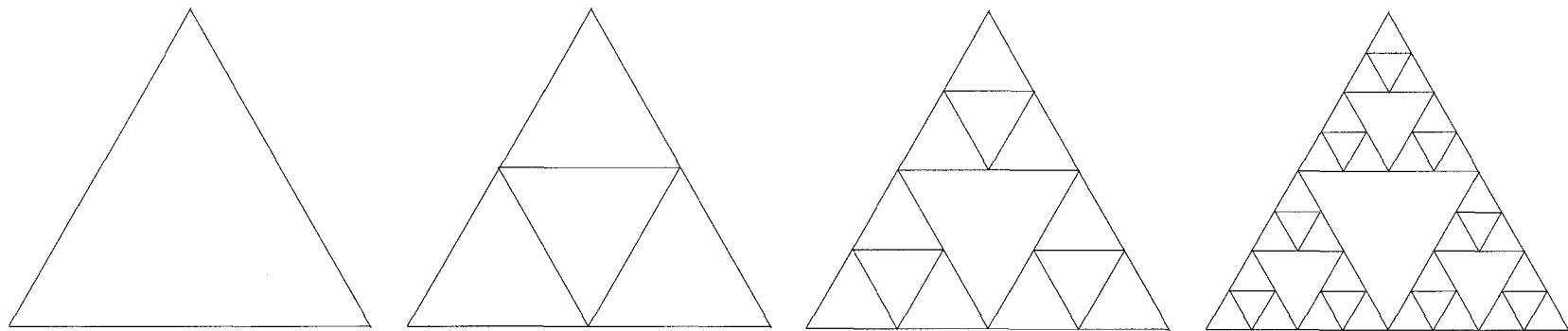


FIGURE 6 : Suite d'itérations de la construction géométrique du comportement d'un automate cellulaire évoluant suivant la règle 90 modulo 2. On observe le caractère fractal de la figure et l'invariance d'échelle (d'après Wolfram).

Chapitre 3 -- Figure 6



FIGURE 7 : L'évolution des automates correspondant aux règles 90 (01011010) et 126 (01111110). Les évolutions sont effectuées ici à partir d'une configuration initiale comportant un unique élément.

correspond alors à la valeur du site après application de la règle au voisinage si ce voisinage se trouve dans l'état 111. La règle 90 correspond alors à :

$$111 \rightarrow 0 ; 110 \rightarrow 1 ; 101 \rightarrow 0 ; 100 \rightarrow 1 ; 011 \rightarrow 1 ; 010 \rightarrow 0 ; 001 \rightarrow 1 ; 000 \rightarrow 0$$

Les trois premiers chiffres précisent la configuration du voisinage d'un site et le chiffre de droite est l'état du site qui en résulte. La suite de ces états, dans l'ordre ci-dessus, donne la représentation binaire de la règle. Le numéro de la règle est la valeur décimale correspondante.

Wolfram ajoute deux autres restrictions importantes. Il considère comme "illégales" les règles qui ne laisseraient pas inchangé un état initial ne comportant que des 0. Ceci explique que toutes les règles légales vérifieront $000 \rightarrow 0$. On appelle *condition d'absence de génération spontanée* cette contrainte. On n'a plus alors que 2^7 règles possibles. La deuxième restriction porte sur *la symétrie des règles*. Toutes les règles en un site doivent être symétriques par rapport à ce site. Par exemple, si $110 \rightarrow 1$, on a alors $011 \rightarrow 1$. Toutes les règles sont alors déterminées par le choix des valeurs résultantes de l'application de la règle aux voisins 111, 110, 101, 100 et 010. Il reste alors 2^5 soit 32 règles légales⁵. Le choix des contraintes de légalité des règles est arbitraire. L'absence de génération spontanée se justifie pour les phénomènes physiques, mais pas *a priori* pour les systèmes symboliques. La condition de symétrie des règles, quant à elle, a pour justification l'esthétique surtout, et le souhait de ne pas privilégier une direction particulière de l'automate. Le choix des conditions de légalité n'est donc pas guidé par des contraintes particulière liées à un phénomène particulier. Les seules règles légales qui subsistent s'écrivent alors en binaire $\alpha_1\alpha_2\alpha_3\alpha_4\alpha_2\alpha_5\alpha_40$, avec les différents α_i , valant 0 ou 1. Par exemple la règle 90 s'écrit en binaire 01011010, ce qui correspond à $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = 0$, $\alpha_4 = 1$, $\alpha_5 = 0$.

Si on applique la règle 90 à la configuration 010110110101011100010, et en supposant que le dernier site est relié au premier, on obtient 100110110000010110101. On représente alors les itérations successives en prenant pour axe des ordonnées orienté vers le bas le nombre d'itérations, et comme axe des abscisses la configuration correspondant à cette itération, en binaire, avec une étoile pour les 1 et un blanc pour les 0. On peut de cette façon itérer les 32 règles légales, en partant d'une configuration initiale comportant un seul site valant 1.

On observe pour la règle 90 que la figure obtenue en partant d'une configuration avec un seul site à la valeur 1 est auto-similaire, c'est-à-dire invariante d'échelle (figures 6 et 7). Cette figure correspond également [Wolfram, 1984b] à une représentation du développement de $(1+x)^n \bmod 2$, c'est-à-dire aussi au **triangle de Pascal** en nombre binaire (figure 5). La règle 150 quant à elle correspond au développement de $(1+x+x^2)^n \bmod 2$. Ainsi, si l'on considère ces règles d'un point de vue géométrique,

⁵Les 32 règles restantes sont les règles 0, 4, 18, 22, 32, 36, 50, 54, 72, 76, 90, 94, 104, 108, 122, 126, 128, 132, 146, 150, 160, 164, 178, 182, 200, 204, 218, 222, 232, 236, 250, 254.

certaines se comportent comme les constructions de Sierpinski qui sont invariantes d'échelle (figures 6 et 7).

On peut également observer le comportement des règles légales lorsque la configuration initiale est aléatoire⁶.

2.1.2. Propriétés locales des automates

On peut d'abord chercher à étudier les propriétés d'une règle en observant le comportement d'un unique site au cours du temps. On parle de propriété locale de l'automate. Pour cela, on étudie l'évolution de la densité de probabilité ρ_t des sites dont l'état vaut 1 après t itérations. On appelle ρ_0 la densité à l'instant initial. On utilise alors une *approximation statistique de champ moyen*⁷, qui ne tient pas compte des corrélations. Nous allons détailler la méthode utilisée sur l'exemple de la règle 90.

On suppose que chaque site admet à l'instant t la probabilité ρ_t d'être dans l'état 1 et donc la probabilité $1 - \rho_t$ d'être dans l'état 0. On suppose dans une première approximation que les probabilités de chaque configuration et donc de chaque voisinage sont indépendantes. On obtient, pour les voisinages 111, 110, 101, 100, 011, 010, 001 et 000, les probabilités respectives $(\rho_t)^3$, $(\rho_t)^2(1 - \rho_t)$, $(\rho_t)^2(1 - \rho_t)$, $(\rho_t)(1 - \rho_t)^2$, $(\rho_t)^2(1 - \rho_t)$, $(\rho_t)(1 - \rho_t)^2$, $(\rho_t)(1 - \rho_t)^2$, $(1 - \rho_t)^3$.

Grâce à l'hypothèse d'indépendance des probabilités d'existence de chaque voisinage, on peut calculer ρ_{t+1} , en fonction de ρ_t . Il suffit pour cela d'écrire que la variation de la probabilité est égale à la probabilité des sites inactifs de devenir actifs $\Gamma(0 \rightarrow 1)$, à laquelle on soustrait la probabilité d'un site actif de devenir inactif $\Gamma(1 \rightarrow 0)$. On écrit cela $\frac{\delta \rho_t}{\delta \tau} = \Gamma(0 \rightarrow 1) - \Gamma(1 \rightarrow 0)$. On obtient alors une équation, dite équation maîtresse⁸, qui nous donne l'évolution de la densité de probabilité ρ_t . Pour la règle 90, les voisinages qui contribuent à $\Gamma(0 \rightarrow 1)$ sont 100

⁶Lorsque l'on parle d'une configuration aléatoire on ne désigne ni une configuration quelconque, ni une configuration particulière, mais une configuration dont on pourrait dire qu'elle correspond à un nombre aléatoire. Pour cet automate cependant, une difficulté se pose, puisqu'un nombre aléatoire possède une description binaire avec un nombre infini de chiffres et n'est pas périodique. Or les automates de Wolfram sont finis, ou peuvent être considérés comme infinis mais périodiques. C'est par abus que l'on peut considérer comme aléatoire une configuration initiale qui correspondrait, au mieux, à un nombre aléatoire tronqué, mais n'aurait alors plus rien d'aléatoire. Il faut noter cependant que le raisonnement en termes de probabilités et en particulier l'approximation en champ moyen pour les automates ne prennent leur pertinence que si l'on a une position initiale aléatoire. Une fois de plus il faut composer avec les contraintes respectives et antagonistes liées aux descriptions finies et infinies. Ceci se fait avec difficulté.

C'est à partir de configurations initiales "aléatoires" que sont réalisées les figures caractéristiques de Wolfram avec des triangles "de toutes tailles".

⁷*Il s'agit de la même méthode que celle présentée dans la première section. Nous la retrouverons encore dans la suite de ce travail.*

⁸On appelle **équation maîtresse** l'équation d'évolution d'une densité de probabilité.

et 001. On a donc $\Gamma(0 \rightarrow 1) = 2(\rho_t)(1 - \rho_t)^2$. Les voisinages qui contribuent à $\Gamma(1 \rightarrow 0)$ sont 111 et 010. On a donc $\Gamma(1 \rightarrow 0) = (\rho_t)^3 + (\rho_t)(1 - \rho_t)^2$. Finalement on obtient l'équation maîtresse $\frac{\delta\rho_t}{\delta\tau} = (\rho_t)(1 - \rho_t)^2 - (\rho_t)^3 = (\rho_t)(1 - 2\rho_t)$.

On peut ensuite chercher la valeur stationnaire de cette densité de probabilité ρ_t , notée ρ^* . Il suffit d'écrire $\frac{\delta\rho^*}{\delta\tau} = 0$. On obtient alors les valeurs théoriques des ρ^* . Dans le cas de la règle 90, on obtient $(\rho^*)(1 - \rho^*)^2 - (\rho^*)^3 = 0$, soit $\rho^* = 0$ ou $\rho^* = 1/2$. Wolfram a calculé toutes les valeurs stationnaires des densités de probabilités pour toutes les règles légales.

On peut aussi étudier les corrélations entre deux sites distincts décalés de r sites. On définit pour cela la fonction $C^{(2)}(r) = \langle S(m)S(m+r) \rangle - \langle S(m) \rangle^2$, où les moyennes⁹ notées $\langle \rangle$ sont calculées sur toutes les positions m possibles dans l'automate à un instant fixé, $S(m)$ valant -1 si l'état du site m est 0 et 1 si l'état du site est $+1$. La fonction $C^{(2)}(r)$ est nulle pour $r > 0$ si la configuration est désordonnée. Pour certaines règles, comme la règle 18, on observe que $C^{(2)}(r)$ est non nulle, ce qui est le signe de présence d'une structure. Plus précisément, pour la règle 18, on observe une décroissance exponentielle de la corrélation $C^{(2)}(r)$ qui est de la forme $e^{-\frac{r}{\lambda}}$. On appelle alors **longueur de corrélation** la valeur de λ . Pour la règle 18 la longueur de corrélation est approximativement égale à 2 . Le fait de voir apparaître une longueur de corrélation non nulle est une première indication permettant de constater l'apparition d'ordre dans l'évolution de l'automate cellulaire. Les études sur la densité de probabilité et sur la fonction de corrélation indiquent que la transition vers le désordre, pour la règle 126 tout au moins, est continue et qu'elle ne correspond donc pas à un phénomène analogue à une transition de phase du deuxième ordre.

2.1.3. Propriétés globales des automates

L'approche précédente consistait à se donner les outils d'analyse de l'évolution d'une configuration au cours du temps. Une autre approche alternative consiste à étudier les propriétés statistiques de l'ensemble de toutes les configurations possibles d'un automate, ce qui constitue *l'analogie de l'espace des phases* en mécanique statistique classique. Cette approche est étroitement liée à la théorie des systèmes dynamiques. Dans cette partie, on ne considère que des automates finis avec des conditions initiales périodiques aux bords¹⁰. On utilise ensuite une **distance**, par exemple la

⁹On aurait pu noter la fonction corrélation $C^{(2)}(r) = \langle S(m)S(m+r) \rangle - \langle S(m) \rangle \langle S(m+r) \rangle$. Sous cette forme on voit mieux les rôles symétriques que jouent les sites m et $m + r$. Les corrélations d'ordre 3, c'est-à-dire entre trois sites, de même que les fonctions de corrélation suivantes, peuvent facilement être définies à partir des fonctions symétriques des racines d'un polynôme. On obtient par exemple : $C^{(3)}(p, q) = \langle S(m)S(m+p)S(m+q) \rangle - \langle S(m) \rangle [\langle S(m)S(m+p) \rangle + \langle S(m)S(m+q) \rangle + \langle S(m+p)S(m+q) \rangle] + 2 \langle S(m) \rangle^3$

¹⁰La représentation sous forme d'un tore amène naturellement ces conditions de périodicité aux bords.

distance de Hamming¹¹ dans l'espace des configurations de l'automate. On peut alors calculer une probabilité d'obtention de chaque état accessible, considérée pour l'ensemble de l'automate, après un pas de temps fini. On définit enfin une **entropie** S de l'automate, par exemple $S = \sum_i p_i \log_2(p_i)$, avec i qui décrit l'ensemble des états possibles, et p_i la probabilité de l'état i . Cette entropie est celle de la théorie de l'information¹².

L'entropie constitue une mesure de l'état global de l'automate, mais elle est en partie arbitraire. "Il est clair que le contenu brut en information¹³ ne détermine pas la valeur de l'information. C'est une évidence : la valeur d'une information est quelque chose de plus compliqué et de relatif. Et c'est bien parce que la valeur en information d'une chaîne de caractères est relative à un certain but et à une certaine situation qu'il y a plusieurs théories de l'information et qu'aucune ne peut vraiment traiter tous les problèmes que pose la notion d'information" [Delahaye, 1994][p. 16]¹⁴. De la même manière, on peut calculer une probabilité d'apparition de cycles de

¹¹Soient deux nombres a et b . Soient (a_i) et (b_i) les chiffres qui permettent d'écrire a et b en binaire. a_i est donc le $(i+1)^{\text{ème}}$ chiffre dans l'écriture de a en binaire (en partant de la droite). La **distance de Hamming** $d(a, b)$ entre deux nombres a et b est définie alors comme suit :

$$d(a, b) = \sum_i |a_i - b_i|$$

¹²On peut définir le manque d'information, comme l'entropie statistique pour un nombre fini d'événements à partir d'un nombre très faible d'hypothèses. On se donne une situation qui comporte N événements possibles. Chaque événement i a la probabilité P_i de se produire. On définit alors le manque d'information $S(P_1, \dots, P_N)$ de la façon suivante :

- (i) le manque d'information S est nul si un événement i est certain, c'est-à-dire si $P_i = 1$ et $P_j = 0$ pour $j \neq i$
- (ii) le manque d'information S est une fonction *positive, continue et symétrique* des variables P_i
- (iii) le manque d'information S est indépendant des événements impossibles, c'est-à-dire des événements qui ne font pas partie des N événements possibles
- (iv) le manque d'information S est croissant en fonction de N pour des événements équiprobables
- (v) le manque d'information S est *additif*

On obtient alors une unique expression du manque d'information ou de l'entropie statistique :

$$S(P_1, P_2, \dots, P_N) = -\lambda \sum_{i=1}^N P_i \ln P_i \text{ avec } \lambda \in \mathbb{R}$$

En théorie de l'information, on choisit de prendre $\lambda = \frac{1}{\ln 2}$, ce qui permet d'obtenir un manque d'information égal à 1 pour une situation comportant deux événements équiprobables, comme dans un jeu de pile ou face.

En physique statistique, on choisit pour λ la valeur de la constante de Boltzmann.

¹³"[...]un contenu brut d'information [...] est donné en digits (ou bits) et on peut le définir comme le nombre de mémoires élémentaires que la chaîne occupe dans la mémoire d'un ordinateur quand on ne lui fait subir aucun traitement particulier autre que la mise sous un format compatible avec le système de l'ordinateur" [Delahaye, 1994][p. 16].

¹⁴Cité par Patrice Prez dans [Prez, 1996][p. 3].

longueur n et définir une mesure, à partir cette probabilité, que l'on appellera encore entropie.

L'essentiel ici est de bien comprendre que lorsque l'on souhaite observer un système dans son ensemble il est nécessaire d'avoir une mesure globale sur l'état de ce système. On appelle toujours **entropie** cette mesure globale. L'entropie, c'est-à-dire la mesure globale est toujours en partie arbitraire. En fait, lorsque l'on souhaite observer un système global, ce qui est observé est pratiquement toujours déterminé de l'extérieur. On pourrait souhaiter au contraire qu'un paramètre mesurant l'état global du système soit défini de manière intrinsèque, de l'intérieur, par le système. Ceci est concevable mais n'est jamais pratiqué à l'heure actuelle. Pour cela il faudrait que l'entropie, la mesure globale, soit un paramètre déterminé comme un invariant interne correspondant à des états globaux stables du système dans son ensemble.

2.1.4. Extensions et applications

Wolfram a proposé différentes extensions des automates précédents¹⁵. Il a proposé par exemple de prendre comme voisinages les éléments à deux pas. On a alors cinq voisins par voisinage¹⁶. Il a également étudié les automates en dimension supérieure et en particulier en dimension deux. On peut alors définir les voisinages d'un site i soit comme les éléments comportant une seule coordonnée au plus qui diffère de i d'une unité, soit comme les éléments comportant des coordonnées qui diffèrent de celles de i d'au plus une unité. En dimension deux¹⁷ on obtient alors respectivement des voisinages à cinq¹⁸ ou neuf¹⁹ voisins pour un réseau carré. Les différentes simulations [Wolfram, 1984c] utilisant les outils que nous avons présentés précédemment mettent en évidence quatre classes et quatre types de comportements pour les automates en dimension une pour cinq voisins. La première classe fait apparaître des motifs ponctuels et homogènes. La seconde classe montre des états limites qui sont des structures simples, séparées²⁰ et périodiques. La troisième classe comporte des motifs apériodiques et chaotiques. La quatrième classe, enfin, évolue vers des motifs complexes de structures localisées. Sur les 32 règles autorisées en dimension une

¹⁵Une première extension consiste, en dimension une, avec trois éléments par voisinage, à considérer que chaque site peut prendre k valeurs. Il s'agit du même type de généralisation que celle de Potts par rapport au modèle d'Ising. Dans ce cas on obtient $k^{\left[\frac{k^2(1+k)}{2}-1\right]}$ règles légales.

¹⁶C'est également dans ce cadre que Wolfram a défini ses quatre classes d'automates initialement [Martin *et al.*, 1984] [Wolfram, 1984c].

¹⁷Les règles sont légales si elles sont invariantes par rotation et par symétrie et si de plus elles vérifient la condition d'absorption $000 \rightarrow 0$.

¹⁸On les appelle voisinages de von Neumann [von Neumann, 1966].

¹⁹On les appelle voisinages de Moore.

²⁰Séparées signifie ici que les structures ne s'enchevêtrent pas, ne s'emmêlent pas, qu'elles sont bien distinguables.

avec cinq voisins, on en dénombre 9 dans la classe 1 (les règles²¹ 0, 4, 16, 32, 36, 48, 54, 60 et 62) ; 5 dans la classe 2 (les règles 8, 24, 40, 56 et 58) ; 16 dans la classe 3 (les règles 2, 6, 10, 12, 14, 18, 22, 26, 28, 30, 34, 38, 42, 44, 46 et 50) et 2 dans la classe 4 (les règles 20 et 52).

Langton, dans son PhD puis dans différents articles [Langton, 1990], a approfondi la classification proposée par Wolfram, en introduisant un paramètre λ . Il travaille dans l'espace des règles possibles d'un automate à K états (chez Wolfram, on a $K = 2$, puisque les états possibles sont 0 et 1 ou $K = 3$, lorsque les états possibles sont 0, 1 et 2) et V voisins (chez Wolfram, le nombre de voisins est $V = 3$ ou $V = 5$). On définit λ comme suit. Soit s un état quelconque parmi les états possibles (on dira que cet état est *absorbant* ou *quiescent*). Soit n le nombre de transitions qui amènent à s dans la règle considérée. On a alors $K^V - n$ transitions de la même règle qui n'aboutissent pas à s . On pose alors $\lambda = \frac{K^V - n}{K^V}$. Lorsque $\lambda = 0$, on a une règle dont toutes les transitions aboutissent à l'état quiescent qui est alors l'état homogène d'arrivée. Si $\lambda = 1$, on a une règle pour laquelle aucune transition ne mène à l'état quiescent. L'état le plus hétérogène est obtenu lorsque tous les états sont représentés uniformément comme états d'arrivée des transitions et en particulier pour les transitions qui mènent à l'état quiescent. On a alors $n = \frac{1}{K}$ et $\lambda = 1 - \frac{1}{K}$. Les règles sont ainsi classées en fonction de ce paramètre λ que l'on peut faire varier progressivement. Langton constate que lorsque le paramètre λ augmente on voit apparaître successivement les différentes classes proposées par Wolfram. Cependant, contrairement à ce que pourrait faire croire la numérotation de Wolfram, la classe 1 est suivie de la classe 2, de la classe 4 puis de la classe 3. Ainsi la classe 4 intervient avant la classe 3. Elle constituerait ainsi une classe intermédiaire entre l'“ordre” qui caractérise les classes 1 et 2 et le “chaos” qui caractérise la classe 3. On a pu parler de “bord du chaos” [Langton, 1990] pour cette classe 4. La classe 4 serait la plus riche par la variété de ses comportements. Langton [Langton, 1992] pense même que cette classe caractérise l'état dans lequel se trouvent les systèmes vivants, entre cristal (classe 2) et fumée (classe 3).

2.1.5. Conclusion

Nous avons pu constater que malgré leur extrême simplicité les automates présentés ici ont des comportements extrêmement variés et difficiles à cerner. On peut distinguer deux grandes classes pour ces automates, l'une exhibant des comportements simples, et correspondant aux classes 1 et 2, et l'autre exhibant des comportements compliqués et correspondant aux classes 3 et 4. Ceci justifie *a posteriori* de considérer que les automates relèvent de la complexité, dans le premier sens que nous avons donné, c'est-à-dire comme quelque chose qui ne se réduit pas à un seul élément. En effet on observe plusieurs types de comportements très différenciés.

²¹Les numéros correspondent aux règles en dimension deux.

On peut dire également que ces études relèvent de la complexité parce qu'elles amènent certains comportements, comme ceux de la classe 3 et de la classe 4, qui sont compliqués, parce qu'on ne sait pas les expliquer facilement et scientifiquement aujourd'hui. Nous avons vu que, pour les classes compliquées, on observait des figures auto-similaires.

Nous préférons choisir de dire que ces études relèvent de la complexité parce qu'elles relèvent de la complexité/quantité. En effet, l'hypothèse de mise en tore, c'est-à-dire de périodicité, est une "ruse" pour contourner les difficultés de la mise en pratique de la quantité et de l'infini qui est nécessaire et qui lui est associée. En effet l'hypothèse de champ moyen et donc l'écriture de l'équation maîtresse qui permet de prévoir le comportement asymptotique de l'automate n'est possible qu'en supposant une décorrélation des probabilités des différentes configurations. Cette hypothèse ne peut être justifiée que dans le cas où l'on suppose un nombre infini d'éléments.

2.2. Une nouvelle classification des problèmes sur les automates

2.2.1. Vingt problèmes sur les automates cellulaires

Dans un article de 1985 [Wolfram, 1985], Wolfram a proposé vingt questions centrales sur les automates cellulaires. Ces questions permettent d'éclairer les recherches actuelles sur les automates cellulaires et les voies qu'elles se proposent d'explorer. Nous reprenons dans l'annexe 2 les intitulés des problèmes posés par Wolfram. Nous avons préféré laisser les énoncés de ces problèmes en annexe pour ne pas perdre notre fil conducteur sur la complexité/quantité et en particulier les conséquences que cette conception peut avoir sur la constitution des outils utilisés et des questions posées. Nous avons conservé les énoncés de Wolfram, mais ils restent relativement confus et peu explicites. Aussi la présentation de ces problèmes est-elle en bonne partie personnelle, ainsi que la discussion qui suit pour chacun d'entre eux. Ces problèmes sont posés de manière relativement large et ne sont certainement pas simples à résoudre. De plus il est probable que la solution complète d'un seul de ces problèmes ne peut être envisagée sans le traitement préalable d'un nombre important de points qui relèveront plutôt d'autres problèmes. On peut tout de même considérer ces problèmes comme très importants pour l'étude des automates cellulaires. Ils permettent de mettre en évidence les aspects techniques des questions en jeu sur la complexité/quantité. On observe ainsi que les questions et les outils développés sur les automates cellulaires sont complètement imprégnés d'une complexité considérée essentiellement sous l'angle de la quantité.

Nous proposons d'établir des liens entre ces différents problèmes, à travers une nouvelle classification des difficultés abordées. L'objectif de notre classification est

de mieux mettre en évidence les questions “génériques” auxquelles se trouve confrontée la complexité/quantité.

2.2.2. Les trois structures des problèmes sur les automates

On peut tout d'abord cerner le contexte, général et préformalisé, des problèmes sur les automates. Il faut pour cela préciser que ces problèmes sont associés à la définition et à l'étude des “bonnes structures” permettant d'étudier certains aspects d'un automate. Trois structures doivent en fait être précisées. La notion de structure ici doit être comprise comme une structure mathématique, c'est-à-dire un ensemble de symboles et de leurs relations. Ce pourrait être par exemple un espace métrique, un espace topologique, un graphe, un espace vectoriel normé, etc.

La première structure à définir dans le cas des automates, que nous appellerons S_1 , est celle de graphe, c'est-à-dire l'ensemble des relations entre les sites de l'automate. Cette première structure est directement associée à la définition de l'automate étudié. Dans le cas des automates cellulaires, nous avons vu que Wolfram choisit comme structure un réseau carré ou une chaîne discrète, finie et mise en tore. Mathématiquement l'ensemble ainsi défini est donc soit l'ensemble des entiers de Gauss $Z[i]$, soit l'ensemble Z/nZ . On peut définir sur ces ensembles une structure d'anneau²² et éventuellement une structure d'espace métrique en précisant une distance.

La deuxième structure à définir, S_2 , est celle des règles. Les règles sont ici identifiées à des chaînes de caractères, et la structure S_2 constitue l'ensemble des relations entre les règles. La structure S_2 est précisément définie par la donnée d'une distance sur l'espace des règles. Généralement on choisit la distance de Hamming qui donne à S_2 une structure d'espace métrique. Cette distance permet de préciser le degré de proximité entre deux règles. Ceci est nécessaire dès que l'on veut pouvoir comparer deux règles l'une à l'autre, par exemple pour en générer une troisième.

La troisième structure S_3 porte sur les comportements possibles de l'automate, c'est-à-dire sur les suites de configurations de l'automate. Contrairement aux sites ou aux règles auxquels on a associé respectivement les structures S_1 et S_2 , les comportements de l'automate ne sont pas clairement et formellement définis à partir des configurations. La première difficulté consiste donc à préciser clairement la notion de comportement et les caractéristiques permettant de le définir. La deuxième difficulté porte ensuite sur le type de structure en jeu. Nous avons vu que pour S_1 il s'agissait d'un anneau et pour S_2 d'un espace métrique. La question se pose alors de savoir quelle est la structure de l'espace S_3 . Généralement, on se pose la question sous la forme du problème inverse. On se donne les structures S_1 et S_2 .

²²Au sens mathématique, mais on peut aussi leur associer l'image d'une structure en anneau, c'est-à-dire sous forme de tore.

Ces structures vont être susceptibles de générer un certain comportement. On suppose que l'on peut associer à ce comportement une structure S_3 . On cherche alors à retrouver, c'est-à-dire à reconstruire, cette structure S_3 .

2.2.3. Définition des éléments de l'espace S_3

La première classe de problèmes de Wolfram est liée à la définition des éléments de l'espace S_3 (pb 1)²³. Les éléments de l'espace S_3 sont définis systématiquement grâce à une “mesure” globale des configurations de l'automate : l'entropie. La difficulté consiste alors à proposer une bonne définition de cette entropie pour préciser les éléments de S_3 (pb 4 et pb 14). Pour cela il faut par exemple chercher à établir les relations entre les différentes mesures globales qui sont possibles (pb 2 et pb 6).

Certaines mesures sont spatiales, dans le sens où elles portent sur une seule configuration, alors que d'autres sont temporelles, dans le sens où elles portent sur un ensemble de configurations, c'est-à-dire sur des configurations que l'on considère comme “successives”, puisque générées les unes par les autres. La question se pose alors de la relation entre mesures globales spatiales et temporelles (pb 2 et pb 12).

La redondance, comme répétition de morceaux de configurations, permet d'obtenir de manière “opérationnelle” une solution “approchée” au problème de la recherche des éléments de l'espace S_3 . La difficulté porte sur le fait qu'on ne connaît pas toujours la validité et la généralité de ces méthodes de recherche de redondance. Les méthodes d'analyse de la redondance peuvent être des méthodes bayesiennes, ou du même type que les MDL (Minimum Description Length). L'objectif est alors de réduire la redondance pour obtenir les éléments de l'espace S_3 .

2.2.4. Obtention de la structure de l'espace S_3

Si l'on suppose désormais que les éléments de l'espace S_3 sont bien définis, il faut maintenant étudier les relations entre ces éléments, c'est-à-dire déterminer la structure de l'espace S_3 et son type (pb 3). Pour obtenir cette structure, on va étudier les régularités de l'espace S_3 , afin de déterminer les relations entre les éléments de S_3 . De nouveau la stratégie consiste à repérer les régularités, les invariances et à réduire le bruit et les redondances pour déterminer une structure minimale (pb 5 et pb 8). Dans la question précédente, nous avons vu que cette stratégie permet de repérer les éléments de l'espace S_3 . Elle permet également de repérer les relations entre ces éléments. En ce sens la recherche de la structure de S_3 nécessite le passage d'un nombre potentiellement infini de configurations successives à une réduction finie de telles configurations.

²³Je précise entre parenthèses les différents problèmes de Wolfram que l'on peut associer à ces questions, et je les note **pb**.

2.2.5. La structure S_3 et le temps infini

Une nouvelle difficulté se pose alors, car la réduction et le passage de l'infini au fini n'est pas toujours possible et nécessite quoi qu'il arrive la définition de l'aspect particulier qui justifie une telle réduction. Tout passage de l'infini au fini nécessite l'explicitation et donc la justification de la réduction. Toute réduction est orientée vers un but. C'est en ce sens qu'il faut comprendre l'intérêt des problèmes liés à l'irréductibilité du comportement des automates (pb 18).

Plus généralement, le comportement des automates au bout d'un temps infini et donc les liens entre la structure S_3 et la configuration obtenue à partir d'une règle au bout d'un temps infini constituent une bonne part des interrogations sur le comportement des automates (pb 13).

2.2.6. Relations et stabilité des relations entre les différentes structures des espaces S_1 , S_2 et S_3

Les espaces S_1 , S_2 et S_3 sont définis les uns par rapport aux autres. Nous avons vu que jusqu'à présent on se donnait S_1 et S_2 et on cherchait S_3 . On peut se demander alors quelles sont les relations entre ces différentes structures. Wolfram cherche à étudier la répartition dans la structure S_2 des règles qui génèrent la structure S_3 . *Le problème se pose plus généralement des relations entre les structures des espaces S_1 , S_2 et S_3 .*

Nous avons vu que le temps infini était réduit par l'intermédiaire de la recherche des invariants temporels. L'espace quant à lui est étudié comme fini grâce à l'hypothèse préalable de périodicité. On définit la structure S_1 comme infinie mais périodique en espace. Pour la structure S_3 , que l'on doit extraire d'un ensemble infini de configurations successives, on va supposer qu'elle est finie. Ceci impose que l'ensemble infini des configurations successives admet une autre forme de périodicité, **la redondance**. Comme on a défini au préalable S_1 comme périodique, le problème se pose des modifications de la structure S_3 , en fonction de la période considérée, c'est-à-dire quand on change la taille de l'espace S_1 et en particulier lorsque l'on fait tendre cette taille vers l'infini. On avait supposé au départ que l'espace S_1 était périodique pour ne pas avoir à traiter directement l'infini de cet espace, il faut désormais étudier les modifications du comportement quand on change cette taille et qu'on la fait tendre éventuellement vers l'infini (pb 17).

On définit la structure S_3 à partir des modifications de la structure S_2 (les règles) construite sur S_1 (l'espace). Ne peut-on pas alors aller plus loin et définir une nouvelle structure de règles S_4 sur la structure S_3 , qui générera des superclasses avec une nouvelle structure S_5 . Wolfram s'interroge sur cette généralisation des constructions successives de structures (pb 20).

En ce qui concerne la relation entre les structures, la question se pose non seulement de leur dépendance, mais aussi de la stabilité de ces dépendances (pb 11).

2.2.7. Comparaison des structures S_1 , S_2 et S_3 , et des structures mathématiques de référence

Les problèmes posés par Wolfram ne se posent pas uniquement en termes de relations entre les structures S_1 , S_2 et S_3 . Un autre aspect important consiste à savoir situer ces trois espaces et leurs interrelations au sein des structures mathématiques de référence. C'est dans cette optique que l'on peut interpréter les différents problèmes de correspondance que pose Wolfram (pb 9, 10, 15, 16, 18, 19).

Tous les outils élaborés sont ainsi issus de l'infini du nombre des configurations successives, dont il faut extraire des quantités finies, et en particulier les structures pertinentes . Les outils utilisés dans ce cadre et les questions qui les motivent portent donc implicitement et constamment le poids de la quantité. Ils sont conçus et utilisés pour obtenir un compromis entre fini et infini.

3. TROIS MÉTHODOLOGIES DE LA COMPLEXITÉ/QUANTITÉ

Dans cette section, l'objectif principal est de montrer que la complexité/quantité peut comporter des méthodologies très différentes. Pour cela nous présentons dans cette section trois méthodologies différentes, à travers trois modèles distincts qui s'appuient tous les trois sur l'utilisation des automates. *Le premier modèle* présenté permet de cerner comment un "mathématicien appliqué" peut utiliser les outils de la complexité/quantité pour obtenir des solutions approchées d'équations aux dérivées partielles. Le modèle est celui des gaz sur réseau présenté dans la sous-section 1. Il permet d'étudier et de simuler le comportement de certains fluides. *Le deuxième modèle* présenté permet d'établir comment un "physicien" peut utiliser les outils de la complexité/quantité pour créer de nouveaux concepts, à partir de résultats de simulations. Le modèle est celui du tas de sable présenté dans la sous-section 2. Il permet d'introduire le concept d'état critique auto-organisé. *Le troisième modèle* présenté permet de comprendre comment un "informaticien" peut utiliser les outils de la complexité/quantité pour créer une simulation informatique qu'il étudiera pour elle-même ou grâce à une analyse mathématique dans l'approximation de champ moyen. Nous présentons pour cela un modèle de coévolution de différentes espèces dans la sous-section 3. Nous établissons dans la sous-section 4 que les méthodologies associées à chacun de ces modèles sont radicalement différentes. Nous discutons alors les implications que ces différentes méthodologies peuvent avoir pour notre compréhension de la complexité/quantité.

3.1. Les modèles de gaz sur réseau

3.1.1. Introduction aux gaz sur réseau

Les modèles de gaz sur réseau¹ servent à simuler le comportement des fluides à partir de modèles proches des automates cellulaires. Wolfram fait partie des premiers à avoir proposé cette approche pour étudier certains problèmes d'hydrodynamique et en particulier de turbulence [Wolfram et Salem, 1985][Wolfram, 1986a]. On peut essayer de comprendre la complexité générale des fluides avec des systèmes dynamiques

¹Les principaux articles concernant les méthodes de gaz sur réseau ont été rassemblés dans [Doolen *et al.*, 1990] et [Doolen, 1991].

non linéaires comportant peu de degrés de liberté. Cependant, lorsque l'on cherche à étudier des systèmes comportant de nombreux degrés de liberté, par exemple pour des fluides dans un état où leur nombre de Reynolds est important², on possède peu de résultats théoriques, et seules des simulations numériques peuvent être envisagées. Les architectures massivement parallèles, avec des algorithmes adaptés, sont nécessaires pour ces simulations.

Lorsque l'espace des paramètres doit être parcouru rapidement et qu'une précision extrême n'est pas nécessaire, l'utilisation des nombres réels n'est pas obligatoire. Or, comme la représentation réelle est une des causes majeures de l'instabilité des algorithmes³, on peut se contenter de construire ici des algorithmes discrets. On construit donc des algorithmes discrets de gaz sur réseau qui tendent asymptotiquement vers les solutions de l'équation de Navier-Stokes pour les fluides incompressibles⁴ en dimension deux ou trois. Cela permet de vérifier en fait que certaines équations de champs intéressantes peuvent être approximées par le comportement à grande échelle d'automates cellulaires adéquats. Cette attitude permet un parallélisme naturel et un traitement égal de tous les bits⁵ pour des systèmes qui évoluent selon des règles d'automates cellulaires discrets avec seulement des interactions locales.

Trois niveaux de description sont possibles pour étudier les fluides. Le premier niveau est moléculaire, avec l'écriture d'un hamiltonien réversible. Le deuxième niveau est cinétique dans l'approximation de Boltzmann⁶ irréversible avec une faible

²De 10.000 à 10.000.000, ce qui correspond à des phénomènes d'extrême turbulence.

³Du fait du bruit lié aux arrondis, les calculs effectués avec des nombres réels représentés en virgule flottante peuvent aboutir à des situations qui ne sont pas physiquement réalistes. C'est pourquoi il vaut mieux avoir un traitement égal de tous les bits. "En principe, les schèmes qui donnent un poids égal à chaque bit seraient préférables" [Frisch *et al.*, 1986].

⁴L'équation standard de Navier-Stokes pour les fluides incompressibles s'écrit $\frac{\partial \vec{u}}{\partial t} + (\vec{u} \cdot \vec{\nabla}) \vec{u} = -\vec{\nabla} p + \nu \Delta \vec{u}$, avec \vec{u} la vitesse moyenne du fluide, p sa pression et ν la viscosité (on choisit classiquement de prendre une valeur de ν autour de 0,2 en dimension 2). L'équation standard et générale de Navier-Stokes en dimension d s'écrit $\frac{\partial(\rho \vec{u})}{\partial t} + \rho(\vec{u} \cdot \vec{\nabla}) \vec{u} = -\vec{\nabla} p + \eta \Delta \vec{u} + (\frac{1}{d}\eta) \vec{\nabla}(div \vec{u})$, avec ρ la masse volumique et η la viscosité. On a $\eta = \rho\nu$. C'est sous cette forme que l'équation a été étudiée dans le cadre des automates cellulaires [Wolfram, 1986a].

⁵Ceci est possible avec un traitement numérique parallèle et discret, mais pas avec des manipulations de nombres réels. En effet, lorsque l'on manipule des nombres réels codés sur 64 bits et que l'on cherche à obtenir une précision de 5 bits, seule une partie de ces 64 bits est effectivement utilisée de manière systématique dans les simulations.

⁶L'approximation de Boltzmann [Frisch *et al.*, 1987] consiste à supposer que les particules n'ont pas de corrélation antérieure au moment où elles participent à un choc. Pour les gaz ordinaires, cette approximation est associée aux situations de faible densité, pour lesquelles le libre parcours moyen est suffisamment grand pour que les particules entrant en collision proviennent presque sûrement de régions distantes et décorrélées. On constate que l'approximation de Boltzmann admet en fait un domaine de validité beaucoup plus large qui ne se restreint pas aux faibles densités. L'approximation de Boltzmann n'est pas non plus valable dans tous les cas de faible densité.

densité. Le troisième niveau est macroscopique avec une approximation continue du fluide. Les deux premières approches sont valables pour un fluide proche de l'équilibre thermodynamique. La troisième description possède des variables thermodynamiques libres comme la densité locale, le moment, la température ... Les modèles de Boltzmann microscopiques sont déterministes, discrets pour la vitesse, mais continus en espace et en temps. On aboutit à des équations continues données *a priori*. Au contraire, les gaz sur réseau sont des modèles discrets pour la vitesse, l'espace des phases et le temps. On peut parler alors de molécules booléennes. L'exemple le plus simple est le modèle de Hardy, Pazzis et Pomeau, ou modèle HPP [Hardy *et al.*, 1973][Hardy *et al.*, 1976].

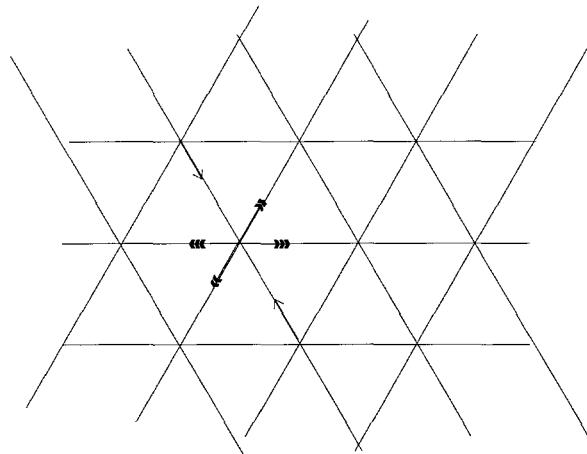
3.1.2. Le modèle fondateur de Frisch, Hasslacher et Pomeau

Le modèle HPP a été repris et adapté par Frisch, Hasslacher et Pomeau [Frisch *et al.*, 1986] en vue d'une implémentation sur réseau. Le modèle HPP s'appuie sur un réseau sous-jacent régulier et carré en deux dimensions. En chaque site on peut trouver jusqu'à quatre molécules de masse égale, chacune ayant alors une vitesse selon chacune des directions du réseau. Un même site ne peut pas être occupé par deux particules identiques avec des vitesses ayant même direction. Ce principe sera appelé **principe d'exclusion**. Le temps est discret, et à chaque pas de temps une molécule ne peut se déplacer que d'un lien dans la direction de sa vitesse. On trouve également des **règles de collision**. On appelle collision la rencontre de deux particules qui se trouvent au même instant en un même site du réseau. En cas de collision, on remplace deux molécules se déplaçant dans des directions opposées par deux nouvelles molécules dont la nouvelle vitesse est à angle droit de l'ancienne. Toutes les autres configurations sont laissées inchangées. C'est donc la seule règle de collision.

Le modèle HPP a différentes propriétés. La relaxation vers l'équilibre a été montrée numériquement. On ne connaît pas de théorème ergodique, qui nous assurerait théoriquement que l'on se situe bien à l'équilibre thermodynamique. Ces équilibres possèdent des paramètres libres et continus, comme la **densité moyenne** de particules et le **moment**. Cependant, lorsque ces paramètres varient lentement dans l'espace et dans le temps, les équations "macrodynamiques" qui apparaissent diffèrent des équations non linéaires de Navier-Stokes sur trois points que nous allons discuter successivement [Frisch *et al.*, 1986]. Nous retrouverons cette discussion sur ces trois points dans tous les modèles de gaz sur réseau : ils sont au cœur de ces modèles et bien spécifiques aux problèmes de physique étudiés. En particulier ces trois points ne se retrouvent pas directement sous cette forme dans les études sur les automates cellulaires.

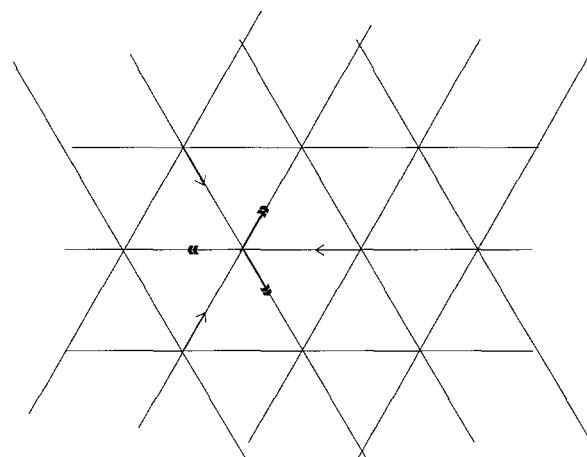
Le premier problème du modèle HPP est l'absence dans ce modèle d'une **invariance galiléenne**. Le second est un problème de **croisement de dimension**,

Les règles de collisions à deux corps



Deux particules, représentées ici avec une seule flèche, viennent l'une du lien i et l'autre du lien opposé $(i+3)$ et entrent en collision. Après la collision les deux particules repartent dans des directions opposées. Soit les deux particules repartent suivant les liens $(i-1,i+2)$ représentés ici avec les doubles flèches, soit elles repartent suivant les liens $(i+1,i-2)$ représentés ici avec les triples flèches. Les deux possibilités ont lieu avec une même probabilité $1/2$.

Les règles de collisions à trois corps



Trois particules, représentées ici avec une seule flèche, viennent l'une du lien i , la seconde du lien $(i+2)$ et la troisième du lien $(i-2)$ et entrent en collision. Après la collision les trois particules repartent suivant les liens $(i-1,i+1,i+3)$ représentés ici avec les doubles flèches. D'autres règles de collisions ont été proposées dans les versions suivantes du modèle HLG.

FIGURE 8 : Les règles de collision pour le modèle HLG sur un réseau triangulaire aux symétries hexagonales.

et le troisième est l'absence d'**isotropie**. La discussion de ces différents points a amené Frisch, Hasslacher et Pomeau à reconsidérer le modèle HPP, pour le modifier en résolvant ces trois questions.

L'**invariance galiléenne** est rompue par le réseau⁷. En conséquence il n'est pas possible d'établir une relation simple permettant de relier des équilibres thermodynamiques avec des vitesses différentes. Ceci se retrouve dans le terme non linéaire de l'équation du moment. Elle contient un tenseur de flux du moment, qui possède non seulement des termes quadratiques en la vitesse \vec{u} , comme il se doit dans l'équation de Navier-Stokes, mais aussi des termes de correction non linéaires de tout ordre, fonctions de $\|\vec{u}\|$. Cependant, ces termes sont négligeables pour des fluides dans un état où le nombre de Mach est faible, ce qui garantit également l'incompressibilité du fluide. Il faut noter également que l'invariance galiléenne qui ne se retrouve pas au niveau du réseau peut-être restaurée au niveau macroscopique.

Le problème du croisement de dimension se résume en l'observation d'une dépendance d'échelle de la viscosité, qui est logarithmique et non linéaire. C'est une propriété générale de l'hydrodynamique en dimension deux, lorsque l'on ajoute un bruit thermique à l'équation de Navier-Stokes ou au modèle HPP. Ce problème n'intervient pas en trois dimensions. Ce phénomène de croisement de dimension est fréquent en théorie des champs et dans la théorie des transitions de phase.

En ce qui concerne l'**isotropie**, on peut noter que le modèle HPP est invariant pour une rotation d'un angle $\frac{\pi}{2}$. Un tel réseau est insuffisant pour assurer l'isotropie du tenseur à quatre degrés de liberté reliant le flux du moment aux termes quadratiques en \vec{u} . On peut observer que ce tenseur est isotrope en dimension deux, pour une invariance sous une rotation du groupe hexagonal, c'est-à-dire multiple d'un angle de $\frac{\pi}{3}$. On obtient alors la forme correcte du terme correspondant dans l'équation de Navier-Stokes pour un nombre de Mach faible.

Suite à cette remarque, il est naturel d'utiliser un réseau triangulaire plutôt qu'un réseau carré. Chaque site possède alors six voisins. Ce modèle, proposé par Frisch, Hasslacher et Pomeau, est appelé HLG, pour Hexagonal Lattice Gas. Les règles de collision du modèle HPP doivent alors être modifiées. On numérote les six liens de chaque site, de 0 à 5. On peut alors avoir des collisions à deux corps et à trois corps (figure 8). Une collision à deux corps, venant l'un du lien i et l'autre du lien $i + 3$, c'est-à-dire de $(i, i + 3)$ amènera les deux protagonistes sur les liens $(i + 1, i - 2)$ ou $(i - 1, i + 2)$, ces deux sorties possibles pouvant apparaître avec une probabilité égale. Pour une collision à trois corps, on a $(i, i + 2, i - 2)$ qui se retrouve en $(i + 3, i + 1, i - 1)$. Ces règles de collision ont été conçues pour conserver le nombre de particules et le moment en chaque site. On obtient ainsi un total de trois relations de conservation scalaires.

⁷L'invariance par transformation galiléenne fait partie des hypothèses qui ont permis l'écriture de l'équation de Navier-Stokes.

Il faut noter qu'en dimension trois l'équivalent du réseau hexagonal serait le réseau cubique faces centrées, qui ne convient pas pour la nouvelle expression du tenseur car elle rompt une des symétries de l'équation de Navier-Stokes. Cet inconvénient peut être résolu par une méthode de partage, qui consiste à décomposer le terme non linéaire de l'équation de Navier-Stokes en une somme de deux termes, qui peuvent, pour chacun, être obtenus sur des réseaux différents, par exemple sur un réseau cubique faces centrées pour l'un et sur un réseau cubique régulier pour l'autre.

3.1.3. Les propriétés des gaz sur réseaux

Dans ces modèles, les particules sont des points pouvant se déplacer sur les sites d'un réseau régulier et fixe. Le problème est discret en espace, en temps et pour les vitesses. Les particules sont caractérisées par leur position, leur vitesse, en particulier la direction de cette vitesse, et enfin par d'autres paramètres éventuels comme la masse, etc. La dynamique se déroule en deux étapes : la première est la **propagation des particules** selon la direction de la vitesse de ces particules ; la seconde est celle de la **collision des molécules** avec des règles qui sont fonction des lois de conservation du nombre de particules, du moment, de l'énergie, etc. Il s'agit bien d'un modèle formel en lui-même et non d'une version hypersimplifiée d'un système de dynamiques moléculaires.

On peut extraire quatre principes de ces modèles. Le premier est lié à la **discrétisation**. Le second est le **principe d'exclusion** qui interdit à deux particules de mêmes caractéristiques de se trouver au même moment en un même site. Le troisième principe est celui de la **synchronisation** des déplacements. Le quatrième principe demande une grande **simplicité des règles de collisions**.

On peut distinguer trois avantages à ce type de modèles. Le premier est lié au fait qu'on travaille uniquement sur des entiers : c'est ce que l'on appelle la "démocratie des bits", puisque chaque bit a la même importance dans la simulation numérique. Le deuxième avantage est la complète parallélisation du modèle et des algorithmes qui en résultent. Le troisième avantage est lié aux facilités de simulation de ces modèles : le codage, les règles de collision, les conditions aux bords, l'absence de traitement spécifique des aspects non linéaires, la facilité de visualisation, la possibilité d'explorer simplement l'espace d'état, l'absence de traitement spécifique des régions à variation rapide sont des éléments qui permettent un traitement simple à mettre en place et complètement parallèle. Ces avantages permettent également des modifications rapides des programmes informatiques de simulation. On peut ainsi réaliser rapidement de nombreux tests de ces automates dans des conditions très variées. Des simulations ont été effectuées comportant jusqu'à cinq milliards de cellules. Les techniques de gaz sur réseau permettent, à l'aide d'automates cellulaires, d'approximer les solutions des équations aux dérivées par-

tielles de l'hydrodynamique, et d'étudier grâce à des simulations le comportement de fluides dans certaines conditions de turbulence.

3.1.4. Extension de l'applicabilité en physique

Nous avons déjà vu que le modèle HLG s'applique pour des fluides incompressibles, avec un nombre de Mach suffisamment faible. Par la suite d'autres études ont été menées pour étendre l'applicabilité des modèles de gaz sur réseau à d'autres types de fluides. L'instabilité de Kelvin-Helmholtz⁸ a été étudiée, ainsi que d'autres modèles pour d'autres nombres de Mach. Des études ont permis d'établir d'autres invariants que ceux proposés dans le modèle HLG. Pour les problèmes périodiques et sur les machines actuelles, les méthodes de gaz sur réseau sont moins rapides que les méthodes spectrales. Cependant ces méthodes sont intéressantes lorsque les conditions aux bords ne sont pas simples et ne sont pas solubles avec les autres méthodes.

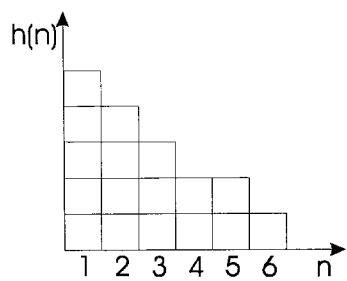
On peut distinguer actuellement quatre communautés scientifiques qui utilisent les gaz sur réseau. La première étudie les dynamiques moléculaires ; elle recherche des stratégies minimales en nombre de bits utilisés pour résoudre l'équation de Newton du mouvement ; elle emploie beaucoup plus de cellules dans les simulations que les autres méthodes. La deuxième communauté est celle des théoriciens des algorithmes de différences finies, qui trouvent avec les gaz sur réseau une méthode simplifiée. La troisième communauté est celle des chercheurs en mécanique statistique qui utilisent cette méthode pour mieux comprendre la relation entre microdynamique et macrodynamique. La quatrième et dernière communauté est constituée des chercheurs qui étudient les machines massivement parallèles et qui considèrent la méthode des gaz sur réseau comme la plus simple et la plus rapide totalement parallèle, avec des applications très variées. Comme nous l'avons vu les problèmes plus particulièrement étudiés concernent l'hydrodynamique et les questions de viscosité, l'étude spécifique de l'équation de Navier-Stokes, les problèmes à phases multiples et en milieux poreux, les phénomènes de diffusion et les réactions chimiques ainsi que l'étude des couplages et de l'auto-corrélation.

3.2. Un modèle du tas de sable

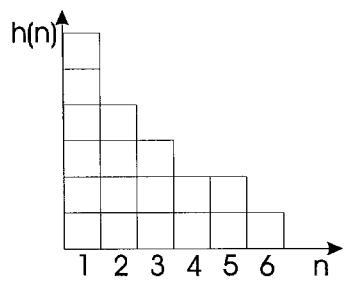
3.2.1. Le modèle fondateur de Bak, Tang et Wiesenfeld

Le modèle du tas de sable peut-être considéré comme l'*exemple paradigmique des modèles d'états critiques auto-organisés*. Il a été introduit comme exemple par Per

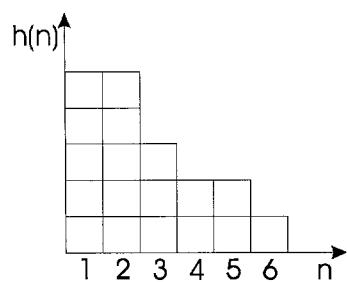
⁸L'instabilité de Kelvin-Helmholtz est une instabilité du fluide qui se développe lorsque la vitesse du fluide est tangentielle à une fine pellicule du fluide, une couche, et si de plus cette vitesse change en amplitude au sein de cette couche. On observe une instabilité issue d'une sorte de "cisaillement" du fluide.



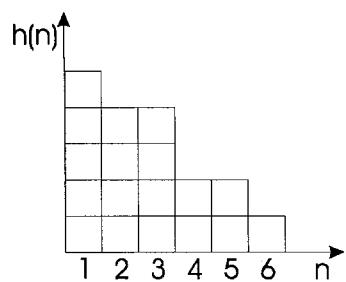
Configuration initiale, avec six colonnes de sable ; chaque carré correspond à un grain



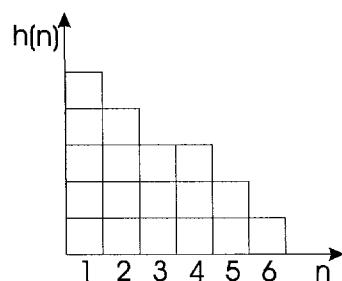
Configuration après application de la règle d'ajout en 1



Configuration après application de la règle de diffusion en 1



Configuration après application de la règle de diffusion en 2



Configuration après application de la règle de diffusion en 3. La taille de l'avalanche et sa durée sont ici égales à 3. La dernière configuration est métastable. Si on applique de nouveau la règle d'ajout en 1, on obtiendra une nouvelle avalanche.

FIGURE 9 : La dynamique du tas de sable en dimension une pour une pente critique égale à un ($z_c=1$). Comme la pente critique est égale à un, la règle de diffusion s'applique dès que la différence de hauteur entre deux colonnes de sable successives est égale à deux.

Chapitre 3 -- Figure 9

Bak [Bak *et al.*, 1988]. Les liens entre ce modèle et le comportement réel des tas de sable n'est pas clair aujourd'hui. Cependant cet exemple reste le plus parlant pour une introduction à la notion d'état critique auto-organisé. Dans des expériences sur des tas de sable réels, on observe une pente critique caractéristique. Si l'on ajoute du sable, la pente pourra être légèrement inférieure (on dira **sous-critique**) ou supérieure (on dira **sur-critique**) à la pente critique. On observe que les tas de sable avec des pentes sous-critiques ou sur-critiques tendent naturellement après relaxation vers l'état critique avec sa pente critique caractéristique. Le tas est instable en de nombreux endroits, mais l'état critique reste un attracteur global. Lorsque l'on perturbe le tas au voisinage de l'état critique, on peut observer son caractère attracteur. On observe également des invariants statistiques comme la distribution des avalanches en taille et en durée.

Le modèle proposé par Per Bak, Tang et Wiesenfeld pour rendre compte du tas de sable est un automate cellulaire (figure 9). En dimension une, une pile de sable est modélisée par N colonnes de sable juxtaposées. Chaque colonne n à l'instant t est caractérisée par sa hauteur $h(n, t)$ qui possède une valeur entière positive ou nulle. On a $h(n, t) \in \mathbb{N}$. La dynamique de l'automate cellulaire s'exprime alors en fonction de $z(n, t) = h(n, t) - h(n + 1, t)$. La valeur $z(n, t)$ peut être interprétée comme la pente du tas de sable en n à l'instant t . Un grain de sable pourra tomber de la colonne n à la colonne $n + 1$, en fonction de $z(n)$.

La dynamique comporte deux règles. La première règle ou **règle d'ajout** correspond à une *perturbation* du tas de sable. Elle consiste à ajouter un grain de sable dans une colonne n à l'instant t . On a alors une modification de la hauteur de la colonne n à l'instant $t + 1$:

$$h(n, t + 1) = h(n, t) + 1$$

Cette règle peut être exprimée en fonction de $z(n, t)$. On obtient alors le système qui traduit l'évolution des $z(n, t)$ lorsque l'on ajoute un grain en n à l'instant t :

$$\begin{cases} z(n - 1, t + 1) = z(n - 1, t) - 1 \\ z(n, t + 1) = z(n, t) + 1 \end{cases}$$

La deuxième règle ou **règle de diffusion** correspond à une *relaxation du tas de sable* si la pente en n est trop importante. La règle exprime le basculement d'un grain de la colonne n vers la colonne $n + 1$. On obtient alors le système :

$$\begin{cases} h(n, t + 1) = h(n, t) - 1 \\ h(n + 1, t + 1) = h(n + 1, t) + 1 \end{cases}$$

Cette règle peut être exprimée en fonction de $z(n, t)$, comme précédemment. On obtient alors le système qui traduit l'évolution des $z(n, t)$ par diffusion d'un grain de sable de n vers $n + 1$ lorsque la pente en n soit $z(n, t)$ a dépassé un seuil critique z_c , c'est-à-dire si $z(n, t) > z_c$:

$$\left\{ \begin{array}{l} \text{Si } z(n, t) > z_c \\ z(n-1, t+1) = z(n-1, t) + 1 \\ z(n, t+1) = z(n, t) - 2 \\ z(n+1, t+1) = z(n+1, t) + 1 \end{array} \right.$$

Un tel système possède en fait un nombre très important d'états méta-stables, à savoir tous les états accessibles. L'état critique est atteint lorsque tous les $z(n, t)$ valent z_c . Les configurations évoluent donc par addition de grains vers l'état critique qui est globalement attractif. Cependant, lorsque l'on se trouve exactement à l'état critique, l'ajout d'un grain en n'importe quelle position introduit une instabilité qui va entraîner une relaxation qui suivra la deuxième règle. *L'état critique qui est attracteur est donc également l'état le plus instable.*

En dimension deux et trois, on peut définir des règles équivalentes à celles proposées en dimension une et qui traduisent de nouveau l'évolution d'une pente du tas de sable. On observe que l'état attracteur, qui est aussi l'état critique, apparaît lorsque les perturbations peuvent se propager à longue portée. C'est-à-dire, dans un vocabulaire que nous avons déjà utilisé précédemment, lorsque la longueur de corrélation est infinie. Il n'y a alors plus d'échelle caractéristique de l'automate, ni en temps, ni en espace. On peut observer un déclenchement de la deuxième règle, c'est-à-dire de la diffusion en n'importe quel site du réseau et un nombre quelconque de fois.

On définit alors la taille d'une avalanche comme le nombre de grains s qui ont basculé entre deux ajouts successifs d'un grain. La taille d'une avalanche peut donc être obtenue par le calcul du nombre de fois où l'on a itéré la règle de diffusion entre deux applications de la règle d'ajout. À l'état critique, en dimension deux, on observe des avalanches de toute taille. Plus précisément, au voisinage de l'état critique, les distributions des avalanches en taille et en durée obéissent à des lois puissances. Si l'on appelle $D(s)$ la fonction de distribution des avalanches de taille s , on observe une loi puissance et si l'on appelle τ l'exposant critique associé, on obtient $D(s) \sim s^{-\tau}$. Soit T la durée des avalanches, c'est-à-dire le nombre de pas de temps que dure une avalanche. On a généralement T qui est inférieur à s , puisqu'à un instant t plusieurs grains issus de sites différents peuvent tomber en même temps. On obtient alors pour la fonction de distribution $D(T)$ des avalanches de durée T pondérée par la réponse moyenne s/T , une loi puissance dont l'exposant critique est α . On a alors $D(T) \sim T^{-\alpha}$. On appelle $f(x, t)$, l'énergie dissipée au site x à l'instant t et qui vaut 1 de plus pour chaque chute d'un nouveau grain, et $F(t)$ l'énergie totale dissipée à l'instant t sur l'ensemble du système. On peut alors caractériser le spectre $S(f)$ de $F(t)$. On obtient de nouveau une loi puissance d'exposant critique β . On a $S(f) \sim f^{-\beta}$. On observe enfin une autre loi puissance pour s en fonction de T d'exposant critique γ . On a alors $s \sim T^{-\gamma}$.

Nous reprenons ici les valeurs numériques obtenues pour les exposants critiques de ces lois puissances par Per Bak grâce à des simulations en dimensions 2 et 3. En dimension 2, les expériences numériques ont été réalisées sur un réseau 50×50 avec 200 essais. En dimension 3, les expériences numériques ont été réalisées sur un réseau $20 \times 20 \times 20$ avec 200 essais. Les nombres dans le tableau suivant sont donc respectivement les valeurs de τ , α , β et γ .

lois puissances associées	$D(s) \sim s^{-\tau}$	$D(T) \sim T^{-\alpha}$	$D(f) \sim f^{-\beta}$	$s \sim T^{-\gamma}$
dimension 2	$\tau = 1,00$	$\alpha = 0,43$	$\beta = 1,57$	$\gamma = 0,57$
dimension 3	$\tau = 1,37$	$\alpha = 0,92$	$\beta = 1,08$	$\gamma = 0,71$

On obtient alors la loi d'échelle suivante : $\alpha = 2 - \beta = (\gamma + 1)\tau - 2\gamma$. On peut définir d'autres lois d'échelle en définissant d'autres exposants critiques. Quoi qu'il arrive, les lois d'échelle laisseront exactement deux exposants critiques indépendants. Les valeurs de ces deux exposants critiques (par exemple α et τ) caractériseront l'ensemble des lois puissances observables sur le tas de sable.

3.2.2. La notion d'état critique auto-organisé

Le terme **d'état critique auto-organisé** a été introduit dans [Bak *et al.*, 1988] dans le cadre du modèle du tas de sable présenté ci-dessus. En effet, pour ce modèle, on observe que certaines grandeurs du système, comme la taille des avalanches, suivent des lois puissances pour une certaine configuration méta-stable du système, définie par une pente caractéristique. Les avalanches peuvent donc être de toutes tailles et pour cette grandeur on peut définir un exposant critique α . De ce constat, on infère que cette configuration méta-stable du système est un état critique. De plus la dynamique d'évolution du tas de sable nous ramène vers l'état critique naturellement et de façon interne lorsqu'on l'a quitté. On parle alors d'*auto-organisation du tas de stable*, l'état critique étant auto-organisé.

Cette notion d'état critique revient donc à une sorte d'*internalisation du paramètre de contrôle*. La mécanique serait la suivante. D'une part *le paramètre de contrôle agit sur le paramètre d'ordre*, comme c'est toujours le cas lorsque l'on se trouve à l'état critique et comme nous l'avons déjà précisé dans la première section. Mais d'autre part on peut supposer que dans le cas des états critiques auto-organisés, *le paramètre d'ordre rétroagit sur le paramètre de contrôle*. On n'est pas capable à l'heure actuelle de spécifier comment se déroule cette rétroaction.

Il est à noter que la notion d'état critique auto-organisé est d'abord introduite par l'observation de lois puissances caractéristiques des phénomènes critiques. On ne commence pas par définir des phases et une transition entre ces phases. On observe d'abord sur une simulation un comportement intéressant, que l'on caractérise par

une loi puissance. Cette approche ne va pas sans soulever encore à l'heure actuelle de vifs débats. Peut-on se contenter de caractériser la structure de fonctionnement d'un automate à partir simplement de quelques mesures sur une simulation informatique ? Il me semblait particulièrement important de présenter ici les états critiques auto-organisés car ils partagent, en dehors de leur intérêt propre, au moins deux points communs avec le but que nous poursuivons. Tout d'abord ils ont été introduits grâce à des outils propres à la physique statistique, ce qui montre la puissance de ces outils et leur universalité. De plus le modèle présenté par Per Bak a été conçu en premier lieu comme un automate cellulaire. La notion d'état critique auto-organisé définie par Per Bak n'est pourtant pas *a priori* spécifique aux automates. De nouveau est en jeu la facilité d'étude et de simulation des automates, ainsi que la possibilité relativement aisée d'utiliser de manière systématique des outils puissants issus de la physique. C'est pourquoi la notion d'état critique auto-organisé a d'abord été comprise dans le cadre des automates cellulaires.

Par la suite le concept d'état critique auto-organisé a été très étudié au début des années quatre-vingt-dix [Prado et Olami, 1992]. On le retrouve désormais comme un concept fédérateur dans des domaines extrêmement variés : en économie [Bak *et al.*, 1992a], en hydrodynamique, dans le cadre des gaz sur réseau [Andersen *et al.*, 1991][Bussemaker et Ernst, 1993], pour étudier des feux de forêt [Drossel et Schwabl, 1992], des tremblements de terre [Sornette et Sornette, 1989], ou pour des modèles de trafic routier [Biham *et al.*, 1992].

3.3. Un modèle d'évolution

Bak, Flyvbjerg et Lautrup, ont proposé un modèle de coévolution [Bak *et al.*, 1992b] qu'ils ont étudié dans une série d'articles [Flyvbjerg et Lautrup, 1992] [Flyvbjerg *et al.*, 1993]. L'étude de ce modèle s'appuie essentiellement sur le concept d'état critique auto-organisé et sur une approximation en champ moyen.

Le but de ce modèle est d'étudier l'évolution d'espèces interdépendantes. Le modèle proposé s'inscrit dans la suite des études du modèle NK [Kauffman, 1989] puis du modèle NKC [Kauffman et Johnsen, 1991] de Kauffman, pour lequel il a été montré que l'on observe une transition d'un état ordonné vers un état désordonné lorsque la dépendance entre les espèces augmente. Cependant toutes les études entreprises montrent qu'un paramètre externe est nécessaire pour franchir la transition, et l'on ne peut donc pas parler d'état critique auto-organisé. Per Bak, Flyvbjerg et Lautrup [Bak *et al.*, 1992b] présentent un nouveau modèle qui suit la même philosophie.

Comme dans tous ces modèles, les espèces sont représentées par un "code génétique", c'est-à-dire une chaîne de bits. Chaque espèce est représentée par une **fonction d'adéquation à l'environnement** (*fitness function*). La fonction d'adéquation de chaque espèce est couplée aux fonctions d'adéquation des autres

espèces. Le **paysage d'adéquation** (*fitness landscape*) représente la capacité de chaque espèce selon son code génétique à survivre dans son environnement. Lorsque la fonction d'adéquation d'une espèce change, cela modifie à la fois la fonction d'adéquation et le paysage d'adéquation des espèces en coévolution avec elle. Dans le modèle on accepte les mutations d'une espèce, c'est-à-dire les modifications locales, si et seulement si elles permettent d'améliorer la fonction d'adéquation de cette espèce dans le paysage. Ainsi chaque espèce possède une fonction d'adéquation qui correspond à un maximum local du paysage. On suppose que les mutations se font rapidement lorsqu'elles amènent chaque espèce vers un maximum local du paysage. Au contraire les mutations qui font passer une espèce d'un maximum local à un autre maximum local du paysage sont beaucoup plus lentes.

La stabilité de chaque espèce est représentée par un nombre réel, la hauteur d'une barrière à franchir pour passer de son maximum local d'adéquation actuel à un maximum local meilleur. On peut par exemple la mesurer comme fonction de la partie du code génétique à modifier. Plus la modification à effectuer pour améliorer sa fonction d'adéquation est importante, c'est-à-dire plus la barrière est élevée, et plus cette modification apparaît rarement. Dans le modèle on suppose que le temps de mutation dépend exponentiellement de la barrière à franchir. En effet, lorsque l'adéquation est bonne, il est difficile de trouver un maximum local meilleur à proximité.

On considère N espèces reliées entre elles par exemple sur un **graphe aléatoire**, chaque espèce possédant $K - 1$ voisines. Le choix du graphe aléatoire permet d'avoir une approximation en champ moyen du modèle qui est très proche du comportement du modèle lui-même. On peut également choisir de placer ces espèces aux noeuds d'un réseau cubique de dimension d avec des interactions uniquement entre les voisins. À chaque espèce i on associe sa plus petite barrière x_i , les barrières étant choisies initialement selon une distribution uniforme sur l'intervalle $[0 ; 1]$. À chaque pas de temps, le système écologique est modifié comme suit : on choisit l'espèce j dont la barrière est la plus petite et on modifie sa barrière ainsi que celle de chacune de ses $K - 1$ voisines en leur attribuant une nouvelle valeur prise au hasard entre 0 et 1. Ainsi, K espèces ont vu leur barrière modifiée. Il est à remarquer que sur un réseau cubique de dimension d , on a $K = 2d + 1$. Le fait de sélectionner l'espèce ayant la plus petite barrière se justifie grâce à l'hypothèse de dépendance exponentielle du temps entre deux mutations en fonction de la barrière. On classe ensuite les espèces à l'instant t en fonction de la hauteur de leur barrière. On note x_i l'espèce dont la barrière est la $i^{\text{ème}}$ plus petite. On appelle $p(x, t)$ la distribution des barrières des N espèces et $p_i(x, t)$ la distribution de la barrière de l'espèce i .

Per Bak développe ensuite une approximation en champ moyen qui permet d'obtenir l'équation maîtresse que vérifie $p(x, t)$. On peut alors calculer la valeur stationnaire de $p(x)$, c'est-à-dire le point fixe de l'équation maîtresse. De cette manière on peut également calculer $p_1(x)$, distribution stationnaire de l'espèce de plus petite

barrière. Les simulations montrent que l'hypothèse du champ moyen correspond bien aux simulations avec des voisins placés sur un graphe aléatoire, et mieux que les simulations sur un réseau cubique. Avec la dynamique définie précédemment, l'hypothèse du champ moyen permet de montrer qu'il existe un seuil pour $p(x)$ en $x = \frac{1}{K}$. En dessous de ce seuil, $p(x)$ est quasiment constant et égal à $\frac{K}{N}$. Per Bak choisit de prendre un grand nombre d'espèces, donc N est grand. Le nombre K est fixe et égal à trois dans les simulations de Per Bak. En dessous du seuil, on a donc $p(x)$ qui est pratiquement nul et qui tend vers 0 lorsque N tend vers $+\infty$. Lorsque x est au-dessus de ce seuil, $p(x)$ est pratiquement constant et égal à $\frac{K}{K-1}$, c'est-à-dire 1,5 si $K = 3$. On peut tracer la figure montrant la distribution des hauteurs des barrières à l'équilibre $p(x)$ et la distribution de la hauteur de la plus petite barrière $p_1(x)$ pour $K = 3$ et $N = 100$.

On peut alors représenter sur une même figure les résultats de la simulation du modèle avec des voisins choisis aléatoirement et le résultat théorique correspondant dans l'hypothèse du champ moyen. On observe pour $p(x)$ un seuil égal à $\frac{1}{K}$, en dessous duquel p est quasiment nul et au-dessus duquel il est constant et égal à $\frac{K}{K-1}$. Il est remarquable de noter alors que les deux courbes sont extrêmement proches l'une de l'autre. La dynamique amène alors progressivement les espèces dont la barrière est inférieure au seuil, les unes après les autres, mais non de manière déterministe, vers une barrière supérieure au seuil.

Per Bak appelle alors **avalanche** la suite d'événements successifs qui fait intervenir une barrière inférieure au seuil. L'avalanche s'arrête lorsque la plus petite barrière est supérieure à $\frac{1}{K}$. La taille de l'avalanche est définie comme le nombre d'espèces dont la barrière est modifiée au cours de l'avalanche. On peut identifier ce processus de propagation des modifications de barrières à un processus de branchement dont le branchement moyen est 1 (puisque le nombre d'espèces modifiées à chaque pas de temps est K) et dont la probabilité moyenne de modification est $\frac{1}{K}$ (la valeur du seuil). Si N est grand, on observe des avalanches de toute taille, et on dit que ces avalanches sont critiques puisque le processus de branchement l'est. De plus l'ensemble des valeurs des N barrières est transformé et se déplace vers la seule distribution point fixe de la dynamique.

Pour ce point fixe, les avalanches sont à l'état critique, et dans la mesure où la dynamique elle-même du système nous ramène vers l'état critique sans intervention d'un paramètre externe, on peut parler d'état critique auto-organisé pour l'évolution de ce modèle biologique. Ce modèle agit sur un graphe avec des connexions locales. On pourrait attribuer à chaque élément un nombre rationnel, au lieu d'un nombre réel, sans modifier pour autant le comportement global du système. Ceci n'a pas été montré numériquement et resterait à vérifier expérimentalement. On peut donc considérer ce modèle comme un modèle d'automate avec un nombre dénombrable d'états pour chaque site, les sites étant placés sur un réseau défini par un graphe.

Le modèle que nous venons de présenter n'est pas à proprement parler un modèle

d'automate. Il entre cependant tout à fait dans le cadre plus général de Farmer, dans lequel nous voulons pouvoir utiliser les outils de la complexité/quantité définis jusqu'ici. Le seul point qui ne permet pas de parler d'automate est la dynamique sur le graphe. En effet, le graphe change chaque fois qu'une espèce disparaît et qu'une autre apparaît. Or un des principes des automates est que le voisinage du réseau sous-jacent reste fixe. On dépasse donc déjà ici la notion d'automate pour se rapprocher des modèles connexionnistes tels que définis par Farmer. Cependant, notre détour par les automates cellulaires nous a permis de nous familiariser avec la notion d'approximation en champ moyen et d'équation maîtresse. Ces outils ont été de fait bien utiles pour analyser ce modèle.

3.4. L'apport de ces modèles à la compréhension de la complexité/quantité

Il importe de montrer ce qu'a pu apporter l'étude de ces différents modèles à notre compréhension de la complexité. Tout d'abord ils *relèvent bien de la complexité/quantité* telle que nous l'avons définie dans la première section. De plus leur étude nous a permis de *voir comment se passe la mise en pratique des outils théoriques* présentés dans les sections précédentes. Nous avons constaté comment l'observation de lois puissances et l'approximation en champ moyen était utilisée pour étudier des modèles.

La présentation de trois modèles très différents issus de la complexité/quantité met également en évidence la multiplicité des facettes que peut prendre la complexité/quantité. Les modèles et les concepts qui en proviennent s'appliquent à de très nombreux domaines de la science. La multiplicité des aspects que peut présenter la complexité/quantité n'est pas seulement liée au grand nombre des domaines d'application de ces modèles. Un autre aspect de cette multiplicité est lié aux conceptions et aux méthodologies qui sont sous-jacentes et liées à ces modèles. Elles aussi sont extrêmement différentes et variées. Cette diversité de points de vue explique pourquoi *on pourrait croire que la complexité se réduit à la complexité/quantité*.

La diversité des domaines d'application de ces modèles n'est plus à montrer. Nous l'avons déjà illustrée en expliquant par exemple que le concept d'état critique auto-organisé était employé en économie [Bak *et al.*, 1992a], en hydrodynamique, dans le cadre des gaz sur réseau [Andersen *et al.*, 1991][Bussemaker et Ernst, 1993], pour étudier des feux de forêt [Drossel et Schwabl, 1992], des tremblements de terre [Sornette et Sornette, 1989], ou pour des modèles de trafic routier [Biham *et al.*, 1992].

Nous voulons insister ici sur la diversité méthodologique et conceptuelle associée aux trois modèles présentés. Le premier modèle, celui du cadre des gaz sur réseau,

n'a pas pour objectif premier de constituer une modélisation du fluide ou de son comportement. L'objectif initial est de construire des solutions numériques de l'équation de Navier-Stokes. *Les gaz sur réseau constituent donc d'abord une méthode plutôt qu'un modèle.* D'ailleurs le titre du recueil réalisé par Doolen est bien "Méthodes de gaz sur réseau pour les équations aux dérivées partielles" [Doolen et al., 1990]. La modélisation elle-même reste en arrière-plan comme un des aboutissements que peuvent permettre les gaz sur réseau. La relation entre les techniques de gaz sur réseau et les phénomènes de mécanique des fluides intervient lorsque l'on confronte les solutions obtenues aux phénomènes effectivement observés pour des fluides possédant le même nombre de Reynolds et les mêmes viscosités. Par contre la constitution du modèle de gaz sur réseau n'est pas réalisée directement par une étude du fluide associé. L'équation de Navier-Stokes et ses différentes variantes vont jouer systématiquement le rôle d'intermédiaire entre le phénomène et la construction du modèle de gaz sur réseau associé. Finalement on peut dire que la technique de gaz sur réseau participe à la modélisation du fluide, mais qu'elle n'en constitue qu'un maillon. La modélisation en jeu fait appel ici à trois termes : le phénomène de mécanique des fluides étudié, l'équation de Navier-Stokes avec ses spécifications, qui joue le rôle de modèle formel, et le gaz sur réseau, qui joue le rôle d'un schème numérique particulier permettant d'obtenir des solutions approchées des équations de Navier-Stokes, ce schème étant simulable "facilement" sur une machine finie et discrète.

Le deuxième modèle présenté dans cette section était celui du tas de sable, qui a permis la constitution du concept d'état critique auto-organisé. Ce concept est issu d'une observation et d'une mesure sur les simulations effectuées. La distribution des avalanches dans la simulation permet de tracer un graphe log – log par lequel on peut faire passer une droite⁹. La pente de cette droite est alors la valeur du paramètre critique. On peut alors affirmer que l'on a observé une loi puissance et un phénomène critique. Dans ce modèle, ce n'est pas l'existence des phases qui permet l'introduction de la notion de criticalité, contrairement aux modèles magnétiques initiaux. En l'occurrence, c'est l'observation de la loi puissance caractéristique des phénomènes critiques qui justifie le terme de criticalité du modèle et donc la notion de "transition de phase", alors même que les différentes phases n'ont pas été spécifiées et ne sont sans doute pas spécifiables dans ce modèle. *Ainsi, pour ce modèle, c'est la simulation numérique qui est à l'origine du concept abstrait d'état critique auto-organisé, et non un phénomène réel observé qui serait premier.*

⁹On peut se demander si l'observation de si nombreuses lois puissances ne serait pas liée à un nouveau "théorème" très utilisé en physique, mais surtout par ceux qui se contentent de travailler sur des simulations, à savoir : *Par un petit ensemble de points passe une et une seule droite.* Ceci est d'autant plus biaisé que les phénomènes observés ont une certaine "régularité", les mesures et points successifs ne pouvant alors être très éloignés. Le diagramme log-log sur lequel on place les points "rectifie" alors les fluctuations, ce qui permet l'observation d'une droite "caractéristique", avec sa pente caractéristique.

La simulation numérique permet d'introduire le concept et de défendre le modèle en montrant grâce à la langue naturelle que le tas de sable a un comportement similaire.

Dans cet exemple, le terme de modélisation semble un peu abusif. En effet, la relation entre le modèle formel et le phénomène étudié, le tas de sable, reste très distendu. On peut remarquer tout d'abord que des expériences réalisées sur des tas de sable réels ont donné des résultats qui sont loin de correspondre à ceux obtenus à partir du modèle [Jaeger et Nagel, 1992][p. 1387]. On peut défendre cependant le modèle pour différentes raisons. La première consiste à dire qu'un modèle est toujours une idéalisation qui ne rend compte qu'approximativement du phénomène étudié, et de certains aspects seulement. En l'espèce, on dirait par exemple que dans les expériences réelles le sable était trop mouillé et que les grains restaient collés. On constate aussi que différents "sables" donnent des résultats différents. On peut alors considérer le modèle de Per Bak comme un modèle générique qui englobe ces différents tas de sables. *Mais si le modèle de tas de sable ne correspond à aucun tas de sable particulier, peut-on encore dire qu'il y a modélisation ? Le modèle est-il générique de quelque chose ?*

Ceci ne remet pourtant pas en cause la validité du concept d'état critique auto-organisé. Ce concept reste valable dans le cadre des systèmes formels. **Le modèle de tas de sable devient alors un exemple paradigmique d'un modèle formel qui se clôt sur sa simulation numérique et ne renvoie à aucune réalité observée.**

Le troisième modèle correspond lui aussi à une autre conception de la notion de modèle, ces trois conceptions étant possibles dans le cadre de la complexité. Le procédé qui amène à la notion de coévolution est lié presque exclusivement à l'utilisation de la machine et au sens commun. Comme le note Geoffrey Miller [Miller, 1995], les chercheurs qui travaillent sur ce type de modèles invoquent la biologie mais ont une connaissance extrêmement fruste de la discipline. La biologie sert d'argument, de prétexte, de lecture, de "sémantique", d'"illusion nécessaire" à des simulations numériques. On parle ici d'espèces, mais il s'agit de bits, c'est-à-dire d'une succession de 0 et de 1. On peut affirmer qu'il s'agit d'une chaîne de caractères comme les acides aminés sont des caractères. Le langage métaphorique de la biologie est alors réutilisé comme métaphore d'un ensemble de bits. Peut-on parler de multiréalisabilité des métaphores comme on parle de multiréalisabilité des modèles ? La métaphore biologique est prise à la lettre. *Qu'on le veuille ou non, cette conception ne se distingue pas d'une prise de position ontologique symbolique qui affirmerait que le monde est construit sur des éléments logiquement insécables, des atomes au sens propre, et que ces éléments fondateurs sont des symboles.*

Dans le cadre du tas de sable, un concept était dégagé et un modèle formel était proposé. L'articulation, la clôture, se faisait entre le modèle formel et la simulation numérique. *Dans le cas des systèmes de coévolution, il faut constater que la clôture a lieu uniquement autour de la simulation. C'est la simulation elle-même qui est*

le centre sur lequel se focalisent ensuite les études. La simulation est le lieu qui permet de dégager une fois encore les lois puissances observées. L'étude formelle et mathématique est également là pour analyser la simulation et non pour analyser un système formel. L'étude mathématique permet de confirmer avec les outils formels les observations effectuées sur la simulation, qui devient le cœur véritable de l'étude.

Il ne faudrait cependant pas sous-estimer la qualité du travail effectué : réalisation, construction et analyse formelle d'une simulation. Le travail est impeccable si l'on accepte de le considérer uniquement dans le cadre des constructions artificielles effectuées sur un ordinateur. La difficulté consiste à ne pas tirer en retour des conclusions issues d'un monde clos sur une simulation que l'on chercherait ensuite à appliquer au monde des métaphores qui a permis la constitution de ces simulations. Il est clair que ces précautions ne sont pas toujours prises.

Langton, pour ne citer que lui, parle de "la vie au bord du chaos", alors qu'il travaille exclusivement sur des simulations informatiques et n'établit aucun lien concret éventuellement "falsifiable" ou s'appuyant sur des faits biologiques [Langton, 1992]. L'acte fondateur du courant Artificial Life est également significatif à cet égard : "Artificial Life est l'étude de systèmes construits par des hommes qui présentent des comportements caractéristiques des systèmes vivants naturels. Elle complète les sciences biologiques traditionnelles concernées par l'analyse des organismes vivants en essayant de synthétiser des comportements similaires à la vie (*life-like*) à l'intérieur de l'ordinateur et d'autres *media* artificiels. En élargissant la fondation empirique sur laquelle la biologie s'appuie au-delà de la vie à partir de chaînes de carbone comme elle a évolué sur Terre, Artificial Life peut apporter sa contribution à la biologie théorique en situant la vie-telle-qu'elle-est au sein du champ plus large de la vie-telle-qu'elle-pourrait-être" [Langton, 1988][p. 1]. Peut-on caractériser les comportements des systèmes vivants dans l'absolu, sans faire référence au vivant construit sur des "chaînes de carbone" ? Nul ne peut préjuger de ce que sera la science. Il est clair cependant que répondre à cette question positivement est une condition nécessaire et préalable à toute investigation d'une Artificial Life. Or répondre à une telle question nécessiterait une étude fine et poussée de la "vie-telle-qu'elle-est". Cette étude n'a évidemment pas été faite et ne le sera sans doute pas, puisque les travaux se tournent désormais uniquement vers la "vie-telle-qu'elle-pourrait-être" "à l'intérieur de l'ordinateur".

Les simulations se closent sur elles-mêmes et peuvent espérer au mieux une analyse mathématique, comme c'est le cas pour le modèle de coévolution présenté précédemment. *L'étude des simulations est l'étude d'un univers en soi aussi difficile à analyser que peuvent l'être les phénomènes "réels".* On pourrait adapter le propos de J.-P. Dupuy à cette nouvelle situation : "le modèle est tellement plus pur, tellement mieux maîtrisable que le monde des phénomènes : le risque existe qu'il devienne l'objet exclusif de l'attention du savant. Des théories, voire des disciplines entières, peuvent s'organiser autour de l'étude des propriétés d'un

modèle” [Dupuy, 1994][p. 20]. Un renversement analogue à celui observé par J.-P. Dupuy sur les modèles a lieu aujourd’hui dans la communauté anglo-saxonne des Computer Sciences autour *des simulations qui deviennent objet d’étude à part entière, remplaçant tout aussi impénétrable des phénomènes évoqués dont elles doivent rendre compte.* L’ordinateur comme outil prend la même place vis-à-vis des simulations que les mathématiques ont eue pour les modèles formels. On peut cependant noter une différence. La simulation est elle aussi un phénomène, alors que les mathématiques restent des manipulations de symboles abstraits. La tentation est alors encore plus grande de mettre en parallèle, voire d’identifier, la simulation et le phénomène étudié, plutôt que le modèle formel et le phénomène dont il doit rendre compte. La défense d’une position scientifique se confond alors rapidement avec la défense d’une position métaphysique.

En conclusion

Les méthodologies et les stratégies de modélisation mises en place dans le cadre de la complexité/quantité possèdent de nombreux points communs apparents. Elles utilisent dans leur argumentation des théories, des modèles, des simulations, des expériences. Cependant, comme nous l’avons montré, les méthodologies mises en place n’attribuent pas le même rôle à ces différents éléments de justification. En fait, du point de vue de la méthodologie, la complexité/quantité recouvre des conceptions très différentes et souvent antagonistes de ce que doit être une argumentation scientifique.

ANNEXE A : PRÉSENTATION DE QUELQUES DYNAMIQUES DES MODÈLES D'ISING

A.1. Les dynamiques de spin-flip

Les dynamiques de spin-flip, sont exactement toutes les dynamiques individuelles ; on trouve parmi elles les dynamiques de naissance et de mort, qui sont des dynamiques sitistes, donc locales, qui ne dépendent même pas du voisinage, puisque $p(i)$ et $q(i)$ ne dépendent pas de S_j , même si $j \in v(i)$. Pour les dynamiques de naissance et de mort, on a :

$$\begin{cases} \text{si } S_i(t) = 0 \text{ alors } S_i(t+1) = 1 \\ \quad \text{avec la probabilité } p(i) \text{ (taux de naissance)} \\ \text{si } S_i(t) = 1 \text{ alors } S_i(t+1) = 0 \\ \quad \text{avec la probabilité } q(i) \text{ (taux de mort)} \end{cases}$$

A.2. Une dynamique de spin-flip particulière

Une dynamique de spin-flip individuelle, mais ubiquiste, est plus particulièrement liée au modèle d'Ising :

$$S_i(t+1) \neq S_i(t) \text{ avec la probabilité } \exp(-J \sum_{\langle i,j \rangle} S_i S_j)$$

A.3. La dynamique du votant

La dynamique du votant [Liggett, 1985] est une dynamique individuelle et ubiquiste. On se donne pour chaque site i une loi de probabilité p_i sur tous les autres sites du système R :

$$S_i(t+1) \neq S_i(t) \text{ avec la probabilité } \sum_{j \in R} p_i(j) \delta(|S_i(t) - S_j(t)|, 1)$$

où $\delta(a, b)$ désigne le symbole de Kronecker¹. La dynamique du votant consiste donc à être influencé par ceux qui votent autrement².

¹ $\delta(a, b) = 0$ si $a \neq b$ et $\delta(a, b) = 1$ si $a = b$.

²Cela se traduit par $|S_i(t) - S_j(t)| = 1$, dans le cas du modèle d'Ising.

A.4. La dynamique de contact

La dynamique de contact est individuelle et ubiquiste. Soit C un sous-ensemble du système R ; si on suppose que R est invariant par translation³, alors :

$$\begin{cases} \text{si } S_i(t) = 1 \text{ alors } S_i(t+1) = 0 \\ \quad \text{avec la probabilité 1 (guérison sûre avec l'échelle de temps choisie)} \\ \text{si } S_i(t) = 0 \text{ alors } S_i(t+1) = 1 \\ \quad \text{avec la probabilité } \lambda \sum_{j \in C} (i+j) \text{ (propagation de la maladie par contact)} \end{cases}$$

avec $\lambda \in \mathbb{R}$.

A.5. La dynamique de renouvellement

La dynamique de renouvellement est individuelle et ubiquiste :

$$S_i(t+1) \neq S_i(t) \text{ avec la probabilité } \lambda d(i, A_i)$$

avec $\lambda \in \mathbb{R}$ et $d(i, A_i)$ la distance⁴ du site i au sous-ensemble A_i de R qui contient tous les sites j de R pour lesquels $S_i \neq S_j$.

A.6. La dynamique de Kawasaki

La dynamique de Kawasaki est un exemple de dynamique par paquets qui est sitiste. Les paquets en jeu ici comportent deux éléments. Il s'agit d'une dynamique d'“échange” ou de retournement simultané de deux spins qui sont opposés :

$$\begin{aligned} \text{si } S_i(t) \neq S_j(t) \text{ alors } S_i(t+1) &= S_j(t) \text{ et } S_j(t+1) = S_i(t) \\ &\text{avec une probabilité qui ne dépendra que des sites } i \text{ et } j. \end{aligned}$$

³Ce qui est possible dans le cas fini si l'on a un ensemble de sites qui est périodique et si l'on traduit la translation en prenant le résultat modulo la période.

⁴Différentes distances sont évidemment envisageables, certaines étant plus “naturelles” que d'autres.

ANNEXE B : ANALYSE DES VINGT PROBLÈMES DE WOLFRAM

Problème 1 : Quelle classification générale peut-on proposer du comportement des automates cellulaires ?

Pour ce type de classification, une première analyse est généralement expérimentale. La méthode est essentiellement inductive ; elle consiste à rechercher des règles qui amèneront des comportements inattendus.

Par exemple, pour les automates cellulaires en dimension une avec cinq voisins et une configuration initiale aléatoire, on a vu la classification en quatre classes proposée par Wolfram et améliorée par Langton. Cette classification constitue certainement le résultat le plus connu sur les automates cellulaires. Les systèmes dynamiques continus permettent, au moins qualitativement, de trouver des équivalents à ces différentes classes. La classe 1 est alors celle qui évolue vers des points limites, la classe 2 vers des cycles limites. La classe 3 possède un comportement chaotique, analogue à celui des attracteurs étranges. La classe 4, quant à elle, possède de très longues durées transitoires, et l'on ne retrouve pas d'équivalent direct dans la théorie des systèmes dynamiques continus.

Un procédé général permet de proposer une caractérisation quantitative de ces quatre classes. Pour cela on définit une fonction, que l'on appelle "entropie". En calculant la valeur de cette fonction pour une suite de configurations de l'automate, on obtient une mesure quantitative. La définition d'une entropie pour les automates cellulaires n'est pas unique, on peut faire des choix très différents. Cependant toutes les définitions de l'entropie utilisent le même principe. Elles consistent à compter le nombre de suites possibles de valeurs pour chaque site, dans une région de l'espace. On peut également définir l'entropie en terme de probabilité de ces suites de valeurs possibles. Cette entropie peut être définie spatialement ou temporellement. L'entropie est spatiale si l'on considère un ensemble de sites à un moment particulier. Elle est temporelle si l'on considère le même site pendant une certaine durée. On peut ainsi faire une mesure soit à temps fixe sur un ensemble de sites variable, soit sur un ensemble de sites fixe pendant une durée variable.

Quoi qu'il en soit, la classe 1 possède des mesures d'entropie temporelle et d'entropie spatiale nulles. La classe 2 montre une mesure d'entropie temporelle nulle, liée à la périodicité des structures ; son entropie spatiale est positive. Les entropies spatiale et temporelle sont positives pour la classe 3. Lorsque l'on perturbe

la classe 1, les différences observées finissent par disparaître. Pour la classe 2, ces différences persistent, mais elles restent localisées. La classe 3 voit ces différences s'étendre sur l'ensemble de l'automate, avec un taux asymptotique constant. Ce taux d'expansion fournit l'exposant de Lyapounov¹ de l'automate et permet de mesurer la vitesse de propagation de l'automate initial. Lorsqu'on les perturbe, les automates de la classe 4 donnent naissance à des motifs qui se propagent de manière irrégulière dans le temps.

Pour plusieurs raisons, cette description n'est pas encore complètement satisfaisante. En effet certains automates présentent des comportements difficiles à classer. Ainsi, on observe des automates dont différentes régions sont compartimentées. Ces différentes régions évoluent alors indépendamment les unes des autres, sans jamais s'influencer. On observe que chacune de ces différentes régions peut être considérée comme un automate appartenant à l'une des classes. Or dans un même automate des régions différentes se comportent comme des automates de classes différentes. D'un point de vue qualitatif, il est donc possible d'observer des comportements intermédiaires et de définir des sous-classes des quatre classes initiales. Comme nous avons pu le constater, cette classification est avant tout qualitative. Elle s'appuie sur la visualisation des comportements sur un écran d'ordinateur. Le point de vue quantitatif de cette description est possible grâce à la définition d'une mesure relativement arbitraire, en terme d'entropie.

Ces mesures sont donc "expérimentales", et il n'est pas possible dans tous les cas de dire dans quelle classe se trouve un automate simplement à partir de la donnée de ses règles. On utilise le terme d'expérience, mais c'est en fait tout à fait abusif. En effet il s'agit d'une programmation sur ordinateur. Cette "expérience computationnelle" ne fait que dérouler un algorithme parfaitement déterminé. Au contraire, ce que nous appelons expérience habituellement est une perturbation et une exploration de phénomènes qui sont supposés "faire partie du réel" et être indépendants de l'observateur, qui ne fait que les révéler mais ne les crée pas. On peut bien sûr discuter cette conception de l'expérimentation. Il me semble cependant nécessaire de bien différencier les expériences où tout est contrôlé par l'expérimentateur et devient donc son fait, comme c'est le cas pour les "expériences computationnelles", des

¹L'exposant de Lyapounov a été défini initialement pour la solution $x(t) \neq 0$ d'un système linéaire d'équations différentielles ordinaires. L'exposant de Lyapounov vaut alors

$$\lambda_x = \overline{\lim_{t \rightarrow +\infty}} \frac{1}{t} \ln |x(t)|$$

Désormais on considère l'exposant de Lyapounov pour une courbe suffisamment régulière dont on suppose que le comportement à l'infini est équivalent à une exponentielle. On a alors $|x(t)| \sim \exp(\lambda_x t)$, avec λ_x l'exposant de Lyapounov selon la direction ou la trajectoire $x(t)$. En particulier, si λ_x est négatif, alors la courbe s'amortit très fortement et rapidement. Si λ_x est positif, alors on contraire la courbe augmente très rapidement. Seule une valeur de λ_x égale à zéro permet d'éviter des phénomènes d'"explosion" des valeurs vers zéro ou l'infini au cours du temps.

On étudie ensuite la valeur des exposants de Lyapounov suivant chacune des directions possibles.

expériences habituelles. La distinction est importante lorsque l'on souhaite discuter du cerveau. En effet, il semble bien que l'observation de son fonctionnement nous échappe complètement.

Sans essayer d'être exhaustif quant aux arguments qui justifient cette différence, on peut déjà affirmer que lors d'une expérience computationnelle tous les états des automates sont accessibles. Au contraire, lors d'une expérience mettant en jeu un cerveau, même considéré comme un simple système neuronal, nous n'avons pas accès actuellement à l'état particulier de chaque neurone. Plus généralement, lors d'une expérience computationnelle, tous les éléments susceptibles de jouer un rôle dans le comportement global sont définis *a priori*, et les valeurs sont toutes accessibles *a posteriori*. Le propre des expériences habituelles, et de celles sur le cerveau en particulier, est de ne pas nous fournir *a priori* l'ensemble des éléments susceptibles d'avoir leur importance et qu'il faudrait enregistrer, surveiller et mesurer. Étudier le cerveau aujourd'hui, c'est s'intéresser essentiellement aux neurones, et les expériences laissent de côté les cellules gliales ou les hormones. Mais peut-être, demain, serons-nous amenés à observer aussi ces éléments. Expérimenter sur le cerveau et sur le "réel" en général, c'est trancher et choisir. Au contraire, les expériences computationnelles peuvent être effectuées en fournissant à l'"expérimentateur" l'ensemble des données qui ont été effectivement utilisées. On ne peut donc pas mettre sur le même plan et utiliser le même terme pour parler des "expériences computationnelles" et des expériences sur le cerveau.

Les mesures d'entropie proposées sont opérationnelles seulement *a posteriori*, quand on a itérée les règles un nombre de fois important sur un ordinateur, à partir de configurations initiales différentes. On ne possède pas actuellement de justification mathématique et théorique, à partir des règles, de la classification des automates de Wolfram-Langton. Il n'existe pas à l'heure actuelle de classification *a priori* du comportement des automates exclusivement à partir des règles. Nous verrons cependant par la suite la classification en terme de chaînes de Markov proposée par Gutowitz [Gutowitz, 1990]. Elle fournit un début de réponse à cette question.

Problème 2 : Quelles sont les relations exactes entre les différentes entropies et les exposants de Lyapounov des automates cellulaires ?

On peut obtenir des inégalités entre l'entropie et les exposants de Lyapounov. Cependant il semble, à partir des estimations numériques obtenues par simulation, que certaines de ces inégalités soient en fait des égalités. Ceci permettrait d'établir une relation importante entre une mesure statique des automates, l'entropie, et une mesure dynamique, l'exposant de Lyapounov.

Ce travail est délicat car il n'existe pas réellement de méthodes adaptées pour l'étude de l'entropie. L'exposant de Lyapounov, quant à lui, permet de mesurer la

divergence d'une trajectoire dans l'espace des configurations. On peut ainsi définir des exposants de Lyapounov selon différentes directions.

Ces égalités n'ont pas de raison particulière d'être des égalités. Il nous semble qu'en fait les égalités recherchées sont sans doute fausses. Nos estimations numériques sont très proches parce qu'elles sont faites sur un nombre faible d'éléments ou d'itérations. Dans ces conditions, les mesures des exposants de Lyapounov et de l'entropie n'ont pas encore eu le temps de se distinguer. Ceci proviendrait du fait que les exposants de Lyapounov sont calculés comme les exposants d'une puissance et que leur évolution en fonction du nombre d'itérations est donc logarithmique. Il est très difficile de distinguer numériquement deux nombres proches en faible dimension et pour un nombre faible d'itérations. C'est le cas de l'exposant de Lyapounov et de l'entropie.

Problème 3 : Quel est l'analogie de la géométrie pour l'espace de configuration d'un automate cellulaire ?

La question posée ici porte sur la géométrie de l'espace de configuration et non pas sur celle de l'espace des voisinages qui, elle, est fixée au départ. La première chose à remarquer est que les exposants de Lyapounov, définis selon différentes directions du réseau carré, sont très proches. Ceci est vrai même lorsque les règles d'évolution sont asymétriques. Ce point conforte un peu plus notre point de vue sur le Problème 2. Il semble que l'évolution des automates cellulaires introduit une forme de géométrie dans l'espace des configurations. Cependant les détails restent obscurs. On ne sait pas par exemple quel est l'équivalent de l'espace tangent pour les systèmes dynamiques continus.

Problème 4 : Quelles grandeurs statistiques caractérisent le comportement d'un automate cellulaire ?

On peut faire différents types de mesures, comme des calculs de densité des sites ou blocs de sites possédant une valeur particulière. D'autres mesures statistiques peuvent être faites à partir de la fonction de corrélation ou d'auto-corrélation, pour décrire l'interdépendance des états entre différents sites. Nous avons déjà discuté ces mesures à propos du Problème 1.

La transformation de Fourier et l'analyse spectrale permettent également d'obtenir des informations sur le système. Une analyse complètement discrète serait d'ailleurs plus adaptée. Les formes obtenues en fonction du comportement des automates ne sont pas connues actuellement.

En dehors de l'entropie et des exposants de Lyapounov, la théorie des systèmes dynamiques suggère que la fonction ζ pourrait fournir une caractérisation du comportement global des automates cellulaires. Elle mesure la densité des suites périodiques dans les configurations d'un automate, et elle pourrait être reliée à la

transformée de Fourier. Il n'est pas vraiment étonnant que les fonctions spéciales soient d'un large intérêt pour les automates, puisque Wolfram a montré certaines relations que l'on peut obtenir aisément en une dimension avec le triangle de Pascal.

D'autres travaux que nous avons pu mener dans cette direction montrent également que les fonctions spéciales interviennent naturellement dans d'autres types de problèmes apparemment sans rapport, comme la recherche des équipotentielles électriques dans un hypercube en dimension n , où tous les points sont reliés par des résistances identiques. Le problème formel se traduit ensuite dans des termes identiques.

Problème 5 : Quels sont les invariants lors de l'évolution d'un automate cellulaire ?

La recherche d'invariants au cours de la dynamique est un des objectifs principaux de l'étude des automates cellulaires. On retrouve cette investigation dans l'étude des gaz sur réseau et plus généralement dans toute modélisation physique qui nécessite un comportement dynamique. Dans ce type de problèmes, les spécifications et les contraintes physiques sont données par des hypothèses formulées sous forme d'invariance ou de conservation de certaines quantités².

Ces invariances peuvent être spatiales, comme les hypothèses d'isotropie, ou temporelles, comme la conservation de la "charge" au cours du temps. Elles peuvent également être spatio-temporelles, comme la conservation des invariants galiléens.

La recherche d'invariants à partir de simulations est très difficile en général, que ce soit pour un automate ou un système dynamique non linéaire. On peut facilement identifier certains invariants pour les automates des classes 1 et 2. Dans ce cas le problème est tronqué, car les invariants sont des points ou des cycles limites. La difficulté est de réussir à obtenir des invariants qui soient eux-mêmes des états dynamiques et ne constituent pas tout l'automate comme c'est le cas pour les cycles limites. Lorsque l'on recherche d'autres invariants, le but est de trouver des règles *ad hoc*. Il est possible de trouver des partitionnements de l'espace qui restent invariants lors de l'évolution de l'automate³. On peut donner d'autres invariants dynamiques⁴

²Pour les gaz sur réseaux, on aura l'hypothèse d'invariance galiléenne, ainsi que la conservation du nombre de particules et du moment.

³On l'observe par exemple pour un automate en dimension une avec trois états par site et des voisinages à trois éléments, pour la règle 00000000000001011002010010. Pour cette règle on observe des "membranes" entre régions différentes, c'est-à-dire une partition des sites de l'automate pour laquelle les comportements vont rester séparés, certaines régions ne comportant que des 0 et des 1 et d'autres que des 0 et des 2 par exemple.

⁴Le jeu de la vie a été proposé par J. Conway en 1970, dans un article non publié. Il a été diffusé et popularisé par Martin Gardner [Gardner, 1970] [Gardner, 1971b] [Gardner, 1971c] [Gardner, 1971a] [Gardner, 1972].

Le jeu de la vie est un automate à deux états sur un réseau carré, en dimension deux. Les

comme les “gliders⁵” pour le jeu de la vie.

Problème 6 : Comment la thermodynamique s’applique-t-elle aux automates cellulaires ?

La thermodynamique décrit le comportement général moyen d’un système comportant de très nombreux éléments. Tous les automates cellulaires vérifient cette première propriété. De plus les dynamiques microscopiques de ces systèmes doivent être réversibles pour que les hypothèses soient similaires à celles de la thermodynamique. Seuls les automates dont les règles sont bijectives ou réversibles⁶ peuvent donc être étudiés dans le même cadre que la thermodynamique.

On connaît toutes les règles réversibles pour les automates en une dimension lorsque les états des sites peuvent prendre deux valeurs et que l’on possède moins de cinq voisins, ou lorsque les états peuvent prendre trois valeurs et que le voisinage est réduit à trois valeurs⁷. Cependant on ne connaît pas actuellement de procédure qui permette de déterminer toutes les règles réversibles, même s’il est possible de construire des règles particulières qui le soient.

Pour se situer dans le cadre de la thermodynamique, il faut également regrouper de nombreux sites dans des états différents afin de mimer l’effet de mesures imprécises. Cette procédure peut être opérée par une règle d’automate, irréversible cette fois. De plus l’entropie de ces règles doit suivre un équivalent du second principe de la thermodynamique.

Problème 7 : Comment les différents comportements sont-ils distribués dans l'espace des règles d'un automate cellulaire ?

Ce problème reste très lié à l’étude du comportement des automates et à leur classification. En effet, toute classification induit de manière implicite la constitution de relations de proximité entre les éléments. La classification des comportements de l’automate permet de structurer l’espace des règles. Le problème consiste alors à

voisinages comportent neuf éléments, ils sont donc du type de Moore. Les règles de l’automate sont simples : (i) si un site possède deux voisins dont l’état est 1, son état reste inchangé ; (ii) si un site possède trois voisins dont l’état est 1, son état passe à 1 ; (iii) sinon le site passe à l’état 0.

On dit qu’un site est actif, ou “vivant”, s’il est à l’état 1, et inactif, ou “mort”, s’il est à l’état 0.

⁵Le glider est un ensemble dynamique comportant cinq activations potentielles. Cet ensemble d’activations se manifeste de site en site et donne visuellement l’impression de constituer un élément autonome qui se déplace sur le réseau en dimension deux. De plus cet ensemble dynamique considéré comme autonome suit lui-même une période propre égale à deux, avec ses cinq éléments propres qui sont actifs ou non actifs.

⁶On dit qu’un automate possède une règle réversible si la connaissance de l’état des sites de l’automate après application des règles permet de retrouver un unique état des sites de l’automate antérieur à l’application des règles.

⁷Cette étude a été menée par Wolfram dans [Wolfram, 1983] et [Martin *et al.*, 1984].

étudier cette structure et à voir comment elle se retrouve dans la structure des règles des automates. La question que pose Wolfram est donc de savoir comment on peut mettre en parallèle la structure des comportements observés computationnellement et celle des règles.

La question de fond qui se pose est de savoir d'abord quelle est la nature structurelle qu'il faut attribuer aux différents espaces intervenant dans les problèmes d'automates. Quelle nature doit-on par exemple attribuer à l'espace des règles, celle d'un espace topologique, d'un espace métrique, d'un espace normé, ou d'une autre structure encore ? La question est de savoir comment il faut structurer cet espace pour que le passage de l'espace des règles à l'espace des comportements des automates, c'est-à-dire à l'espace des états des sites dans leur succession, permette la conservation de certaines propriétés topologiques, métriques ou normées.

Les premiers résultats obtenus sont d'ordre empirique et proviennent de différentes simulations. Il semble que les classes 1 et 2 de Wolfram deviennent de moins en moins fréquentes pour les automates en une dimension lorsque le nombre d'états possibles pour chaque site augmente, ou lorsque la taille du voisinage augmente. Il en est de même pour la classe 4. Au contraire la classe 3 devient plus importante. Dans son article [Wolfram, 1984c], Wolfram propose un tableau rassemblant la proportion des différentes classes, en dimension une, en fonction du nombre de voisins $2r + 1$ et du nombre d'états possibles par site k :

classes	$k = 2$ et $r = 1$	$k = 2$ et $r = 2$	$k = 2$ et $r = 3$	$k = 3$ et $r = 1$
classe 1	0,50	0,25	0,09	0,12
classe 2	0,25	0,16	0,11	0,19
classe 3	0,25	0,53	0,73	0,60
classe 4	0	0,006	0,06	0,07

Le problème général est très difficile à aborder d'un point de vue théorique, comme c'est généralement le cas lorsque la structure sous-jacente est en jeu. En effet il existe plusieurs moyens de définir une distance dans l'espace des règles d'un automate cellulaire, même si la plus répandue est la distance de Hamming. Certaines règles peuvent être proches au sens de cette distance et exhiber des comportements tout à fait différents. Pour étudier certains comportements, la définition des règles en fonction d'un paramètre continu permet une étude plus fine. C'est en particulier ce que permettent les travaux de Langton [Langton, 1990] avec l'introduction du paramètre λ .

Problème 8 : Quelles sont les propriétés d'échelle des automates cellulaires ?

Cette question est importante pour celui qui souhaite utiliser les automates cellulaires. Le premier enjeu est de savoir quand il est possible d'observer certaines mesures invariantes d'échelle pour l'automate. Ceci s'observe, en particulier, si l'on

peut définir une dimension, éventuellement non entière, de la figure qui représente au cours du temps le comportement de l'automate (figure 6). Il est alors possible de faire certaines mesures sur l'automate qui suivent des lois puissances. L'habitude est alors de parler d'état critique lorsque l'on observe ces lois puissances. Le raisonnement est inverse du raisonnement initial en physique. Au départ on observait des lois puissances pour les phénomènes critiques. Désormais on considère que l'observation d'une loi puissance est le signe de présence d'une transition de phase, avant même d'avoir défini quelles étaient les différentes phases en jeu.

Le second enjeu porte sur la possibilité de transformer l'automate étudié par changement d'échelle en regroupant par paquets certains sites en un nouveau site de taille plus grande. Il faut alors être capable de déterminer les règles de ce nouvel automate pour que la mesure effectuée sur lui soit la même que celle qui aurait été effectuée sur l'automate plus petit en regroupant les valeurs des sites. Nous appellerons renormalisation ce second problème. Nous l'avons déjà dans le cadre des modèles de type Ising. On sait que, dans les conditions de criticalité, il est possible d'effectuer ces transformations, c'est-à-dire de renormaliser le système. Or, comme nous l'avons dit, on considère désormais que les conditions de criticalité sont remplies si l'on observe des lois puissances.

Le but est de renormaliser l'automate. Aussi l'observation de fractales ou de lois puissances est une étape intermédiaire. Elle doit permettre de trouver comment il faut transformer les règles au cours de la renormalisation. Certaines règles des automates en une dimension sont connues pour créer des motifs invariants d'échelle, comme les règles 90 et 150 dans la description de Wolfram (figures 5 et 6).

Cependant il semble que, lorsque Wolfram a posé son problème 8, il n'avait pas étudié en détail la situation en physique de la matière condensée. En effet il semble un peu confondre la notion de renormalisation et celle de fractales. Il ne parle jamais d'état critique, même si le mot "phase" est plusieurs fois prononcé. De plus Wolfram n'explique pas pourquoi ce qu'il définit comme une phase pour les automates serait l'équivalent d'une phase en physique. Certains articles, comme ceux de Langton [Langton, 1990], s'intéressent au problème de la définition des phases d'un automate. Cependant, même si implicitement cela constitue un des centres d'intérêt de la communauté, je n'ai pas trouvé d'article actuel traitant explicitement de la renormalisation des automates.

On peut développer cette question en insistant sur le fait que le problème est de savoir comment la quantité se répartit spatialement. Plus précisément, on peut se demander, pour un automate à deux états, comment les sites dont l'état vaut 1 se distribuent dans l'espace de l'automate. Les travaux de Takahashi, en utilisant la notion de multifractales, permettent d'aller dans ce sens [Takahashi, 1990].

Dans son article de 1990 [Takahashi, 1990], Takahashi réalise une analyse multifractale des règles 90 et 150. Pour cela il définit des outils théoriques permettant

de réaliser une analyse computationnelle des caractéristiques multifractales des automates. Il fournit également quelques résultats théoriques, puis il les applique à d'autres règles.

Takahashi reprend d'abord la définition proposée par Willson de la matrice de transition d'un automate. La dimension de Hausdorff est alors obtenue à partir de la plus grande valeur propre de cette matrice. On définit ensuite la dimension spectrale $H(\delta)$ d'un ensemble limite issu d'un automate. Soit T le temps supposé fini au bout duquel on atteint cet ensemble limite. On divise alors la variable qui correspond au nombre d'itérations par T pour la normaliser. La dimension $H(\delta)$ est alors la dimension de l'ensemble des instants τ normalisés au cours desquels la configuration de l'automate a eu la dimension d'Hausdorff δ . Takahashi fournit également une méthode numérique pour calculer $H(\delta)$.

Takahashi explique comment on peut travailler sur les aspects multifractals d'un automate. Il rappelle la définition du spectre multifractal $f(\alpha)$ qui représente la distribution d'une mesure sur un multifractal. Ce spectre permet non seulement de connaître la dimension d'un objet fractal, mais aussi la répartition spatiale des différents éléments constituant cet objet fractal. On affecte un poids ou une mesure à chacun des éléments constituant l'objet étudié. Le spectre permet d'obtenir la dimension des différents objets définis par les moments d'ordre successif de l'objet de départ. Si, quel que soit l'ordre du moment, on obtient encore un objet fractal, on parle d'objet multifractal et on peut calculer le spectre multifractal. Takahashi applique ensuite une méthode usuelle en analyse multifractale à l'analyse des règles 90 et 150.

Problème 9 : Quelle correspondance peut-on établir entre les automates cellulaires et les systèmes continus ?

L'intérêt premier des automates cellulaires est de permettre des simulations rapides, parallèles et sans erreurs d'arrondis dues à la manipulation de nombres réels⁸. C'est ainsi l'aspect discret des automates qui est mis en avant.

Cependant, une autre façon de travailler avec des éléments discrets consiste tout simplement à discréteriser un problème continu. Ainsi, on peut discréteriser les

⁸Attention, ceci devient faux si une énergie doit être mesurée dont les valeurs peuvent être réelles. C'est le cas par exemple pour les modèles d'Ising et de Heisenberg où l'énergie calculée influence ensuite l'ensemble du système.

Les manipulations exactes avec des nombres rationnels sont possibles en général. Cependant les fonctions transcendantes qui apparaissent naturellement en physique, et en particulier pour le calcul de la fonction de partition et de l'énergie, ne permettent pas dans le cas des modèles issus de la physique de faire tous les calculs de façon exacte avec des rationnels.

Dans le cadre de modèles ou d'algorithmes qui ne subiraient pas les mêmes contraintes que les modèles physiques, on peut par contre envisager des formulations complètes en termes rationnels, sans fonctions transcendantes, et qui ne nécessiteraient que des calculs, sans aucune approximation.

équations différentielles régissant un système dynamique, afin de simuler son évolution sur un ordinateur. Le cœur du problème devient alors le lien qui peut exister entre un système dynamique continu et que l'on discrétise et un système dynamique essentiellement discret, comme c'est le cas des automates cellulaires. Les problèmes formulés en termes continus peuvent être approchés par des constructions arithmétiques. De plus les ordinateurs symboliques dont les opérations élémentaires sont de type arithmétique ne permettent de travailler directement que sur l'arithmétique. De ce fait, il faut reformuler les problèmes continus directement en termes arithmétiques. La question est alors de savoir comment les automates traitent cette formulation arithmétique des problèmes continus et si les solutions obtenues correspondent aux solutions théoriques qui auraient pu être obtenues pour le problème continu initial. Pour cela, on peut essayer de résoudre directement des équations différentielles formulées en termes arithmétiques, à l'aide d'automates cellulaires⁹.

Une autre façon de travailler consiste à regarder si la discrétisation des équations différentielles nous ramène à des automates cellulaires. Or comme les automates sont discrets en espace et en temps, on peut écrire une formulation en termes de différences finies des équations différentielles. Si le problème est stable, on obtient des approximations qui s'améliorent quand on affine la discrétisation. Ceci nous amène à utiliser l'approximation de Jacobi, dans les cas les plus simples, plutôt que celle de Gauss-Seidel.

Le fait de travailler sans erreurs cumulées sur les réels n'a pas encore été totalement approfondi. On ne sait pas exactement comment cela modifie l'instabilité des équations discrétisées, par rapport à la solution théorique.

Il est difficile de traduire réciproquement en termes continus tous les problèmes liés aux phénomènes à longue distance et à l'invariance d'échelle. Finalement, cette question cruciale pour justifier l'intérêt théorique des automates cellulaires n'est pas du tout résolue actuellement¹⁰. Nous avons déjà insisté par contre sur l'intérêt pratique des automates.

Problème 10 : Quelle correspondance peut-on établir entre les automates cellulaires et les systèmes stochastiques ?

Comme nous l'avons déjà fait remarquer, l'apprentissage par renforcement fonctionne bien lorsque l'on peut faire des approximations des probabilités de transition d'un état à un autre en termes de chaînes de Markov. La relation qui peut exister entre automates et systèmes stochastiques devient alors essentielle. On retrouve ceci

⁹On peut comprendre dans ce sens les méthodes de gaz sur réseau, comme nous le voyons dans la section 3.

¹⁰Cette question impose implicitement de retourner à une étude des fondements et des paradoxes de la théorie des ensembles.

en particulier dans la classification de Gutowitz présentée et commentée en annexe 3. En fait, même si les automates suivent des règles déterministes, du fait de la configuration initiale aléatoire, leur comportement peut être très étroitement relié à celui de systèmes dynamiques habituellement décrits en termes de bruit ou de probabilités. Comme en physique, par exemple dans le cadre du modèle d'Ising, on obtient des règles déterministes. Cependant, une bonne description du comportement du modèle d'Ising se fait en termes probabilistes, comme pour les automates cellulaires. Wolfram ne parle pas du tout des automates stochastiques, pour lesquels les règles suivent des lois de probabilité. Nous tâcherons de montrer par la suite que les automates stochastiques semblent nécessaires pour résoudre des problèmes d'apprentissage dans un monde bruité. Ceci se retrouve par exemple dans l'apprentissage par renforcement.

Problème 11 : Comment les automates cellulaires sont-ils modifiés par le bruit et d'autres types d'imperfections ?

La question qui se pose ici est de savoir si le comportement de l'automate va être ou non fortement affecté par une perturbation faible de ses règles ou de l'état initial de l'automate. Cette question est intrinsèquement liée à la structure des règles et à celle de l'état de l'automate. Cet aspect est crucial et on sait qu'il se trouve également au cœur des problèmes qui se posent dans l'utilisation des algorithmes génétiques. Le problème 11 est donc relativement proche des problèmes 1 et 7.

Les imperfections peuvent être de différents types. Pour un automate déterministe, ce peut être la donnée d'un aspect stochastique de certaines règles. Ainsi, on peut utiliser deux règles déterministes, la première avec la probabilité p et la seconde avec la probabilité $1 - p$.

On peut également avoir des inhomogénéités dans le réseau, ou encore avoir des règles différentes selon les sites du réseau. On lève alors l'isotropie présente dans les automates cellulaires finis et déterministes habituels. Si on introduit des interactions non locales et des remises à jour asynchrones, on obtient des automates analogues aux réseaux booléens.

Un bruit, même faible, peut entraîner des modifications considérables du comportement. Jusqu'à une valeur seuil, les simulations de Wolfram ne permettent d'observer que de faibles modifications du comportement. Au-delà, le comportement change radicalement et brutalement. Aussi, on peut penser que des transitions de phase sont en jeu dans ces phénomènes.

Problème 12 : Pour les automates cellulaires à une dimension, la complexité générique du langage rationnel est-elle non décroissante avec le temps ?

L'ensemble des configurations générées lors de l'évolution d'un automate cellulaire, partant de toutes les configurations initiales possibles, peut être considéré comme

un langage formel. Chaque configuration correspond à un mot du langage obéissant à un ensemble défini de règles grammaticales. Ce mot est formé d'une suite de symboles représentés par l'état en chaque site. L'ensemble des configurations générées après un temps fini forme un langage formel rationnel¹¹, comme l'a montré Wolfram [Wolfram, 1984a]. On peut réaliser une description en terme de graphe de ce langage. Pour un langage rationnel, le nombre de sites du graphe minimal est un nombre caractéristique, qui peut servir par exemple à mesurer la complexité de ce langage. Plus ce nombre est grand, plus le langage est complexe.

Les études empiriques semblent annoncer que, pour les classes 1 et 2, ce nombre augmente lentement et de manière au plus polynômiale en fonction du temps. Pour les classes 3 et 4, les simulations montrent une croissance rapide de ce nombre et il semble qu'il ne peut pratiquement que croître en fonction du temps. Wolfram pense que cette propriété devrait pouvoir se démontrer relativement facilement, uniquement à partir de quelques propriétés des automates.

Dans ce cas, cela donnerait un aspect quantitatif à l'affirmation selon laquelle la complexité serait une fonction croissante du temps. Wolfram propose de considérer ceci comme un équivalent du second principe de la thermodynamique.

Problème 13 : Quels ensembles limites peuvent produire les automates cellulaires ?

On ne peut pas obtenir, avec un automate cellulaire, toutes les configurations comme configurations limites au bout d'un temps infini. Le nombre de configurations totales possibles est non dénombrable, si le nombre de règles est infini dénombrable. À chaque pas de temps de l'évolution d'un automate avec une règle irréversible, une nouvelle configuration est atteinte. Celle-ci ne peut plus alors être atteinte comme configuration limite. Les configurations limites sont donc constituées par l'ensemble des configurations qui ne sont jamais exclues. L'ensemble de toutes les configurations exclues est récursivement énumérable¹², car tous ses éléments sont obtenus après un nombre fini d'itérations. On peut donc en déduire que les ensembles limites font partie du complémentaire dans l'espace des configurations possibles d'ensembles récursivement énumérables. De plus, pour être un ensemble limite, il faut que ces ensembles soient invariants par translation. Les ensembles limites doivent donc contenir soit un seul élément, soit une infinité.

Pour certaines règles, qui sont dans les classes 1 ou 2, l'ensemble limite forme un langage rationnel. Par contre, pour les autres classes, on obtient des langages formels plus complexes. On connaît des exemples explicites pour lesquels ces ensembles

¹¹On se trouve ici, en fait, dans un cas particulier du théorème d'équivalence entre les langages rationnels et les automates finis [Stern, 1990][p. 27].

¹²On dit qu'un ensemble A est récursivement énumérable s'il existe un système formel correct et complet pour les énoncés de la forme “ n est dans A ”[Delahaye, 1994][p. 74].

limites sont des langages dépendants ou indépendants du contexte. De plus certains ensembles limites peuvent être des ensembles non récursifs.

On peut s'interroger aussi sur les propriétés de l'ensemble limite, en fonction des propriétés de la règle correspondante. Il est vraisemblable qu'un certain nombre des propriétés des ensembles limites est indécidable. On peut sans doute aussi construire des automates cellulaires dont les ensembles limites vérifient certaines propriétés données *a priori*. Ils pourraient alors être utilisés comme filtres. Cependant, ceci n'est sans doute possible de façon explicite que dans des cas très particuliers, car il semble difficile de déterminer les automates associés à des ensembles limites à partir de la simple connaissance de ces ensembles limites ou de certaines de leurs propriétés. Il s'agit là d'un problème inverse.

Problème 14 : Quelles relations peut-on établir entre les caractéristiques computationnelles et statistiques des automates cellulaires ?

Le taux d'information transmise affecte à la fois les propriétés computationnelles et statistiques des automates cellulaires. Il permet d'établir les valeurs des exposants de Lyapounov de l'évolution de l'automate cellulaire.

Pour les classes 1 et 2, ces exposants sont nuls, ce qui assure que l'information reste localisée. Ces classes font alors apparaître des langages rationnels. Les configurations peuvent alors être générées par un processus de Markov, pour lequel il n'y a pas de corrélation à longue distance entre les différentes parties de l'automate.

Pour les classes 3 et 4, les exposants de Lyapounov sont positifs, et une légère modification initiale se développe par la suite. Les valeurs en un site particulier dépendent alors d'une région dont la taille est croissante au cours du temps en fonction de l'état initial. Wolfram considère que la classe 3 amène à des langages sensibles au contexte, alors que la classe 4 fait apparaître des langages formels en général non récursifs. Or il semble que le type de langage observé permet de savoir quelles sont les valeurs possibles de l'entropie.

Les automates en deux dimensions avec des exposants de Lyapounov positifs dans toutes les directions génèrent des ensembles de configurations non récursifs, même au bout d'un temps fini. En particulier, le problème qui consiste à trouver des configurations qui évoluent de manière périodique, en dimension deux, est équivalent au problème du pavage avec des dominos, lequel est connu pour être indécidable.

Problème 15 : Dans quelle mesure les suites générées par les automates cellulaires sont-elles aléatoires ?

On ne sait pas caractériser le comportement général des suites d'états apparaissant en un site au cours du temps. Pour les automates de la classe 4, on sait

qu'elles peuvent former un ensemble non récursif. La question est de savoir comment on peut reconstruire un état initial, à partir de la connaissance des séries temporelles de quelques sites. Ce problème, quoique soluble, est tellement difficile que les automates cellulaires se présentent comme de bons candidats pour créer des suites de nombres pseudo-aléatoires et sont en utilisés en cryptographie [Wolfram, 1986c][Wolfram, 1986b].

Certains automates, même partant d'un seul état initial non nul, réussissent à franchir de nombreux tests qui assurent l'aspect pseudo-aléatoire d'une suite. Par exemple les systèmes montrant un comportement chaotique nécessitent une infinie précision de leur configuration initiale. Ce n'est pas le cas pour les automates qui peuvent avoir un comportement chaotique simplement par la nature de leur dynamique. Celle-ci consiste à progressivement mettre au jour les conditions initiales. Il est vraisemblable que l'aspect aléatoire¹³ des nombres générés par les automates cellulaires est du même type que la suite des décimales de π .

Problème 16 : Quelle est la fréquence de la computation universelle et des problèmes d'indécidabilité rencontrés avec les automates cellulaires ?

Si un système est capable de computation universelle, son évolution peut faire apparaître tout processus computationnel fini, si on lui donne des conditions initiales appropriées. On connaît plusieurs exemples d'automates équivalents aux machines de Turing. C'est le cas du "jeu de la vie". En dimension une, ceci a aussi été montré pour des automates pouvant avoir jusqu'à 18 valeurs par site.

Wolfram et Langton pensent que les automates identifiés de manière statistique comme étant de la classe 4 sont en fait capables de computation universelle. Cela prouverait que l'on peut obtenir un automate capable de computation universelle avec seulement deux valeurs par site et un voisinage comportant cinq sites, ou trois valeurs par site et un voisinage comportant trois sites. Cependant il reste à prouver au moins pour une de ces règles que l'automate correspondant est capable de computation universelle. Plusieurs méthodes sont susceptibles d'amener au résultat. Une méthode consiste à retrouver les éléments de la machine de Turing. Cependant, pour cette méthode, une des difficultés est liée au fait qu'actuellement on n'a pas trouvé d'élément correspondant à une horloge et qui fournirait une suite infinie de signaux.

Si effectivement la classe 4 est capable de computation universelle, elle est assez fréquente en une dimension. Par contre elle est beaucoup plus rare en deux dimensions, le "jeu de la vie" étant un des seuls exemples connus actuellement. Il

¹³Ceci mériterait d'être approfondi en fonction des définitions actuelles de l'aléatoire et de sa distinction avec le calculable [Delahaye, 1994]. En effet, le nombre π étant calculable, il n'est plus considéré comme aléatoire à l'heure actuelle.

est également possible que cette computation universelle apparaisse aussi pour des automates de la classe 3, mais seulement pour quelques configurations initiales, de mesure nulle dans l'espace des configurations. Dans ce cas l'ensemble des automates capables de computation universelle serait beaucoup plus large.

Il faut noter que de nombreux problèmes indécidables apparaissent, même lorsque l'on a un système qui n'est pas capable de computation universelle complète. Le théorème de Rice nous dit en particulier que presque tous les énoncés à propos d'un ensemble récursivement énumérable sont indécidables [Rice, 1953]. Un résultat dû à Kari [Kari, 1990] nous dit également que la question générique de la réversibilité de la règle d'un automate en dimension deux est indécidable.

Problème 17 : Quelle est la nature des ensembles limites des automates lorsque la taille de ces automates devient infinie ?

Il est vraisemblable que lorsque la longueur de la période des automates augmente, des processus de plus en plus compliqués apparaissent, comme la capacité d'autoreproduction. Dans ce cas les mesures qui ont été faites pour de petits automates sont remises en cause. En effet, les phénomènes observés et prédominants pour les automates de petite taille seraient en nombre très faible pour des automates de taille infinie. Ils seraient négligeables devant les phénomènes nouveaux et plus complexes.

Ceci nous entraîne vers des problèmes déjà rencontrés. En particulier, trouver la taille de la plus petite structure d'un automate ayant une propriété donnée, comme sa capacité d'autoreproduction, est un problème indécidable. On retrouve de nouveau une question que l'on se pose naturellement dans les termes de la quantité. De plus l'hypothèse de périodicité a permis de laisser de côté l'infini pour travailler sur la quantité à partir d'un nombre fini d'éléments.

Problème 18 : Quelle est la fréquence de l'irréductibilité computationnelle des automates cellulaires ?

On parle d'irréductibilité computationnelle lorsque le seul moyen de connaître l'évolution d'un automate consiste à itérer les règles de cet automate. Ce moyen est toujours possible, mais la question consiste à savoir quand et dans quelle mesure on peut l'éviter. Il est alors possible de compacter le fonctionnement de l'automate et on mesure la complexité/quantité de l'automate en donnant la taille minimale du compactage. Il est parfois possible de traiter le problème de manière algébrique pour obtenir directement l'état d'un site particulier, à un instant donné. De telles méthodes ne peuvent pas être obtenues de manière systématique pour des automates capables de computation universelle. En fait, il semble que ce critère ne soit pas pertinent pour déterminer le caractère irréductible des automates¹⁴.

¹⁴Ces aspects pourraient être repris plus précisément dans le cadre de la "deuxième théorie

Des critères plus précis d'irréductibilité peuvent être obtenus dans le cadre de la théorie de la complexité computationnelle. On pense que les automates des classes 1 et 2 sont réductibles, alors que l'évolution des automates des classes 3 et 4 ne le serait pas. Il faut tout de même noter que l'irréductibilité d'un automate n'interdit pas l'existence éventuelle de certaines procédures rapides permettant d'obtenir des propriétés générales de cet automate.

Problème 19 : Quelle est la fréquence des problèmes sur les automates qui ne peuvent être traités informatiquement ?

La question consiste par exemple à savoir au bout de combien de temps une suite d'états va apparaître en un site, en partant d'un état donné. Si un algorithme peut répondre à cette question, il sera NP complet¹⁵. Le problème¹⁶ deviendra alors rapidement intractable en temps sur un ordinateur. Wolfram considère que le problème est computationnellement analogue à un problème indécidable. Certains problèmes sur les automates sont équivalents à des problèmes NP complets. Wolfram pense que les problèmes NP complets sont fréquents à propos des classes 3 et 4.

Problème 20 : Quelles descriptions de plus haut niveau peut-on donner à propos de l'information traitée par les automates cellulaires ?

Le but est d'organiser les informations fournies par l'automate afin de pouvoir travailler à un niveau supérieur avec des assemblées d'automates. Or les solutions proposées sont généralement sérielles, alors que les automates cellulaires travaillent fondamentalement de manière distribuée et parallèle. Une autre approche pourrait être statistique. Elle consiste à diviser et décrire les attracteurs de l'évolution globale du système. Cette approche est tout à fait valable lorsque les configurations initiales dans un même bassin tendent vers un même attracteur au cours de la dynamique. Ceci est possible quand les bassins d'attraction sont locaux dans l'espace, comme c'est le cas en reconnaissance d'images.

Cependant, lorsque cette caractéristique locale n'est pas vérifiée, le problème s'avère beaucoup plus délicat. Il semble, quoi qu'il arrive, que pour résoudre ce

algorithme de la complexité” [Delahaye, 1994]. En particulier il faudrait discuter les relations entre l'incompressibilité au sens de Kolmogoroff et l'irréductibilité dont parle Wolfram.

¹⁵Un ensemble est NP s'il existe une machine de Turing non déterministe (ou machine de Turing à choix) qui décide l'ensemble (c'est-à-dire permet d'affirmer si les éléments, c'est-à-dire les mots de l'ensemble, font ou non partie du langage) et si cette machine est telle que tout calcul à partir d'un mot de longueur n assez grande a une longueur au plus égale à $P(n)$ avec P un polynôme [Dehornoy, 1993].

Un ensemble est NP complet s'il est NP et si tout problème NP peut être ramené à ce problème en un temps polynômial. Un ensemble est NP complet s'il est au moins “aussi compliqué” que tout autre problème NP.

¹⁶On passe d'un ensemble NP à un problème NP en disant que le problème est NP si l'ensemble de ses solutions est NP.

problème une approche radicalement nouvelle soit nécessaire. On peut penser que des automates distribués sur un graphe et non sur un tore permettraient de mieux cerner quels regroupements sont possibles, en fonction de la topologie. On quitte alors ce que nous avons appelé automates cellulaires pour aller vers le formalisme plus général de Farmer.

ANNEXE C : LA CLASSIFICATION DE GUTOWITZ

C.1. Pour rendre compte du comportement temporel de la quantité : introduction à l'approximation des automates par des chaînes de Markov

La classification des automates cellulaires en quatre classes réalisée par Wolfram possède ses limites. Gutowitz a proposé une classification des automates grâce à une approximation du comportement de ces automates par des chaînes de Markov [Gutowitz, 1990]. L'idée d'étudier le comportement des automates à l'aide des chaînes de Markov n'est pas neuve puisqu'on la trouve déjà dans le livre de Bobrow et Arbib [Bobrow et Arbib, 1974]. Cette classification des comportements proposée par Gutowitz est moins connue que les classifications de Wolfram et Langton. Elle est pourtant beaucoup plus intéressante, plus fine et directement utilisable. Le problème spécifique est lié au fait que, de manière générale, des automates avec des règles tout à fait similaires peuvent avoir des comportements extrêmement différents. Ceci signifie que la structure qui établit la proximité des règles n'est pas en adéquation avec la structure qui établit la proximité des comportements. Réciproquement, des automates comportant des règles apparemment très différentes dans l'espace des règles peuvent avoir des comportements extrêmement proches. La classification de Wolfram permet de distinguer des comportements différents lorsque des règles sont données. Étant donnée une règle, on peut ainsi lui associer une classe. Le problème inverse n'est cependant pas résolu. On n'est pas capable, à l'heure actuelle, de caractériser les ensembles de règles, dans l'espace des règles, associés à chaque classe de comportement.

L'approche proposée par Gutowitz consiste à étudier les règles en fonction de leur action sur des mesures de probabilités. Il a déjà été montré que l'action des automates cellulaires sur des mesures de probabilités peut être approximée par leur action sur des mesures de Markov à n pas. Une mesure de Markov à n pas ou d'ordre n est, grossièrement, une mesure générée par un processus de Markov, avec une "mémoire" de n pas. Les mesures de Markov d'ordre n permettent de décrire les corrélations spatiales entre des cellules, à un moment donné. Le but est ici de résoudre un problème inverse. Étant donné le comportement d'un automate, on cherche à trouver l'ensemble des règles qui permettent d'obtenir un comportement similaire. Le point essentiel est que si la mémoire spatiale est faible, c'est-à-dire si n est petit, de nombreux automates cellulaires seront indistinguables dans une approximation de Markov d'ordre n . Le principe de cette classification consiste à considérer que si deux automates cellulaires ne peuvent être distingués que sur la base d'une corrélation spatiale au-delà d'une distance de n cellules, alors ces

deux automates seront considérés comme étant dans la même classe au $n^{ième}$ ordre. Les résultats numériques semblent indiquer que la prédiction du comportement de chaque classe s'améliore lorsque l'ordre de l'approximation augmente.

C.2. Quelques résultats

Sur l'ensemble des configurations d'un automate, on se donne la "topologie" dérivée de la métrique suivante¹. Soit une origine fixée pour les sites de l'automate, et i l'indice des différents sites, à partir de cette origine². Soient x et y deux configurations quelconques, alors :

$$d(x, y) = \sum_{i=-\infty}^{+\infty} 2^{-|i|} |x_i - y_i|$$

Soit μ une mesure de probabilité sur l'ensemble des configurations, et E un ensemble ouvert de configurations μ -mesurables. On définit alors τ_μ la mesure de probabilité obtenue après application à la probabilité de mesure μ de la règle τ d'un automate. On a alors l'équation

$$\tau_\mu(E) = \mu(\tau^{-1}(E)) \quad (1)$$

ce qui ne pose aucun problème de définition puisque $\tau^{-1}(E)$ est aussi un ensemble ouvert de configurations μ -mesurables. On peut aussi écrire ceci différemment, avec r le rayon³ de la règle τ .

On appelle alors bloc un sous-ensemble de l'ensemble des états correspondant à un sous-ensemble de l'ensemble des sites ne comportant qu'une seule composante connexe. Soit $P^t(a)$ la probabilité d'existence d'un bloc a à l'instant t . Soit b un second bloc, on pose $\delta(\tau(a), b)$, avec δ le symbole de Kronecker, qui vaut donc 1 si $\tau(a) = b$, avec $\tau(a)$ l'application de la règle τ au bloc a , et 0 sinon. On désigne la longueur d'un bloc a par $|a|$. Soit $\mathbf{B}_{|b|+2r}$ l'ensemble des blocs de longueur $|b|+2r$. On peut alors réécrire l'équation (1) comme un système infini pour tous les b possibles, et on obtient l'équation suivante :

$$P^{t+1}(b) = \sum_{a \in \mathbf{B}_{|b|+2r}} \delta(\tau(a), b) P^t(a) \quad (2)$$

Avec cette nouvelle équation, lorsque l'on voudra calculer la probabilité d'existence des blocs d'une longueur donnée, il suffira de la calculer à partir des probabilités d'existence des blocs de longueur inférieure. Plus précisément, on pourra construire

¹On fixe ainsi les relations de voisinage possibles, c'est-à-dire les relations de structures entre sites de l'automate.

²Il est également possible d'indexer les sites d'un automate en dimension quelconque.

³Ceci signifie, comme précédemment, que le voisinage comporte $2r + 1$ éléments en dimension une. On appelle alors $d = 2r + 1$ le diamètre de la règle d'un automate.

une approximation de Markov d'ordre n à partir des approximations successives aux ordres inférieurs, grâce à l'équation (2).

Pour obtenir l'approximation à l'ordre 0, l'équation (2) se simplifie de manière extrême, et il suffit de remplacer $P^t(a)$ dans le terme de droite par $\frac{1}{2^{|a|}}$. On obtient alors l'équation suivante, par exemple pour l'évolution de la probabilité associée à l'existence d'un site dans l'état 1 :

$$P_1^{t+1} = \sum_{a \in \mathbf{B}_{1+2r}} \frac{\delta(\tau(a), 1)}{2^{1+2r}} \quad (3)$$

On peut réécrire cette probabilité plus simplement, si l'on appelle λ le nombre de voisinages qui aboutissent à 1 avec la règle de l'automate. On observe alors que P_1^{t+1} est une constante $P_1 = \frac{\lambda}{2^d}$.

On peut ensuite chercher à calculer l'approximation à l'ordre 1 à partir de l'approximation à l'ordre 0. On se retrouve alors avec une approximation de type champ moyen. L'approximation de champ moyen, qui élimine les corrélations trop fortes, n'est rien d'autre ici qu'une approximation qui considère le processus comme markovien d'ordre un. On note D_a^0 et D_a^1 , le nombre de 0 et de 1 dans le bloc a . La probabilité d'existence d'un bloc a dans l'approximation de champ moyen est donnée par $P(a) = P_1^{D_a^1} P_0^{D_a^0}$. On peut alors substituer cette nouvelle relation dans l'équation (2). On appelle a_i le nombre de blocs de voisinages qui aboutissent à un nouveau 1 avec la règle de l'automate et qui contiennent aussi i fois la valeur 1 (c'est-à-dire qui sont dans la même approximation d'ordre 1). On obtient finalement l'équation d'évolution de la probabilité markovienne d'ordre 1 :

$$P_1^{t+1} = \sum_{i=0}^d a_i (P_1^t)^i (1 - P_1^t)^{d-i} \quad (4)$$

Les coefficients a_i peuvent prendre toute valeur entre 0 et le coefficient du binôme de Newton C_d^i . Le point fixe P_1^* de l'équation d'évolution (4) est une valeur estimée de la densité invariante d'un automate cellulaire dont les règles amènent les coefficients a_i . Avec la même méthode, on peut calculer explicitement l'approximation de Markov à l'ordre 2, ainsi que la relation de récurrence qui fournit l'approximation de Markov d'ordre n . Nous ne développerons pas ces calculs ici.

Les résultats expérimentaux montrent que cette classification correspond assez bien à des comportements similaires à ceux des automates. On observe, de plus, que le nombre d'éléments dans chaque classe est très variable. Ceci correspond qualitativement à la variabilité du nombre de règles appartenant à chacune des classes définies par Wolfram.

RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 3

- [deb, 1990] *La querelle du déterminisme*. Le Débat, Gallimard, 1990.
- [Andersen *et al.*, 1991] ANDERSEN J., JENSEN H. ET MOURITSEN O. Crossover in the power spectrum of a driven diffuse lattice-gas model. *Physical review B*. 1991, vol. 44, no 1, p. 439–442.
- [Bak *et al.*, 1992a] BAK P., CHEN K., SCHEINKMAN J. ET WOODFORD M. *Self-organized criticality and fluctuations in economics*. Santa Fe Institute, 1992. Rapport technique 92-04-018.
- [Bak *et al.*, 1992b] BAK P., FLYVBJERG H. ET LAUTRUP B. Coevolution in a rugged fitness landscape. *Physical review A*. 1992, vol. 46, no 10, p. 6724–6730.
- [Bak *et al.*, 1988] BAK P., TANG C. ET WIESENFIELD K. Self-organized criticality. *Physical review A*. 1988, vol. 38, no 1, p. 364–374.
- [Balian, 1986] BALIAN R. *Physique statistique*. Édition 1986. École Polytechnique, Palaiseau, 1986. Deux volumes, 268 p. et 246 p. ISBN 2-7302-0145-9 et ISBN 2-7302-0146-7.
- [Basdevant, 1987] BASDEVANT J.-L. *Physique. Mécanique quantique*. Édition 1987. École Polytechnique, Palaiseau, 1987. Deux volumes de 332 p. et 328 p. ISBN 2-7302-0154-8 et ISBN 2-7302-0155-6.
- [Bergé *et al.*, 1988] BERGÉ P., POMEAU Y. ET VIDAL C. *L'ordre dans le chaos : vers une approche déterministe de la turbulence*. Hermann, Paris, 1988. 353 p. Volume 33 de la collection Enseignement des sciences. ISBN 2-7056-5980-3.
- [Bersini et Detours, 1994] BERSINI H. ET DETOURS V. Asynchrony induces stability in cellular automata based models. Dans *Artificial life IV*, Brooks R. et Maes P., éditeurs, p. 382–387. MIT Press, Cambridge (Mass.), 1994. ISBN 0-262-52190-3.
- [Biham *et al.*, 1992] BIHAM O., MIDDLETON A. ET LEVINE D. Self-organization and a dynamical transition in traffic-flow models. *Physical review A*. 1992, vol. 46, no 10, p. R6124–R6127.

- [Bobrow et Arbib, 1974] BOBROW L. ET ARBIB M. *Discrete mathematics: applied algebra for computer and information science*. Saunders, Philadelphie, 1974. 719 p. ISBN 0-7216-1768-9.
- [Bussemaker et Ernst, 1993] BUSSEMAKER H. ET ERNST M. Lattice gas automata with self-organization. *Physica A*. 1993, vol. 194, p. 258–270.
- [Codd, 1968] CODD E. F. *Cellular automata*. Academic, New York, 1968.
- [Dehornoy, 1993] DEHORNOY P. *Complexité et décidabilité*. Springer, Paris, 1993. 200 p. Volume 12 de la série Mathématiques et applications. ISBN 2-287-00416-5.
- [Delahaye, 1994] DELAHAYE J.-P. *Information, complexité et hasard*. Hermès, Paris, 1994. 275 p. ISBN 2-86601-410-3.
- [Diu *et al.*, 1989] DIU B., GUTHMANN C., LEDERER D. ET ROULET B. *Physique statistique*. Première édition. Hermann, Paris, 1989. 1018 p. Volume 37 de la collection Enseignement des sciences. ISBN 2-705-66065-8.
- [Doolen, 1991] Doolen G., éditeur. *Lattice gas methods: theory, applications, and hardware*. MIT Press, Cambridge (Mass.), 1991. ISBN 0-262-54063-0.
- [Doolen *et al.*, 1990] Doolen G., Frisch U., Hasslacher B., Orszag S. et Wolfram S., éditeurs. *Lattice gas methods for partial differential equations*. Addison-Wesley, Redwood City (California), 1990. Volume IV de la collection Santa Fe institute studies in the sciences of complexity. ISBN 0-201-13232-X.
- [Drossel et Schwabl, 1992] DROSSEL B. ET SCHWABL F. Self-organised criticality in a forest-fire model. *Physica A*. 1992, vol. 191, p. 47–50.
- [Droz, 1993] DROZ M. *Introduction au groupe de renormalisation*. Université de Genève, Département de physique théorique, 1993. 171 p.
- [Dupuy, 1994] DUPUY J.-P. *Aux origines des sciences cognitives*. Première édition. La découverte, Paris, 1994. 188 p. Collection Textes à l'appui/série sciences cognitives. ISBN 2-7071-2200-9.
- [Enriques, 1941] ENRIQUES F. *Causalité et déterminisme dans la philosophie et l'histoire des sciences*. Hermann, Paris, 1941. 114 p. Volume 899 de la collection Actualités scientifiques et industrielles.
- [Flyvbjerg et Lautrup, 1992] FLYVBJERG H. ET LAUTRUP B. Evolution in a rugged fitness landscape. *Physical review A*. 1992, vol. 46, no 10, p. 6714–6723.
- [Flyvbjerg *et al.*, 1993] FLYVBJERG H., SNEPPEN K. ET BAK P. Mean field theory for a simple model of evolution. *Physical review letters*. 1993, vol. 71, no 24, p. 4087–4090.

- [Frisch *et al.*, 1987] FRISCH U., D'HUMIÈRES D., HASSLACHER B., LALLEMAND P., POMEAU Y. ET RIVET J.-P. Lattice gas hydrodynamics in two and three dimensions. *Complex systems*. 1987, vol. 1, p. 649–707.
- [Frisch *et al.*, 1986] FRISCH U., HASSLACHER B. ET POMEAU Y. Lattice-gas automata for the Navier-Stokes equation. *Physical review letters*. 1986, vol. 56, no 14, p. 1505–1508.
- [Gardner, 1970] GARDNER M. The fantastic combinations of John Conway's new solitaire game "life". *Scientific american*. 1970, vol. 223, p. 120–123.
- [Gardner, 1971a] GARDNER M. Geometric fallacies: hidden errors pave the road to absurd conclusions. *Scientific american*. 1971, vol. 224, p. 114–117.
- [Gardner, 1971b] GARDNER M. On cellular automata, self-reproduction, the garden of Eden and the game "life". *Scientific american*. 1971, vol. 224, p. 112–117.
- [Gardner, 1971c] GARDNER M. The orders of infinity, the topological nature of dimension and "supertasks". *Scientific american*. 1971, vol. 224, p. 106–109.
- [Gardner, 1972] GARDNER M. How to triumph at nim by playing safe, and John Horton Conway's game "Hackenbush". *Scientific american*. 1972, vol. 226, p. 104–107.
- [Gerola et Seiden, 1978] GEROLA H. ET SEIDEN P. Stochastic star formation and spiral structure of galaxies. *Astrophysical journal*. 1978, no 223, p. 129.
- [Gleick, 1989] GLEICK J. *La théorie du chaos : vers une nouvelle science*. Première édition française. A. Michel, Paris, 1989. 424 p. ISBN 2-226-03635-0.
- [Goldstine, 1993] GOLDSTINE H. H. *The computer: from Pascal to von Neumann*. Cinquième édition. Princeton University Press, Princeton (New Jersey), 1993. 378 p. Première édition 1972. ISBN 0-691-02367-0.
- [Greenberg *et al.*, 1978] GREENBERG J., HASSARD B. ET HASTINGS S. Pattern formation and periodic structures in systems modelled by reaction-diffusion equations. *Bulletin of the American Mathematical Society*. 1978, no 84, p. 1296.
- [Gutowitz, 1990] GUTOWITZ H. A. A hierarchical classification of cellular automata. *Physica D*. 1990, vol. 45, p. 136–156.
- [Hardy *et al.*, 1976] HARDY J., DE PAZZIS O. ET POMEAU Y. Molecular dynamics of a classical lattice gas: Transport properties and time correlation functions. *Physical review A*. 1976, vol. 13, p. 1949.

- [Hardy *et al.*, 1973] HARDY J., POMEAU Y. ET DE PAZZIS O. Time evolution of a two-dimensional model system. I. Invariant states and time correlation functions. *Journal of mathematical physic.* 1973, vol. 14, p. 1746.
- [Ising, 1925] ISING E. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik.* 1925, vol. 31, p. 253–258.
- [Jaeger et Nagel, 1992] JAEGER H. ET NAGEL S. La physique de l'état granulaire. *La recherche.* 1992, vol. 23, no 249, p. 1380–1387.
- [Kari, 1990] KARI J. Reversibility of 2D cellular automata is undecidable. *Physica D.* 1990, vol. 45, p. 379–385.
- [Kauffman, 1989] KAUFFMAN S. Adaptation on rugged fitness landscapes. Dans *Lectures in the sciences of complexity*, Stein D., éditeur, p. 527–618. Addison-Wesley, Redwood City (California), 1989.
- [Kauffman et Johnsen, 1991] KAUFFMAN S. ET JOHNSEN S. Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and coevolutionary avalanches. *Journal of theoretical biology.* 1991, vol. 149, p. 467–505.
- [Langer, 1980] LANGER J. Instabilities and pattern formation in crystal growth. *Revue of modern physics.* 1980, no 52, p. 1.
- [Langton, 1988] LANGTON C. Artificial life. Dans *Artificial life*, Langton C., éditeur, p. 1–47. Volume VI de la collection Santa Fe institute studies in the sciences of complexity, Addison-Wesley, Redwood City (California), 1988. ISBN 0-201-09356-1.
- [Langton, 1990] LANGTON C. Computation at the edge of chaos: phase transitions and emergent computation. *Physica D.* 1990, no 42, p. 12–37.
- [Langton, 1992] LANGTON C. Life at the edge of chaos. Dans *Artificial life II*, Langton C., Taylor C., Farmer J. et Rasmussen S., éditeurs, p. 41–91. Volume X de la collection Santa Fe institute studies in the sciences of complexity, Addison-Wesley, Redwood City (California), 1992. ISBN 0-201-52570-4.
- [Le Bellac, 1988] LE BELLAC M. *Des phénomènes critiques aux champs de jauge : une introduction aux méthodes et aux applications de la théorie quantique des champs.* InterÉditions/Éditions du CNRS, Paris, 1988. 639 p. Collection Savoirs actuels. ISBN 2-7296-0197-X 2-222-04026-4.
- [Lestienne, 1990] LESTIENNE R. *Les fils du temps : causalité, entropie, devenir.* Presses du CNRS, Paris, 1990. 292 p. ISBN 2-87682-037-4.

- [Liggett, 1985] LIGGETT T. *Interacting particle systems*. Springer, New York, 1985. 488 p. Volume 276 de la collection Grundlehren der mathematischen Wissenschaften. ISBN 0-387-96069-4 3-540-96069-4.
- [Lumer et Nicolis, 1994] LUMER E. ET NICOLIS G. Synchronous versus asynchronous dynamics in spatially distributed systems. *Physica D*. 1994, vol. 71, p. 440–452.
- [Marco, 1996] MARCO. *Cours d'agrégation sur les séries de Fourier*. Université Pierre et Marie Curie, Paris, 1996.
- [Martin *et al.*, 1984] MARTIN O., ODLYZKO A. ET WOLFRAM S. Algebraic properties of cellular automata. *Communications in mathematical physics*. 1984, vol. 93, p. 219–258. Reproduit dans [Wolfram, 1994], p. 71-113.
- [Mézard *et al.*, 1987] MÉZARD M., PARISI G. ET VIRASORO M. *Spin glass theory and beyond*. World Scientific, Singapour, 1987. 461 p. ISBN 9971-50-116-3.
- [Mézard et Toulouse, 1991] MÉZARD M. ET TOULOUSE G. Des verres de spin aux réseaux de neurones. *La recherche*. 1991, vol. 22, no 232, p. 616–623.
- [Miller, 1995] MILLER G. Artificial life as theoretical biology: How to do real science with computer simulation. *Article en ligne sur internet*. 1995.
- [Mumford, 1991] MUMFORD D. On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological cybernetics*. 1991, vol. 65, p. 135–145.
- [Onsager, 1944] ONSAGER L. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical review*. 1944, vol. 65, no 3 et 4, p. 117–149.
- [Potts, 1952] POTTS R. Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*. 1952, vol. 48, p. 106–109.
- [Prado et Olami, 1992] PRADO C. ET OLAMI Z. Inertia and break of self-organized criticality in sandpile cellular-automata models. *Physical review A*. 1992, vol. 45, no 2, p. 665–669.
- [Prez, 1996] PREZ P. *Voir : de la tâche aveugle au sujet de la vision*. CRÉA, École Polytechnique, Palaiseau, 1996. Rapport technique 9624.
- [Prigogine, 1996] PRIGOGINE I. *La fin des certitudes : temps, chaos et les lois de la nature*. O. Jacob, Paris, 1996. 226 p. ISBN 2-7381-0330-8.
- [Prum, 1986] PRUM B. *Processus sur un réseau et mesures de Gibbs : applications*. Masson, Paris, 1986. 190 p. Collection Techniques stochastiques. ISBN 2-225-80914-3.

- [Rice, 1953] RICE H. G. Classes of recursively enumerable sets and their decision problem. *Transactions of the American Mathematical Society*. 1953, vol. 74, p. 358–366.
- [Rolles, 1993] ROLLES J.-L. Complexity, a theory of quantity. *Theoria*. 1993, vol. 19, p. 123–155.
- [Rosen, 1981] ROSEN R. Pattern generation in networks. *Progress in theoretical biology*. 1981, no 6, p. 161.
- [Saurel, 1992] SAUREL P. *Q-learning et apprentissage par renforcement*. DEA de sciences cognitives, EHESS, Paris, 1992.
- [Saurel, 1996] SAUREL P. *Comment les sciences cognitives conçoivent l'étude des processus*. CRÉA, École Polytechnique, Palaiseau, 1996. Rapport technique 9627.
- [Schewe, 1981] SCHEWE P. Galaxies, the Game of life, and percolation. Dans *Physics news*, Schewe P., éditeur. American Institute of Physics, Pub R-302, 61, 1981.
- [Searle, 1996] SEARLE J. Deux biologistes et un physicien en quête de l'âme : Crick, Penrose et Edelman passés au scalpel de la critique philosophique. *La Recherche*. 1996, no 287, p. 62–77.
- [Sornette et Sornette, 1989] SORNETTE A. ET SORNETTE D. Self-organized criticality and earthquakes. *Europhysics letters*. 1989, vol. 9, no 3, p. 197–202.
- [Stern, 1990] STERN J. *Fondements mathématiques de l'informatique*. McGraw-Hill, Paris, 1990. 320 p. Collection Informatique. ISBN 2-7042-1228-7. ISSN 0989-392-X.
- [Stevens, 1974] STEVENS P. *Patterns in Nature*. Little, Brown, Boston, 1974.
- [Takahashi, 1990] TAKAHASHI S. Cellular automata and multifractals: dimension spectra of linear cellular automata. *Physica D*. 1990, vol. 45, p. 36–48.
- [von Neumann, 1951] VON NEUMANN J. The general and logical theory of automata. *Cerebral mechanisms in behavior - the Hixon symposium, Pasadena, 1948*. 1951, p. 1–31. L. Jeffress, éditeur. J. Wiley, New York. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 288–328. Pergamon Press, Oxford, 1961–63.
- [von Neumann, 1956] VON NEUMANN J. Probabilistic logics and the synthesis of reliable organisms from unreliable components. Dans *Automata studies*, Shannon C. et McCarthy J., éditeurs, p. 43–98. Princeton University Press, Princeton

(New Jersey), 1956. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 329-378. Pergamon Press, Oxford, 1961-63.

[von Neumann, 1966] VON NEUMANN J. *Theory of self-reproducing automata*. A. Burks, éditeur. University of Illinois Press, Urbana et Londres, 1966.

[von Neumann, 1992] VON NEUMANN J. *L'ordinateur et le cerveau*. Première édition française. La découverte, Paris, 1992. 130 p. Collection Textes à l'appui/série sciences cognitives. Traduction de The computer and the brain - the Silliman lectures. Première édition américaine, Yale University Press, New Haven, 1958. ISBN 2-7071-2164-9.

[von Neumann *et al.*, 1959] VON NEUMANN J., BLAIR A., METROPOLIS N., TAUB A. ET TSINGOU M. A study of a numerical solution to a two-dimensional hydrodynamical problem. *Mathematical tables and other aids to computation*. 1959, vol. 13, p. 145-184. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 611-650. Pergamon Press, Oxford, 1961-63.

[Wolfram, 1983] WOLFRAM S. Statistical mechanics of cellular automata. *Reviews of modern physics*. 1983, vol. 55, p. 601-644. Reproduit dans [Wolfram, 1994], p. 3-69.

[Wolfram, 1984a] WOLFRAM S. Computation theory of cellular automata. *Communications in mathematical physics*. 1984, vol. 96, p. 15-57. Reproduit dans [Wolfram, 1994], p. 159-202.

[Wolfram, 1984b] WOLFRAM S. Geometry of binomial coefficients. *American mathematical monthly*. 1984, vol. 91, p. 566-571. Reproduit dans [Wolfram, 1994], p. 451-456.

[Wolfram, 1984c] WOLFRAM S. Universality and complexity in cellular automata. *Physica D*. 1984, vol. 10, p. 1-35. Reproduit dans [Wolfram, 1994], p. 115-157.

[Wolfram, 1985] WOLFRAM S. Twenty problems in the theory of cellular automata. *Physica scripta*. 1985, vol. T9, p. 170-183. Fifty-ninth Nobel symposium. Reproduit dans [Wolfram, 1994], p. 457-485.

[Wolfram, 1986a] WOLFRAM S. Cellular automaton fluids 1: basic theory. *Journal of statistical physics*. 1986, vol. 45, p. 471-526. Reproduit dans [Wolfram, 1994], p. 359-408.

[Wolfram, 1986b] WOLFRAM S. Cryptography with cellular automata. 1986, vol. 218, p. 429–432. Advances in cryptology: crypto'85 proceedings. Collection Lecture notes in computer science. Springer. Reproduit dans [Wolfram, 1994], p. 487-490.

[Wolfram, 1986c] WOLFRAM S. Random sequence generation by cellular automata. *Advances in applied mathematics*. 1986, vol. 7, p. 123–169. Reproduit dans [Wolfram, 1994], p. 267-307.

[Wolfram, 1994] WOLFRAM S. *Cellular automata and complexity*. Première édition. Addison-Wesley, Reading (Mass.), 1994. 600 p. Collected papers de Stephen Wolfram. ISBN 0-201-62664-0.

[Wolfram et Salem, 1985] WOLFRAM S. ET SALEM J. *Thermodynamics and hydrodynamics of cellular automata*. 1985. Rapport technique, Thinking Machines Corporation.

CHAPITRE 4

**La pratique de la modélisation :
la mobilisation de masse
en RDA en 1989**

1. INTRODUCTION À LA PRATIQUE DE LA MODÉLISATION

Ce chapitre constitue une mise en pratique de la théorie de la complexité/quantité étudiée dans le chapitre précédent. Il s'articule en trois sections. La première section donne quelques éléments synthétiques sur le rôle et le fonctionnement de la modélisation. La deuxième section est une présentation des différents éléments de la modélisation d'un phénomène particulier relevant des sciences politiques, **la mobilisation de masse en République démocratique allemande (RDA) en 1989**. Dans une troisième section, qui fait office de conclusion, nous montrons comment il est possible de poursuivre cette modélisation. Cette dernière section permet également de mettre en évidence l'apport de ce travail au modélisateur en sciences cognitives comme au spécialiste de sciences politiques. On établit aussi l'intérêt d'une telle modélisation pour notre discipline, les sciences cognitives, même si des relations évidentes ne sont pas visibles *a priori*.

Dans cette première section nous souhaitons esquisser très rapidement, mais dans un ensemble cohérent, les différents éléments, quelque peu épars, que nous avons rencontrés jusqu'à présent et qui sont liés à la modélisation. Pour cela nous nous appuierons essentiellement sur trois figures, le but du texte étant simplement de fournir le début d'un commentaire très grossier de ces 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 font 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.* De même que nous ne présentons pas 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 cette section est de montrer que la modélisation est une activité qui ne peut pas être réduite à l'analyse d'un modèle, c'est-à-dire d'un système formel.** Le modèle à proprement parler est au centre de la modélisation, mais il n'en constitue qu'un des très nombreux aspects.

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

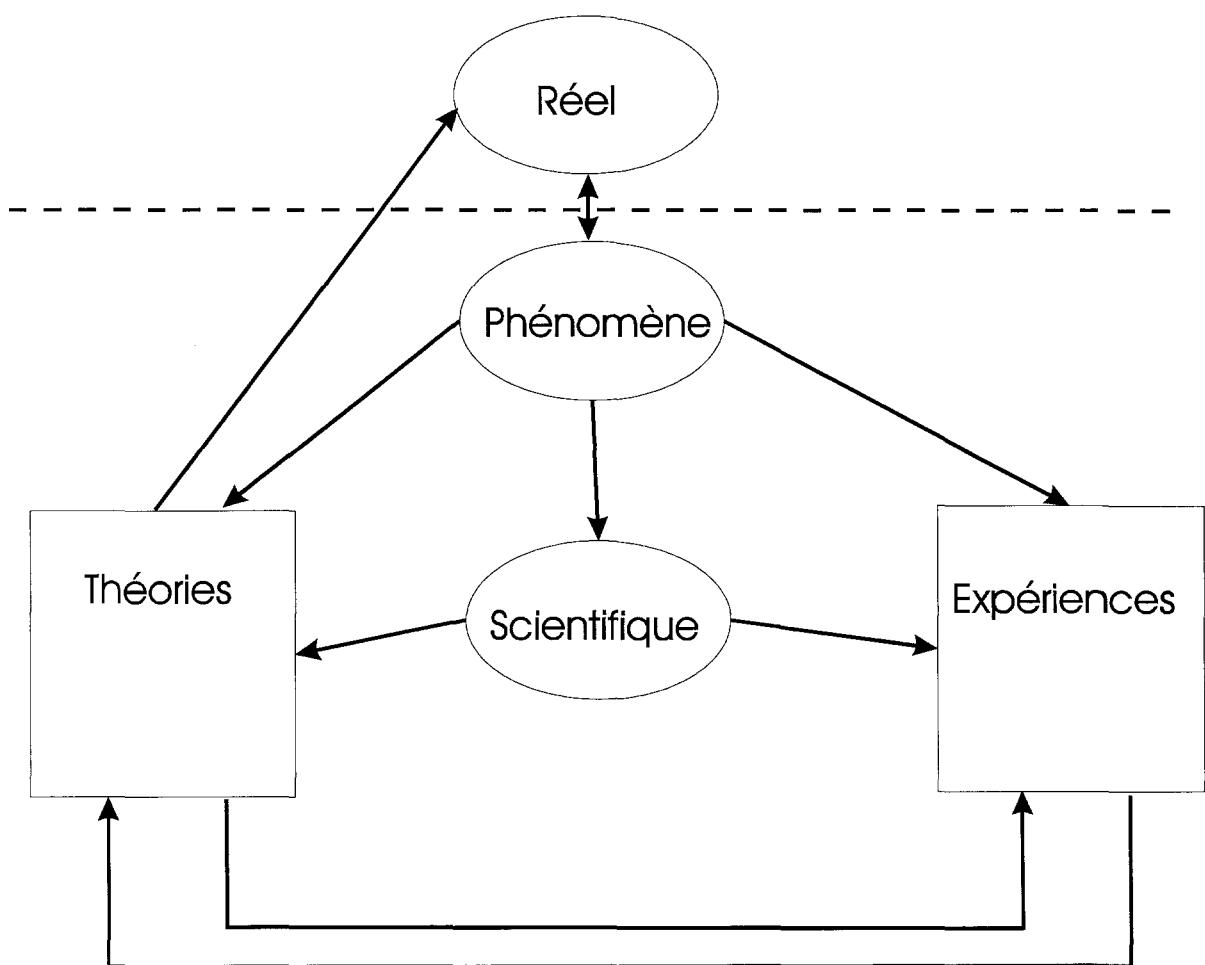


FIGURE 1 : La conception classique de la confrontation des théories aux expériences.

1.1. Une vision classique ou “popperienne” de la justification des théories

La conception de la justification des théories que nous trouvons chez Popper [Popper, 1978], 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 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 justification des théories. Nous ne retiendrons de la conception popperienne, pour notre propos, que les éléments qui apparaissent sur la figure 1.

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 expériences² et d'autre part les théories³. 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.

²Dans 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 prédefini par un observateur humain. Le terme ne saurait renvoyer aux “expériences computationnelles” qui ne partagent que leur nom avec les expériences à proprement parler. Comme nous l'avons signalé dans la section 2 du chapitre précédent, 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. Celles-ci permettent, entre autres, d'explorer facilement le fonctionnement d'un ensemble d'éléments formels dont le jeu a été prédefini. Nous voulons simplement insister sur le fait qu'il faut clairement distinguer ces deux types d’“expériences”.

³Nous 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 Cavaillès, 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.

1.2. Les quatre domaines de la modélisation

1.2.1. Les nouveaux domaines de la modélisation

Dans le contexte présenté précédemment, la notion de modèle au sens propre, c'est-à-dire de système formel ou axiomatique⁴, n'est pas évoquée. Ainsi **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, 1978], Popper n'utilise jamais le terme de modèle. Popper fait référence aux aspects linguistiques des justifications scientifiques, mais il ne s'appuie jamais sur les systèmes formels à proprement parler. Ceci transparaît en particulier dans son utilisation systématique du terme *énoncé*, à propos des produits de la science, alors qu'il ne parle jamais d'énoncés formalisés ou 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 ou "naïve" de la notion d'axiome⁵, comme elle se traduisait chez le premier Hilbert [Hilbert, 1971], ou chez Cantor. L'axiome n'est pas considéré ici comme une assertion purement symbolique qui explicite les règles du jeu symbolique, comme ce sera le cas pour les mathématiciens, après les travaux des années 30 sur la théorie des ensembles.

La conception classique ne tient pas compte non plus de l'utilisation nouvelle et massive de l'ordinateur. En particulier l'ordinateur a permis l'introduction de **la simulation comme quatrième domaine** qui est venu s'ajouter à la conception classique. Désormais la simulation est devenue un élément important de la justification d'une connaissance scientifique. Elle peut apporter différents éléments nouveaux. En particulier, elle permet de donner une véritable effectivité aux systèmes axiomatiques⁶ qui ont été définis. Comme nous l'avons vu dans le chapitre premier, la définition d'un système formel au sens propre, après les travaux de Turing, nécessite d'expliciter une machine.

Dans les conceptions classiques, un système d'axiomes, ou plus généralement d'énoncés, est porteur de tout le sens qu'il est susceptible de générer. Les axiomes sont alors le résumé suffisant de leur propre sens. Aussi, il peut sembler "inutile"

⁴En fait nous ne respectons pas tout à fait ici le terme de système formel tel qu'il a été défini par Gödel, suite aux travaux de Turing (voir page 43 du chapitre premier). Pour être tout à fait rigoureux, il faudrait dire que les modèles sont considérés ici comme des systèmes axiomatiques, alors que les simulations sont des systèmes formels. Afin de ne pas alourdir le texte, nous avons parfois utilisé le terme de système formel dans un sens affaibli par rapport à celui que lui attribue Gödel, mais en accord avec l'usage habituel.

⁵Pour plus de détails sur ces éléments, voir le chapitre 1.

⁶En fait la notion d'effectivité est partie intégrante de la notion de système formel. Nous avons déjà insisté sur ce point dans le premier chapitre.

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 formel, s'il est ou non conséquence du système d'axiomes. Répondre à cette question et déterminer ainsi les conséquences d'un système formel n'est pas toujours chose aisée à réaliser à partir des axiomes eux-mêmes. La question générale est tellement ardue que, comme nous l'avons vu avec les résultats de Gödel dans le chapitre premier, on peut parfois proposer des énoncés écrits dans les termes du système formel dont on ne peut pas dire intrinsèquement, c'est-à-dire dans le cadre du système formel, s'ils sont ou non des conséquences possibles du système d'axiomes. La simulation permet de préciser les comportements qui sont effectivement conséquence du système formel. La simulation permet donc parfois d'analyser les conséquences possibles d'un système d'axiome. Dans les termes de Popper, elle permet 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.

1.2.2. 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 s'est substituée à cette conception la notion de modélisation. **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 modèle⁷. **La modélisation est une activité**⁸. *La 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

⁷Un modèle, c'est-à-dire un système formel, conçu généralement indépendamment de son effectivité.

⁸Il ne s'agit pas là de l'activité du scientifique qui modélise, mais d'une interaction entre les domaines, intrinsèque à la modélisation.

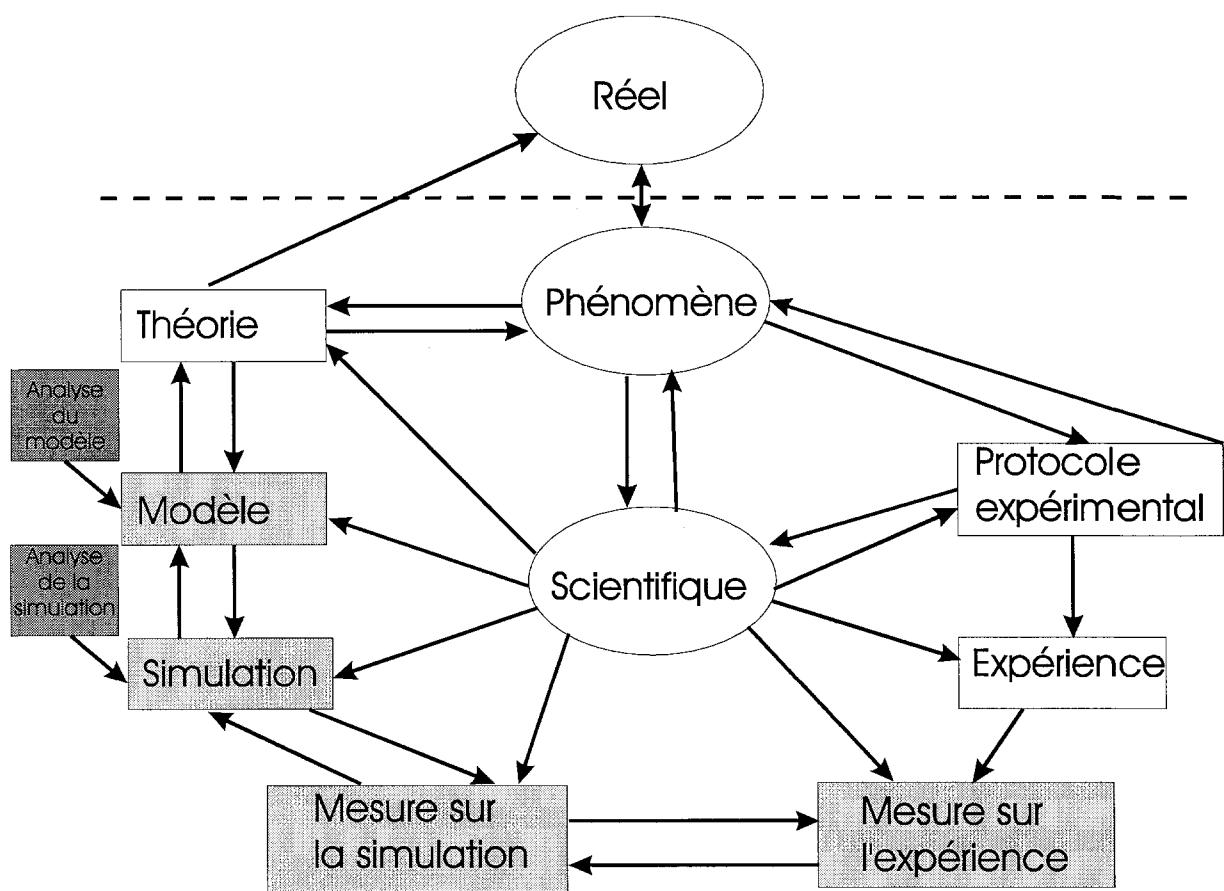


FIGURE 2 : Pratique actuelle de la confrontation des théories aux expériences : les quatre nouveaux domaines ainsi que l'analyse du modèle et l'analyse de la simulation.

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.*

1.2.3. Un cadre d'interprétation

La figure 2 permet d'interpréter facilement les différentes méthodologies que nous avons rencontrées dans les chapitres précédents.

Si on cherche à *interpréter la technique des gaz sur réseau dans ce cadre*, 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 est un peu différent. On est en présence ici d'un système formel, c'est-à-dire d'un modèle. Ce modèle est simulé et cette simulation permet de mettre en évidence un concept théorique. 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 parler de modélisation au sens propre dans le cadre du modèle du tas de sable, même si un modèle est en jeu.**

1.3. Différentes sophistications actuelles

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

1.3.1. L'analyse du modèle et l'analyse de la simulation

Aux quatre domaines précédents, on peut en ajouter 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.

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 initial⁹. 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 *et al.*, 1992] présentées dans le chapitre précédent. 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

⁹Voir à ce propos le chapitre 3, section 1.

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 justifier 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.

1.3.2. La multiréalisabilité

Chacun des différents domaines précédents constitue également un cadre contraint. *L'expression, dans ce cadre, de la situation à modéliser, est d'un type pré-déterminé.* Pour les modèles, par exemple, les expressions devraient se présenter sous forme de systèmes formels, même si dans la pratique elles se font en termes d'expressions mathématiques. 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 formels des 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 format formel, c'est-à-dire un seul outil mathématique pour transcrire dans le domaine des modèles toute situation à modéliser. Ceci permet de comprendre pourquoi certains scientifiques qui maîtrisent essentiellement un seul outil formel 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.

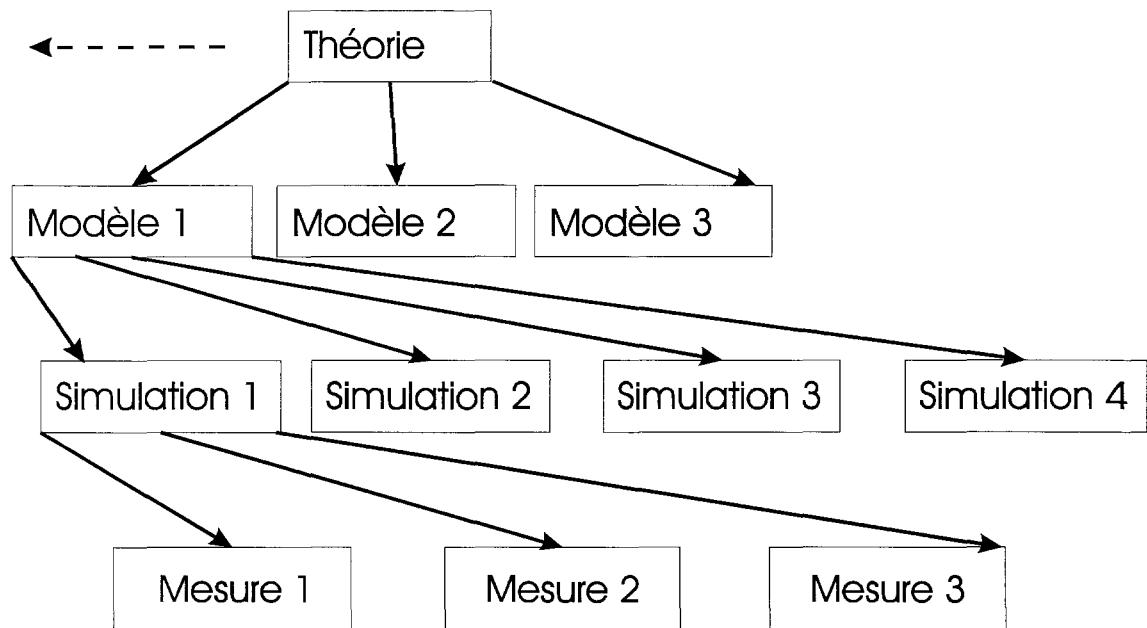


FIGURE 3 : La multiréalisabilité des théories et des modèles, ainsi que la multiplicité des mesures globales possibles sur une simulation. On notera également l'évolution des théories par déplacements de concepts.

Leur intérêt est méthodologique, il consiste à permettre d'exprimer, sous la forme d'un modèle, c'est-à-dire dans un des cadres 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.

On observe également une **multiréalisabilité 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 se rend compte que l'on est encore beaucoup trop loin de quelque chose 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. Il faut bien comprendre que cette multiréalisabilité 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. 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 en effet 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 c’est le changement de format, 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.

1.3.3. 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”. 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 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 acceptations multiples et variées.

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.

Le lecteur se sentira peut-être frustré à la fin de cette section 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 quelle est la référence actuelle pour la modélisation en sciences cognitives. 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 uniquement un problème de mise en correspondance, d'encodage et de décodage, pour passer des systèmes physiques aux systèmes formels.

Nous avons voulu esquisser dans cette section 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.

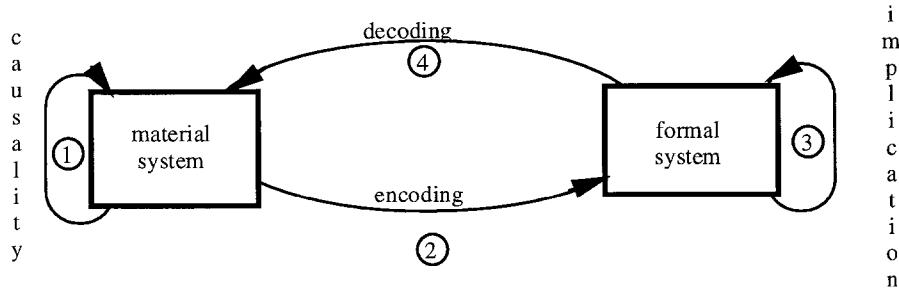


Figure 4a : La figure originale de Rosen. La modélisation selon Rosen [Rosen, 1987] est un problème d'encodage et de décodage des systèmes matériels vers les systèmes formels.

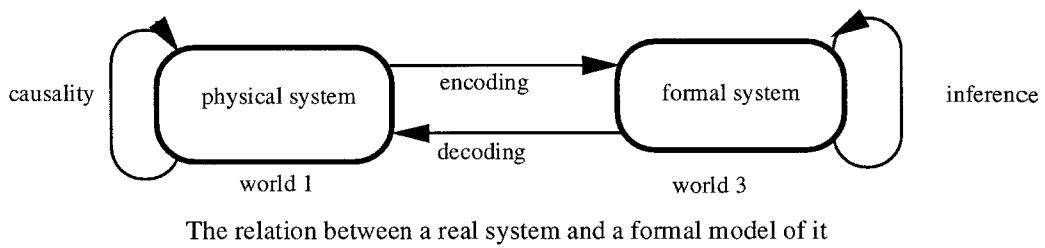


Figure 4b : La figure de Rosen reprise par Hallam et Malcolm [Hallam et Malcolm, 1994].

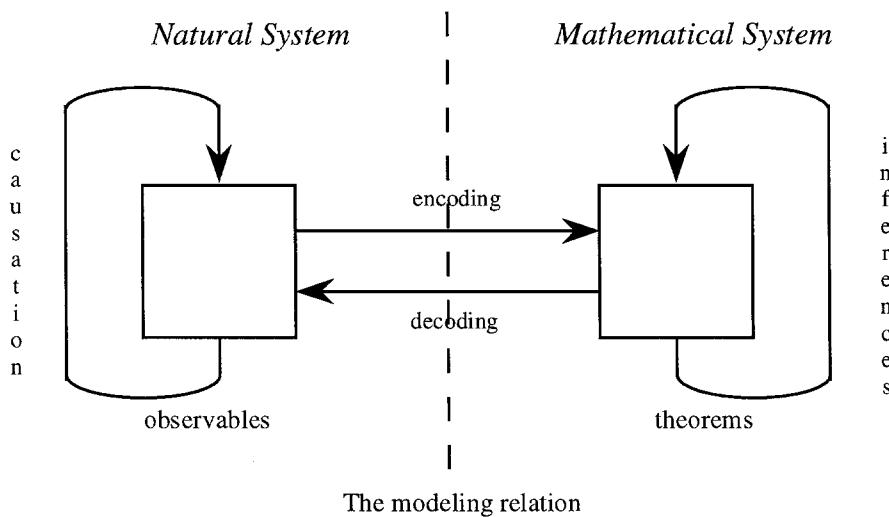


Figure 4c : La figure de Rosen reprise par Casti [Casti, 1996].

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

Chapitre 4 -- Figure 4

2. LA MODÉLISATION MISE EN PRATIQUE : LA MOBILISATION DE MASSE EN RDA EN 1989

2.1. Le phénomène étudié et la théorie qui lui est associée

Nous avons déjà insisté sur le fait que modéliser c'est trancher et choisir. Ce point très important nous amène à présenter le phénomène étudié de manière suffisamment détaillée pour que l'on perçoive l'enchevêtrement des faits devant lequel se retrouve le modélisateur qui va devoir construire son modèle. Seule la richesse et la relative précision de la description permettent de comprendre la nécessaire réflexion méthodologique qui va précéder le travail sur le modèle formel proprement dit. Mais ces faits que nous a présentés Wolf-Dieter Eberwein, un spécialiste de sciences politiques, ne sont ni bruts ni neutres. Ils sont déjà organisés autour de concepts spécifiques aux sciences politiques. C'est dans ce contexte que la modélisation peut commencer. Dans ce cadre les schémas décrits dans la section précédente nous ont permis de proposer un modèle en conservant en mémoire les objectifs de la modélisation.

2.1.1. Narration des événements

Introduction

Nous présentons ici le phénomène qui était susceptible d'être modélisé. Pour comprendre la suite du travail et les enjeux de la modélisation, il faut tout d'abord exposer brièvement ce phénomène¹. Il s'agit grossièrement de modéliser certains aspects de la chute du régime communiste en RDA en 1989. Le changement de régime a eu lieu sous la forme d'une révolution pacifique. Le terme de révolution a même été contesté. En effet, le concept de *révolution* n'est pas très clair. Il est généralement associé à une rupture politique accompagnée d'une violence organisée, comme ce fut le cas pour la Révolution Française en 1789 ou la Révolution Russe en 1917. En ce sens, la violence serait un élément constitutif d'une révolution. D'autres spécialistes proposent de réservier le terme de révolution aux mouvements de contestation organisés qui sont parvenus à leurs fins et ont réussi à obtenir un changement de la nature du régime politique. Cette dernière conception permet de

¹On trouvera des explications plus développées dans [Eberwein et Saurel, 1995].

parler de révolution uniquement lorsque celle-ci a abouti. Elle confond ainsi la phase révolutionnaire et ses conséquences sur la nature du régime politique. Par définition on ne peut alors plus envisager de révolution qui échoue. Ceci est de toute évidence contestable, puisque l'on peut parfois considérer que l'on participe à une révolution alors que l'on ne sait pas encore quelles seront les conséquences à venir, si elles existent, sur la nature du régime politique. Nous avons suivi dans ce travail une autre conception de la révolution, proposée par Kuran. Selon lui, une révolution se définit comme "... un changement fondamental dans l'ordre social, bouleversé dans une brève période de temps, par le biais d'un changement massif des opinions politiques exprimées par la population". Cette définition est plus générale et semble plausible. Si on suit cette définition, le terme de révolution est approprié pour parler des événements qui se sont déroulés en RDA en 1989.

Lorsque l'on souhaite faire un récit précis des événements, il faut garder à l'esprit que cette narration s'appuiera sur le choix de certains faits et de certaines données. Or ces données sont rarement disponibles et lorsqu'elles le sont, elles restent partielles et souvent insuffisantes. On peut cependant dégager les moments principaux autour desquels se sont articulés les événements de 1989. En l'occurrence, il est important de noter deux phases très distinctes. La première est constituée des antécédents historiques qui ont amené la RDA dans une situation prérévolutionnaire. Cette première phase est importante si l'on ne souhaite pas considérer que ce phénomène révolutionnaire s'est produit *ex nihilo*, mais plutôt qu'il s'ancre dans une certaine perspective historique. On tient compte de l'aboutissement de cette première phase dans les conditions initiales du modèle. La deuxième phase est la période révolutionnaire elle-même. Cette seconde phase est celle dont le modèle doit rendre compte par son fonctionnement.

La situation en RDA au début de l'année 1989 et les antécédents historiques

Il nous faut tout d'abord préciser la situation générale de la RDA à la veille des événements de 1989. Parmi les indications que l'on peut obtenir, il faut citer les données publiées par Köhler [Köhler, 1990] sur l'opinion publique en RDA, qui montrent bien que le mouvement de contestation de 1989 s'enracine dans une critique ancienne et n'est donc pas apparu sans raison. Elles indiquent que le niveau de satisfaction sur les conditions de vie et le niveau de vie en RDA dans les années quatre-vingts n'ont cessé de se détériorer de façon systématique, jusqu'à atteindre une situation extrême en 1989. Ceci semble étroitement lié à l'aggravation générale des résultats économiques de la RDA, symptomatiques, selon l'opinion publique, d'un dysfonctionnement du système économique comparable à ce que l'on observait à la même époque dans les autres pays du bloc communiste. Si l'on prend comme indicateur le taux d'investissement (taux général d'accumulation des biens), on observe clairement un déclin systématique dans les années quatre-vingts

jusqu'à 22%, alors que dans les années soixante-dix ce taux était de l'ordre de 26% [Eberwein *et al.*, 1990].

On peut donc résumer la situation en quatre points. Le premier point est la frustration de la population. Le second point est à associer à la dégradation de la situation économique. Le troisième point important est de noter que cette situation n'implique pas que la population considère sa classe dirigeante comme particulièrement incompétente. Ce troisième point nous amène à mettre en évidence deux facteurs qui doivent être considérés comme distincts, le soutien du régime et le soutien du gouvernement. Le premier facteur est "dépersonnalisé" alors que le deuxième est attaché aux personnes au pouvoir. Le quatrième point, dont nous aurons à tenir compte par la suite, est la constitution progressive de petits groupes contestataires, comme le mouvement pacifiste. On tiendra compte dans le modèle d'éléments contestataires irréductibles qui correspondront à ces groupes. Les thèmes qui les animent ne sont pas toujours directement liés à la situation interne en RDA. Ce n'est pas le cas par contre des mouvements écologistes. Il est à noter que les deux mouvements cités précédemment, le mouvement pacifiste, comme le mouvement écologiste ne sont pas centralisés. Ils sont constitués de petits groupes d'activistes qui agissent au niveau local.

Les années quatre-vingts ont joué un rôle déterminant. Elles ont permis la constitution progressive et l'extension de groupes d'opposition. Des groupes pacifistes se sont constitués pour lutter contre les dangers d'une guerre nucléaire. À partir de 1984 d'autres groupes contestataires ont porté leurs revendications vers des thèmes écologiques, puis sur l'égalité et les droits de l'homme dès 1987. En 1989, la contestation portait sur la réforme démocratique de la RDA et sa capacité à ouvrir une "troisième voie", intermédiaire entre capitalisme américain et communisme soviétique. *L'essentiel est de noter qu'au départ ces différents groupes de contestation étaient essentiellement organisés à un échelon local.* Ceci signifie en particulier que *ces groupes étaient isolés les uns des autres*, en partie du fait des moyens de communication. *Cette organisation locale des groupes permet de proposer un modèle dans lequel les districts vont être l'unité géographique de base. Les éléments du district et en particulier les éléments contestataires, dans le modèle, seront uniquement en relation avec d'autres éléments du même district.*

L'église protestante a joué un rôle déterminant pour la constitution de ces groupes. Elle offrait une protection et un toit contre les autorités. Des réseaux ont commencé à se constituer, mais seulement à partir de 1985. Selon les actes de la Stasi², on dénombrait 160 groupes de ce type en 1989, qui comprenaient au total environ 2 500 membres actifs. Toujours selon la Stasi, on estime à 600 le nombre d'individus ayant eu un rôle décisionnel. Le groupe d'irréductibles était constitué d'environ 60 membres, si l'on en croit un document de la Stasi datant de 1990. Sur les

²Stasi est l'abréviation de Staatssicherheit qui désigne les services de sécurité de l'État, c'est-à-dire la police secrète est-allemande.

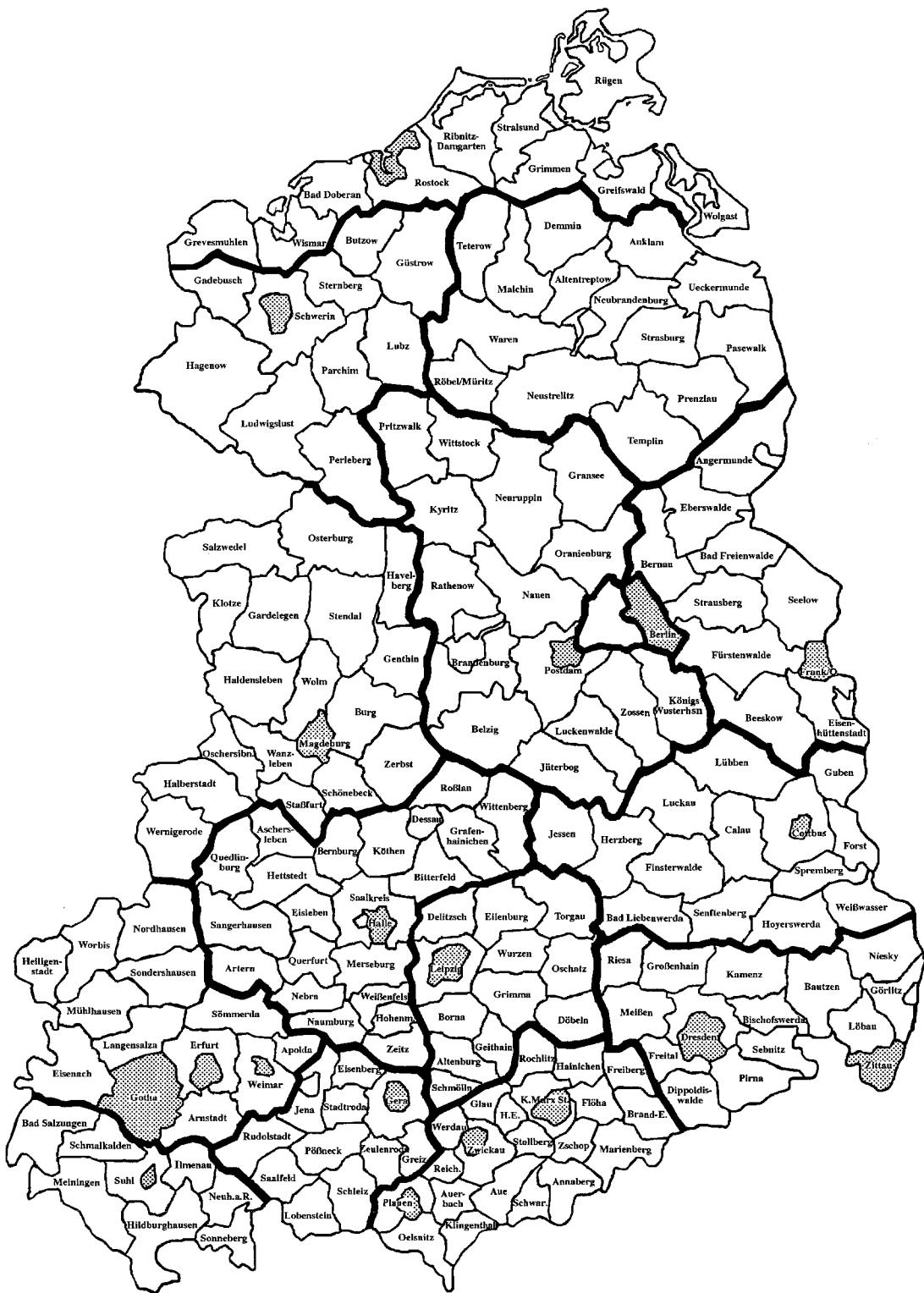


FIGURE 5 : La carte de la RDA avec la localisation des districts où se trouvaient les principaux groupes contestataires (en grisé).

160 groupes, environ 150 étaient ce que l'on a appelé les "groupes de base des églises" (kirchliche Basisgruppen). On connaît en partie la composition sociale de ces groupes. Ils étaient essentiellement composés de personnes âgées de 25 à 40 ans, avec une majorité d'intellectuels possédant une formation universitaire. Rares étaient les ouvriers. Ces groupes étaient concentrés essentiellement dans dix villes. Berlin-est comprenait à elle seule 19 groupes soit 12% de l'ensemble. Tous les autres étaient distribués dans le sud de la RDA à Leipzig, Halle, Wittenberg, Gera, Erfurt, Zwickau, Dresde et Karl-Marx-Stadt (figure 5). Tous ces chiffres sont cependant à prendre avec la plus extrême prudence. En effet ils proviennent de la Stasi et il semble aujourd'hui que le nombre de groupes et de personnes a été largement sous-estimé. On peut comparer ces données à celles plus récentes proposées par Blattert, Rink et Rucht [Blattert *et al.*, 1995] pour Berlin-est. Ces auteurs comptabilisent 964 membres actifs dans les groupes d'opposition et 2 809 membres passifs, soit un total de 3 773 membres. Ces chiffres ne tiennent compte que de Berlin-est et sont déjà supérieurs de 50 % à ceux qui étaient avancés par la Stasi pour la RDA dans son ensemble. On peut donc supposer que les opposants au régime ont été beaucoup plus nombreux que ne l'avait cru la Stasi. Si on conserve la proportion que proposait la Stasi selon laquelle 12 % des groupes d'opposition de toute la RDA se trouvaient à Berlin-est et si on considère que ces groupes ont en moyenne le même effectif dans tout le pays, on obtient plus de 8 000 membres actifs irréductibles sur l'ensemble du territoire (contre 600 selon la Stasi) et un total de 30 000 membres (contre 2 500 selon la Stasi). Globalement, selon ces hypothèses, on peut alors estimer que l'erreur de la Stasi est de l'ordre d'un facteur 10, ce qui est évidemment loin d'être négligeable. Cet exemple montre les difficultés à obtenir des données absolument fiables et provenant de sources très différentes. Ces différentes données sont reprises dans le modèle sous la forme de la proportion initiale de militants du district. Ces données permettent également de connaître par district la proportion de militants selon la catégorie sociale d'origine, facteur dont nous tiendrons compte dans le modèle.

La phase révolutionnaire : l'année 1989

On peut donc supposer qu'une détérioration générale de la performance économique et politique du système a eu lieu en même temps que se manifestait une désaffection croissante de la population pour le régime. De plus des organisations commençaient à émerger qui devenaient le lieu central d'une dissidence organisée. Ces groupes ont ainsi constitué le "groupe levier révolutionnaire", pour reprendre une expression de Kuran ("revolutionary bandwagon"). La question se pose alors de savoir comment ce potentiel révolutionnaire a été capable de s'activer. Deux facteurs sont susceptibles d'expliquer l'évolution du processus de mobilisation au cours de l'année 1989. Le premier facteur explicatif est associé à la capacité de ces groupes à transmettre et diffuser leurs conceptions vers les autres groupes. Ils ont également réussi à

Tableau 6a : Les manifestations

Mois	Nb de manifestants	Maximum	Non disponibles	Nb d'évènements
Janvier	2 157	900	11	17
Février	284	200	5	9
Mars	385	300	6	9
Avril	319	146	6	10
Mai	2 774	1 000	12	21
Juin	4 057	1 500	5	17
Juillet	5 700	4 500	8	12
Août	207	150	1	5
Septembre	27 119	8 000	24	45
Octobre	2 191 517	350 000	44	122
Novembre	3 108 704	1 000 000	25	69
Décembre	914 082	250 000	22	46

Tableau 6a : Les données associées aux manifestations. On trouve dans la première colonne le nombre de manifestants, c'est-à-dire le nombre cumulé de personnes actives dans le mois. La seconde colonne fournit le nombre de personnes qui ont participé à la plus grande manifestation du mois. La troisième colonne donne le nombre d'évènements pour lesquels le nombre de manifestants n'est pas disponible. La dernière colonne enfin donne le nombre de manifestations qui ont été répertoriées au cours du mois.

Tableau 6b : La répression

Mois	Nb de sanctions	Maximum	Non disponibles	Nb d'évènements
Janvier	44	20	1	5
Février	11	7	4	6
Mars	0	0	6	6
Avril	151	146	1	4
Mai	96	40	7	12
Juin	357	300	7	12
Juillet	50	40	3	5
Août	38	23	1	4
Septembre	77	40	6	9
Octobre	338	100	33	43
Novembre	0	0	0	0
Décembre	0	0	0	0

Tableau 6b : Les données associées à la répression. Dans la première colonne, on trouve le nombre total de sanctions répertoriées dans le mois, y compris les arrestations. La deuxième colonne propose le nombre maximum de sanctions réalisées pour un même évènement. La troisième colonne comptabilise le nombre d'évènements pour lesquels le nombre de sanctions n'est pas disponible. La quatrième colonne enfin fournit le nombre d'évènements répertoriés dans le mois et qui ont donné lieu à des sanctions.

Tableau 6c : La levée des sanctions

Mois	Personnes libérées	Maximum	Non disponibles	Nb d'évènements
Janvier	12	12	3	4
Février	0	0	0	0
Mars	0	0	2	2
Avril	20	20	1	2
Mai	324	300	0	2
Juin	200	200	1	2
Juillet	0	0	7	7
Août	0	0	0	0
Septembre	3 098	3 000	3	5
Octobre	11 729	5 000	27	34
Novembre	22	20	36	38
Décembre	865 125	450 000	11	15

Tableau 6c : Les données associées à la levée des sanctions et donc à la libéralisation du régime. Dans la première colonne, on trouve le nombre total de personnes dont la libération a été répertoriée dans le mois. La deuxième colonne propose le nombre maximum de personnes libérées au cours d'une même libération. La troisième colonne comptabilise le nombre de libérations pour lesquelles le nombre de personnes libérées n'est pas disponible. La quatrième colonne enfin fournit le nombre d'événements répertoriés dans le mois, c'est-à-dire le nombre de libérations.

FIGURE 6 : Trois tableaux résument les données obtenues pour les manifestations, la répression et la levée des sanctions.

établir des rencontres croisées entre les membres des différents groupes. Ce premier facteur est donc collectif. Le second facteur explicatif est au contraire individuel. Il est associé au choix de l'exil comme réponse individuelle à cette situation de mécontentement. On peut associer à ce choix personnel un effet de renforcement du mécontentement collectif et de la mobilisation générale. Comme le note Hirschman et contrairement à ce qu'il pensait dans le modèle qu'il avait développé auparavant, le nombre d'exilés et le nombre de mécontents se sont renforcés l'un l'autre (comparer le tableau 6a et la courbe 7b).

L'exil Le concept d'émigration ou d'exil défini par Hirschman est particulièrement adapté au cas de la RDA. En effet, la Constitution de la RFA avait postulé l'unicité de l'état allemand, même si les faits ne correspondaient pas à cette situation. Ceci signifiait en particulier que tout citoyen de la RDA était considéré comme un citoyen de la RFA de droit et à part entière. Ceci permettait aux allemands de l'est de bénéficier dès leur arrivée en RFA des avantages sociaux réservés aux allemands de l'ouest. Ils obtenaient une somme d'argent ainsi que les garanties d'une protection sociale minimale. Ceci a bien entendu facilité grandement la vague d'émigrés de RDA (courbe 7a).

Cette particularité expliquerait déjà l'évolution de la situation de Berlin-est en 1961. En effet, à cette date, les autorités est-allemandes avaient été obligées de construire le Mur entourant Berlin-est pour contenir l'exode massif des allemands de l'est vers la RFA. Dans les années suivantes et spécialement après l'expatriation de Wolf Biermann en 1976, il est apparu clairement que l'exil avait des propriétés particulières qui ne correspondaient pas complètement à l'interprétation de Hirschman. Selon Hirschman deux options alternatives permettaient de s'opposer au régime en place. La première possibilité consistait à choisir l'exil, alors que la seconde était de prôner la contestation au régime sur le sol même de la RDA. L'exil illégal apparaît alors comme un choix individuel d'opposition qui consiste à quitter le pays. Cette solution était devenue extrêmement dangereuse après la construction du Mur.

Hirschman n'a cependant pas mis en évidence l'utilisation de l'exil par le pouvoir pour se débarrasser de ses opposants les plus virulents. Biermann est l'exemple le plus caractéristique d'un opposant très critique que le régime a exilé. Ainsi le régime a lui-même utilisé l'exil comme un outil politique. Comme le montre l'analyse de Hirschman [Hirschman, 1993], le pouvoir est-allemand a également monnayé le renvoi de certains prisonniers politiques vers la RFA. Évidemment, cette stratégie qui consistait à rejeter certains opposants hors du pays s'avéra dépassée lorsque le nombre de prisonniers eut atteint un certain seuil.

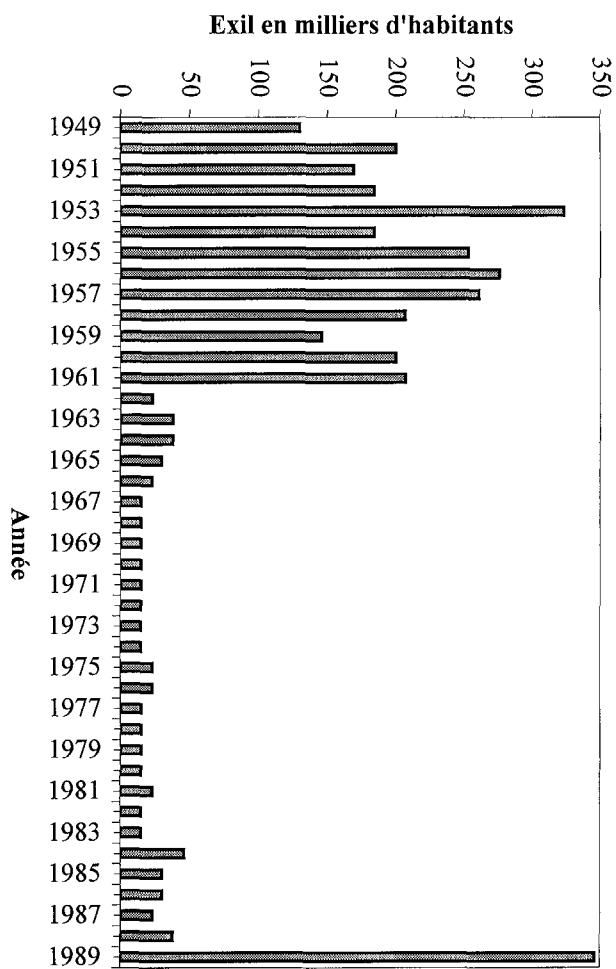
Déjà avant 1989, entre 1949 et 1961, on a constaté des périodes d'augmentation significative du nombre d'émigrés comme le montrent les données rapportées partiellement sur la figure 7. La période de 1949 à 1989 est marquée par différents

événements politiques en RDA (en 1953, 1956 et 1961, par exemple) auxquels on peut associer une modification de la courbe du nombre des exilés (courbe 7a). On observe que ce nombre a augmenté de manière “exponentielle” en 1989 (courbe 7b). Il est vraisemblable que des événements imprévisibles et extérieurs à la RDA permettent d’expliquer le phénomène. En effet le deux mai le gouvernement hongrois décida d’ouvrir sa frontière avec l’Autriche, mais cette mesure ne bénéficia tout d’abord qu’aux citoyens hongrois car Budapest continua d’exiger des autres ressortissants des pays de l’Est une autorisation spéciale de leur gouvernement. Le neuf août la Hongrie décida de ne plus refouler vers la RDA les ressortissants est-allemands surpris en train d’essayer de passer en Autriche et elle renonça même à viser leurs passeports. Enfin, le dix septembre la Hongrie annonça la suspension de son accord bilatéral de 1969 avec l’Allemagne de l’est, prévoyant le rapatriement en RDA des allemands de l’est qui ne seraient pas munis de visas valides. Ces évènements ont permis aux allemands de l’est de se rendre sans restriction en RFA via la Hongrie et l’Autriche. On a pu obtenir des informations sur la composition sociale des personnes qui souhaitaient quitter le pays (tableaux 7c et 7d). Les points essentiels à dégager sont d’une part la région d’origine des candidats à l’exil et d’autre part le groupe social auquel ils appartiennent. On constate en ce qui concerne l’origine sociale que la majorité est composée d’ouvriers. Pour la ville de Berlin-est, on trouve ces informations sur la figure 7d.

Sans entrer dans les détails, ces deux tables suggèrent deux éléments de réflexion. Tout d’abord, on peut suggérer que le nombre important de personnes qui ont quitté le pays pouvait laisser la porte ouverte à une réaction anti-révolutionnaire de la part de la population restante. Ceci est d’autant plus vrai que ceux qui ont quitté le pays provenaient essentiellement des zones géographiques où la contestation était la plus grande. Le second élément à noter est le fait que les personnes les plus instruites étaient moins inclines à partir pour la RFA que les ouvriers. Cette seconde hypothèse est également confirmée par les données des tableaux 7c et 7d, qui incluent uniquement la ville de Berlin-est. Il faut cependant relativiser ces données qui ont été établies sur la base du nombre de personnes exilées par catégorie sociale en fonction du nombre total de personnes exilées. Ce calcul augmente nécessairement la proportion d’ouvriers puisque ceux-ci constituent un groupe social très nombreux en RDA. Méthodologiquement, pour obtenir une proportion plus représentative de cet aspect, il serait plus judicieux de calculer la proportion du nombre d’exilés dans une catégorie sociale par rapport au nombre de personnes relevant de cette catégorie sociale. On peut noter également que les volontaires pour l’exil étaient pour la plupart des personnes âgées de 18 à 40 ans, alors que les personnes âgées étaient les plus nombreuses avant les événements de 1989.

L’élément suivant à analyser au cours de l’année 1989 est lié aux manifestations elles-mêmes. Le tableau 6a montre que les manifestations ont suivi une progression du même type que l’exil. Ces deux mesures du comportement suivent une croissance

Courbe 7a : L'exil de 1949 à 1989



Courbe 7b : L'exil en 1989

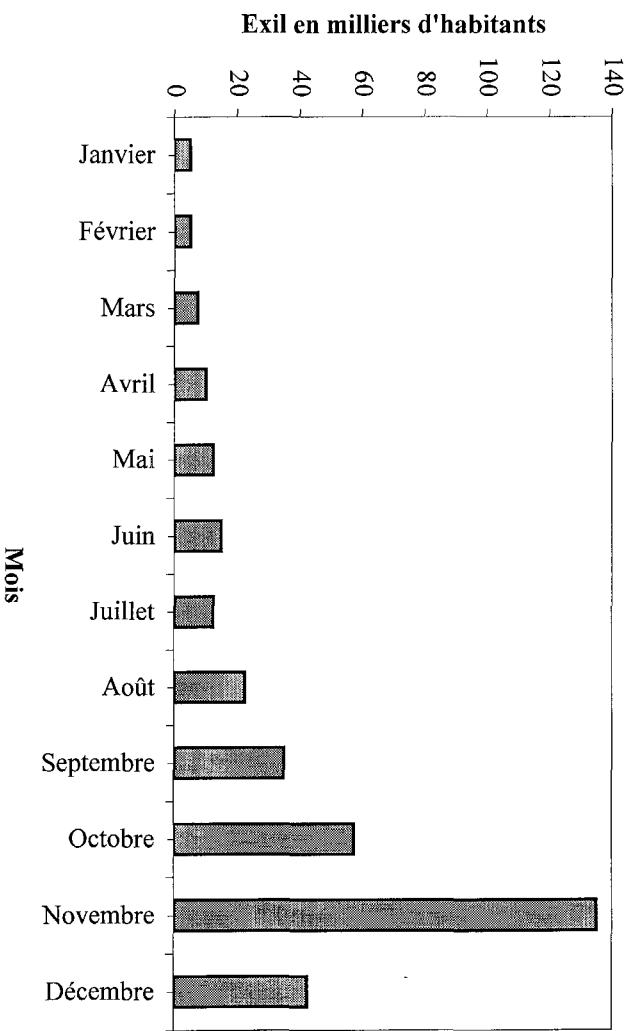


Tableau 7c : L'exil par tranches d'âges

Année	Moins de 18 ans	18-40 ans	40-60/65 ans	Plus de 65 ans
1982	18,7	28,8	14,3	38,2
1986	19,2	37,3	11,8	31,7
1989	29,3	53,9	14,2	2,6

Tableau 7c : La répartition de l'exil par tranches d'âges, pour la ville de Berlin-est. Les données sont fournies en pourcentages pour les différentes classes d'âges. Les plus de 65 ans qui pouvaient émigrer ont été remplacés par les 18-40 ans. (Source : Ines Schmitt, 1990:22).

Tableau 7d : L'exil par classe sociale

Année	Ouvriers	Intellectuels
1982	51,5	25,2
1986	65,0	18,8
1989	69,3	17,7

Tableau 7d : La répartition de l'exil selon les classes sociales, pour la ville de Berlin. Les données sont fournies en pourcentage pour une année. On observe une diminution de la proportion de personnes socialement favorisées ou éduquées qui quittent le pays. Au contraire la proportion d'ouvriers qui s'exilent augmente. (Source : Ines Schmitt, 1990:23).

FIGURE 7 : Deux courbes et deux tableaux résumant les données obtenues pour l'exil.

exponentielle. Hirschman en a conclu avec raison que l'exil et les manifestations avaient dû se renforcer réciproquement au cours de l'année 1989 pour aboutir à la chute du régime.

Cependant les manifestations et les modes d'expression n'étaient pas uniquement l'expression d'une désapprobation du régime. On peut distinguer deux motifs différents pour expliquer ces manifestations, et l'association de ces deux motifs permet sans doute d'expliquer en partie la forte dépendance de la courbe de l'exil et de celle des manifestations. Les gens manifestaient d'abord pour exprimer leur désir d'une certaine démocratisation et d'une libéralisation plus importante du régime. De plus les manifestants souhaitaient exprimer leur désir de pouvoir quitter le pays.

La loyauté La loyauté est en principe une disposition individuelle non observable. En effet la loyauté, définie par Hirschman, consiste à s'abstenir de contester. Cette disposition individuelle n'exprime donc rien sur le fait que cet individu soutient ou non le régime. On peut cependant obtenir des indices sur la loyauté en notant par exemple que la majorité de la population était certainement restée fidèle au régime, au moins au début des événements. Ils s'abstenaient d'exprimer leur désapprobation du régime et pour la plupart respectaient certainement le régime. On ne peut pas affirmer cela de l'ensemble des deux millions d'adhérents au SED, mais certainement d'une majorité d'entre eux.

Lorsque l'on regarde ensuite la répartition géographique des manifestations, on constate que Leipzig est l'endroit où ces manifestations se déroulent le plus souvent et de façon croissante. Les données pour Leipzig montrent que le mois de septembre a été crucial, comme le mettent en évidence les études de Opp [Opp, 1991] [Opp et Gern, 1993] et Lohmann [Lohmann, 1994]. À partir de ce moment, le nombre de manifestants a augmenté progressivement de semaine en semaine. Parmi les autres lieux où ces manifestations ont eu lieu, on peut citer Berlin-est, Dresde, Karl-Marx-Stadt et Halle. Le nord du pays est une région plutôt agricole dans laquelle la densité de population est faible. Au contraire le centre et le sud sont des zones industrielles dans lesquelles la densité de population est importante. Dans ce contexte, l'absence de manifestations massives dans le nord de la RDA est tout à fait remarquable.

Il faut noter également l'absence quasiment totale de toute répression gouvernementale après le dix octobre (tableau 6b). Ceci traduit en fait l'existence de divergences de vues importantes au sein de l'élite au pouvoir. Certains ont proposé une stratégie de répression massive contre les manifestations de Leipzig. La direction générale du parti a préféré favoriser une stratégie non-violente. Ces divergences sont devenues évidentes à partir du remplacement de Honecker par Krenz le dix-huit octobre.

Le neuf novembre 1989 et la chute du régime communiste Le phénomène dont on cherche à réaliser une modélisation est l'évolution du comportement contestataire, c'est-à-dire l'évolution de la mobilisation de masse en RDA en 1989. Lorsque l'on cherche à réaliser une telle modélisation, il est important de délimiter dans le temps de manière assez précise le phénomène auquel on s'intéresse. Une telle délimitation n'est sans doute pas un impératif absolu quand on recherche uniquement une explication qualitative du phénomène. Il n'en est pas de même dans le cadre d'une modélisation. On peut sans doute distinguer trois aspects importants qui marquent la fin de la phase révolutionnaire. Le premier aspect est la fin de la mobilisation de masse contre le régime. Le deuxième aspect est la chute du régime communiste. Le troisième aspect est le changement de régime et l'établissement d'un nouveau régime. On peut se demander dans quelle mesure ces différents aspects sont susceptibles d'être identifiés les uns aux autres et s'ils ne doivent pas être plutôt considérés comme trois moments bien distincts. En fait il est évident qu'une réponse quantitative cherchant à dater ces différents moments comporte des aspects arbitraires. La difficulté est en partie liée à la nature même du changement. On peut se demander dans quelle mesure le nouveau régime ne fait pas aussi partie intégrante du processus révolutionnaire. Peut-on considérer au contraire que le processus de transition est un phénomène à part entière qui nécessite sa propre analyse distincte et systématique ? On considérera par la suite que la fin de la mobilisation de masse contre le régime, c'est-à-dire la fin du phénomène qui nous intéresse, date de l'ouverture du Mur le neuf novembre 1989.

Ce jour et l'événement qui lui est lié n'ont pas été choisis délibérément par le régime. En effet, cet événement est lié à une déclaration de Schabowski, le porte-parole du Bureau Politique du Parti. Schabowski a déclaré publiquement à la télévision, le soir du neuf novembre, que tout le monde pourrait prochainement voyager librement à l'étranger. En annonçant cela il avait mal compris la décision du bureau qui consistait uniquement à autoriser les départs à l'étranger de ceux qui voulaient émigrer. Suite à cet événement fortuit, la population a considéré que la politique de liberté de voyager était devenue la politique officielle. Le soir même tout Berlin-est se précipitait à l'ouest en forçant les barrages, et le Mur était détruit.

Le processus fondamental de démocratisation devenait alors inévitable. Pourtant on ne savait pas à ce moment comment le régime allait évoluer ni si la réunification de l'Allemagne était ou non une option réaliste. Tous ces éléments ont évidemment une très forte incidence sur la constitution de notre modèle. En effet les données elles-mêmes ne permettent pas de préciser à quel moment on peut parler de changement de régime. On peut cependant considérer qu'à partir d'un certain niveau de mobilisation de masse contre le régime on a atteint un point de non retour et que le régime sera effectivement renversé.

On peut donc résumer les faits en affirmant tout d'abord qu'une évolution à long terme du mécontentement en RDA a eu lieu, pour des raisons sociales, économiques

et aussi politiques. Ceci nous mène à l'affirmation qu'un potentiel révolutionnaire s'est développé dans les années quatre-vingts. Certains éléments de la situation politique internationale ont également joué un rôle déterminant, parmi lesquels on peut citer la politique de Gorbatchev. En effet, celui-ci avait déclaré officiellement qu'il n'interviendrait pas dans les affaires intérieures des autres membres du Pacte de Varsovie. L'évolution de la situation en Pologne a sans doute eu un rôle non négligeable également. En ce qui concerne la situation intérieure en RDA, il faut considérer comme un facteur déterminant le fait que les dissidents sont organisés en groupes disséminés dans tout le pays. Pourtant, tout ceci ne permet pas encore de parler du mécanisme de coordination qui s'est établi ni de la dynamique qui s'est mise en place.

La question qui se pose est de savoir si on peut avoir recours à *la théorie de la main invisible* pour expliquer la dynamique qui s'est développée au cours de cette année. Selon cette explication, le comportement rationnel individuel peut entraîner un comportement collectif inattendu. Dans ce cadre, une interaction collective peut avoir un résultat ou des conséquences qui ne sont pas prévus par ses propres acteurs. En RDA initialement, seul un petit nombre de personnes s'est mobilisé dans un contexte purement local. Les motivations étaient de deux sortes. Certains avaient la volonté de quitter le pays. D'autres souhaitaient forcer le régime à se réformer et à se libéraliser, mais ils n'envisageaient pas et ne souhaitaient pas pour autant un changement fondamental du régime. Ainsi les conséquences globales issues des comportements individuels ont été très différentes des attentes individuelles de l'époque. Aucun acteur n'avait souhaité ni même envisagé que ce mouvement pourrait avoir de telles conséquences. On ne peut donc même pas dire que le mouvement collectif a été détourné par une minorité.

Les mécanismes de répression tels qu'ils sont pratiqués habituellement ont parfaitement fonctionné pendant la première phase de mobilisation, comme le montre la comparaison des tableaux 6a et 6b. Les manifestations ont persisté pendant l'été après avoir été particulièrement importantes au mois de mai, du fait des fraudes électorales. Il est intéressant de noter également que les vacances du mois d'août ont stimulé le mouvement d'émigration après un court ralentissement de ce phénomène en juillet. À partir de ce moment, l'émigration et les manifestations se sont développées de façon tout à fait parallèle, se renforçant entre elles. La réponse des autorités a alors consisté en une politique de plus en plus répressive qui a culminé dans les événements du début du mois d'octobre à Berlin-est. La direction du Parti a décidé alors de changer de stratégie en évoluant d'une politique répressive vers une politique de libéralisation du régime (tableaux 6b et 6c). Ce revirement a revigoré le processus de mobilisation et d'émigration, au lieu de le calmer (tableau 6a).

On note également un processus de diffusion de la contestation à partir du triangle Berlin-est—Leipzig—Dresde. Il est clair qu'un changement structurel essentiel a

eu lieu, malgré les motivations individuelles ou celles des groupes impliqués qui ne recherchaient pas *a priori* un changement fondamental d'une telle nature.

2.1.2. Éléments de théorie politique permettant d'expliquer le phénomène

Comme l'ont fait remarquer de nombreux observateurs, le phénomène observé en RDA, même s'il a surpris la communauté internationale, peut être expliqué *a posteriori* à partir de quelques concepts théoriques clés. Les travaux de Hirschman, en particulier, permettent d'identifier les facteurs explicatifs fondamentaux. Ces facteurs sont susceptibles de permettre une description de l'évolution dynamique du phénomène même s'ils ne fournissent pas des éléments systématiques d'explication.

Chez Hirschman [Hirschman, 1993], lors de la phase révolutionnaire, différentes attitudes peuvent être suivies par la population. Chacun peut réagir par une absence de contestation (*loyalty*), la manifestation de son mécontentement (*voice*) ou l'exil (*exit*). La loyauté est l'attitude suivie par une partie de la population. Comme nous l'avons souligné, il s'agit d'une disposition qui n'est pas directement observable et qui se manifeste par une absence de contestation du régime. La manifestation est quant à elle une expression publique de désapprobation du régime. Dans le cas de la RDA, l'exil correspond à une émigration vers la RFA où les individus, comme nous l'avons déjà souligné, étaient assurés d'une certaine protection sociale.

On peut ajouter au travail de Hirschman les recherches de trois auteurs : Kuran, Opp et Lohmann. La question qui se pose pour eux est de déterminer comment fonctionne un schéma comportant trois termes, la loyauté, l'exil et la contestation. L'exil et la contestation se renforcent réciproquement alors que la loyauté diminue à la fois l'exil et la contestation. Ces auteurs ont essayé de décrire le choix rationnel des individus qui ont décidé ou non de s'engager, c'est-à-dire de participer à la contestation. Ils ne précisent pas cependant comment le processus dynamique de mobilisation de masse s'est effectivement déroulé dans le détail en RDA.

Kuran met au centre de la problématique précédente entre loyauté, exil et contestation le choix rationnel d'un individu. Selon lui ce choix individuel est fonction d'une part des préférences privées individuelles qui assurent son intégrité personnelle et d'autre part de ses préférences publiques qui mettent en jeu son image vis-à-vis des autres. Lorsque ces deux préférences ne sont pas compatibles, le choix d'un comportement actif a un coût qui permet d'expliciter la fabrication d'un seuil d'activation de préférences individuelles. En référence à Kuran on peut interpréter la diffusion du mécontentement à partir du triangle Berlin-est—Leipzig—Dresde comme un changement dans les comportements individuels suite au franchissement d'un seuil de mécontentement. Plus la population exprimait son opposition au régime, plus diminuait le caractère individuel de l'engagement qui consistait à s'associer au groupe levier révolutionnaire ("revolutionary bandwagon"). Ainsi l'interprétation de Kuran permet de rendre compte de la diffusion dans le temps

du phénomène de mobilisation de masse. Le mécanisme de préférence individuelle décrit par Kuran met en évidence un effet “pervers” du choix rationnel. En effet, à partir d'un certain seuil, ceux qui soutiennent le régime finissent par s'activer et s'associer avec l'opposition. Il réagissent en fait essentiellement en fonction d'un sentiment et d'une représentation collectifs.

Les travaux de Kuran sont intéressants, car ils apportent une perspective supplémentaire explicite en faisant référence à des facteurs à long terme et en tenant compte de la chute du régime elle-même. Kuran tient compte des raisons qui à long terme expliquent une opposition au régime [Kuran, 1989][Kuran, 1991b][Kuran, 1991a]. Le potentiel révolutionnaire est alors rassemblé sous le terme générique de **sentiment collectif**. Kuran entend par là une disposition qui peut être mobilisée au cours des événements, et en particulier l'expression publique de préférences individuelles comme la manifestation et l'exil.

Pour Opp, le choix rationnel est encore au centre. Cependant, ce choix n'est pas fonction uniquement d'un utilitarisme matériel, mais dépend également de responsabilités et de valeurs collectives.

Lohmann est également utilitariste, mais elle tient compte de groupes hétérogènes qui s'associent afin d'exprimer leurs préférences.

En résumé, on peut dire que les trois modèles théoriques proposés par Kuran, Opp et Lohmann essayent d'expliquer un processus de diffusion du comportement individuel dont Hirschman ne tient pas compte. Néanmoins, ces trois modèles s'appuient uniquement sur un processus de diffusion dans le temps. Ils abandonnent un autre aspect de ce phénomène de mobilisation de masse en RDA, la diffusion dans l'espace. Il s'agit là d'un des éléments clés dont la modélisation que nous proposons essaie de rendre compte.

2.1.3. Les données collectées

Les faits ont été analysés et des données ont été récoltées pour essayer d'affiner cette analyse. Nous allons réaliser ici un rapide survol de ces données qui seront utilisées pour décrire quantitativement les événements. Ces données nous serviront également pour construire le modèle et le valider. Nous tiendrons compte de ces différents travaux théoriques pour décrire les faits.

On peut classer les données collectées en deux types. Le premier sous-ensemble regroupe les variables qui décrivent les conditions socio-économiques, le second mesure les différents comportements observés.

Les données socio-économiques ont été obtenues sur les 226 districts de la RDA. Le district correspond à un découpage administratif suffisamment homogène pour permettre une catégorisation de ces districts en fonction des critères socio-économiques qui nous intéressent. La collecte des données est réalisée sur ces districts en

particulier pour réussir à analyser la dimension spatiale du phénomène. Les données ont été obtenues au bureau des statistiques de la RDA qui est devenu accessible à partir de 1990. Une première série de variables correspond aux conditions de vie, comme la taille moyenne des appartements, la proportion d'appartements avec douche ou bain dans le district. Une seconde série de variables représente des facteurs sociaux de type structurel, comme la taille de la population, les groupes d'âge ou le niveau moyen d'éducation. On trouve également dans ces données la proportion de personnes qui ont émigré en 1989.

Les autres données correspondent aux manifestations du comportement, en particulier l'exil et les manifestations. On peut ajouter à cela les actions de répression du gouvernement, que nous appelons les sanctions. Pour cela le mode de codification des événements est légèrement différent du mode classique défini par Taylor dans le *World Handbook of Social and Political Indicators II*. Pour plus de précisions, on pourra consulter [Eberwein *et al.*, 1991]. Les données correspondant à l'émigration ont été collectées sur une base mensuelle [Ulrich, 1990]. Au contraire les indicateurs pour les manifestations et les sanctions ont été obtenus sur une base quotidienne. Deux sources de renseignements ont servi de référence au départ pour construire les indicateurs des comportements : un quotidien berlinois, le *Tagesspiegel*, et l'équivalent allemand de Keesings Archive, l'*Archiv der Gegenwart*, qui décrit les faits et les événements politiques. Les problèmes liés à l'obtention de données avec ces types de sources sont bien connus. En particulier les données manquantes ont été comblées grâce aux archives de Radio Free Europe Munich. Trois types d'événements ont été pris en compte, **les manifestations**, la répression du régime sous la forme de **sanctions** prises par le gouvernement et la levée des **sanctions** par le gouvernement qui montre la libéralisation du régime. Pour une définition précise de ces termes on pourra se référer à [Eberwein *et al.*, 1990]. La liste des événements obtenus est récapitulée sur la figure 6.

Les différentes données permettent les observations suivantes. Tout d'abord le nombre total de manifestations par mois (tableau 6a) varie de 5 en août, période de vacances, à 122 en novembre. La seconde colonne en partant de la droite nous indique également qu'un nombre important d'événements ont été rapportés avec une information insuffisante sur le nombre de personnes effectivement actives, entre 1 et 44 événements par mois. Cependant l'information disponible révèle une certaine tendance de l'évolution du nombre de manifestations au cours de l'année 1989. Les sanctions prises par le régime (tableau 6b) comprennent différents types d'événements. Tout d'abord, on a dans la colonne de droite le nombre d'actes répressifs réalisés par les autorités. La seconde colonne à gauche correspond au nombre de personnes qui ont été arrêtées. De nombreux événements sont considérés comme manquants, soit que les arrestations n'aient pu être comptabilisées, soit qu'ils aient pris une autre forme, comme une censure par exemple. On constate en particulier qu'en novembre il n'y avait plus d'acte répressif. En fait toute répression s'est arrêtée à partir du

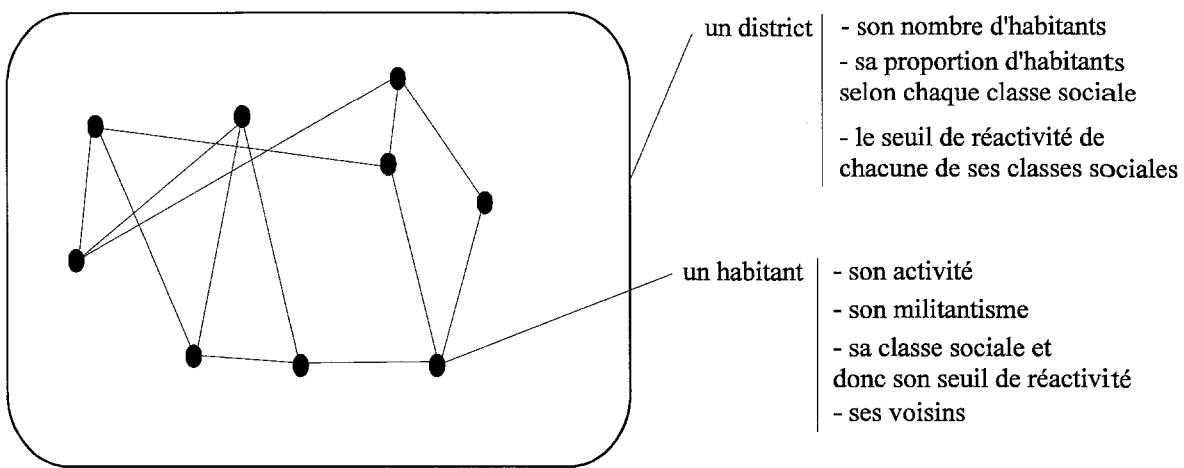


Schéma 8a : Les habitants d'un même district sont reliés et forment un graphe. Ils sont caractérisés par leur classe sociale et donc leur seuil de réactivité, ainsi que par leur activité et leur militantisme. La dynamique des habitants modifie l'activité des habitants. Elle est fonction de l'activité des voisins ainsi que du seuil de réactivité.

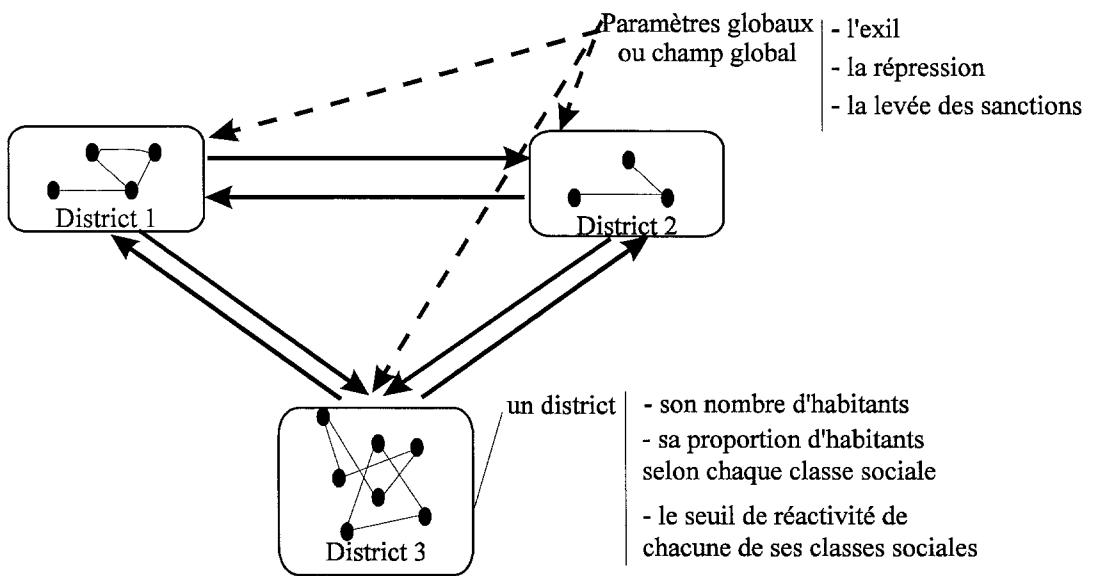


Schéma 8b : Les districts forment un graphe dont les liens sont pondérés par leurs distances géographiques. Ils sont caractérisés par leur nombre d'habitants ainsi que par leur proportion selon chacune des classes sociales. La dynamique des districts modifie les seuils de réactivité de chaque classe sociale au sein du district. Elle est fonction des paramètres globaux ainsi que de l'activité et de la distance géographique aux autres districts. Les paramètres globaux comprennent l'exil, la répression sous la forme du nombre de sanctions, ainsi que la levée des sanctions qui correspond à la libéralisation du régime.

FIGURE 8 : Présentation schématique des deux niveaux du modèle.

11 octobre. On constate le contraire pour la levée des sanctions et la libéralisation du régime (tableau 6c). Pour ce facteur, le nombre d'événements augmente considérablement à partir du mois d'octobre. Ces événements sont constitués soit par des libérations de prison soit par des amnisties.

Pour ce qui concerne le potentiel révolutionnaire, aucune donnée n'est accessible. On peut trouver de nombreuses données descriptives comme l'évolution des groupes d'opposition en RDA [Rink, 1991]. Ces sources ont permis d'obtenir des renseignements comme l'emplacement ou la taille de ces groupes, ainsi que des données sur leur constitution sociologique. Cependant ces données sont extrêmement difficiles à obtenir *a posteriori*. Cela amène certains comme Opp à s'interroger sur l'opportunité de modéliser ces processus de mobilisation de façon systématique. Opp a réalisé ce travail *a posteriori* par sondage, mais la méthode mise en œuvre et les résultats ainsi obtenus sont problématiques. Il est tout à fait notable qu'en fait seules certaines variables nous serviront à réaliser notre modèle. Ces variables auront été choisies avec une certaine part d'arbitraire, mais nous semblent suffisantes pour rendre compte du phénomène qui nous intéresse. Elles doivent également être d'un accès relativement aisé pour celui qui doit quantifier le phénomène étudié.

2.2. Le modèle proposé pour ce phénomène

Comme nous l'avons déjà souligné, le but du travail est désormais de modéliser le processus de mobilisation de masse en RDA au cours de l'année 1989. Ce modèle doit permettre de rendre compte de la dynamique globale résultant d'interactions locales. De plus le phénomène de mobilisation se déroule non seulement dans le temps mais aussi dans l'espace. Aussi le modèle doit être conçu pour rendre compte aussi bien de la dimension spatiale que de la dimension temporelle du phénomène. Ces propriétés sont satisfaites par toute une classe de modèles bien connus en physique, les modèles de transition de phase³. Ces types de modèles relèvent de ce que nous avons appelé les modèles de type Ising et que nous avons présentés dans le chapitre 3, même si des différences parfois substantielles sont susceptibles de les distinguer.

Ce modèle a été conçu au cours d'une étroite collaboration avec Wolf-Dieter Eberwein, spécialiste de sciences politiques au Wissenschaftszentrum für Sozialforschung Berlin. Puisque modéliser c'est choisir, il faut préciser que l'objectif affiché de la modélisation est de rendre compte du phénomène de mobilisation de masse, aussi bien dans le temps que dans l'espace. Ce modèle tente de tenir compte à la fois des hypothèses qui se dégagent des événements et des données et de notre connaissance des modèles de type d'Ising. S'il s'inspire des modèles de type Ising, il en est cependant une adaptation au phénomène étudié. Nous présenterons d'abord le modèle en langue naturelle avant d'en fournir une version plus formelle.

³Voir le chapitre 3 et plus particulièrement la section 1.

2.2.1. Présentation du modèle en langue naturelle

Ce modèle fait partie des modèles de type Ising. Son étude nous permettra d'identifier les problèmes qui se posent quand on cherche à appliquer les modèles de type Ising aux processus de mobilisation de masse et plus généralement aux processus qui mettent en jeu des phénomènes qui se situent à deux niveaux de description différents.

Le modèle proposé est à deux niveaux. Au premier niveau on trouve un ensemble d'éléments, les habitants, en interaction au sein de districts. Au second niveau on trouve un ensemble de districts eux-mêmes en interaction. La figure 8 résume la description du modèle et l'ensemble des paramètres qui interviennent.

L'unité "géographique" de base pour analyser le modèle est le district. Ce choix a été inspiré par la division administrative de la RDA en 226 districts. Nous avons de plus pu opérer ce découpage grâce au fait que les groupes de contestation sont distribués sur l'ensemble du territoire et ont une organisation locale. Il est alors raisonnable de considérer que les contestataires ne peuvent influencer directement que d'autres éléments à l'intérieur du même district.

Dans chaque district vivent un certain nombre d'habitants qui sont les unités de base du modèle. Nous cherchons à analyser la dynamique du processus de mobilisation de masse dans chaque district. Chaque habitant peut exprimer sa désapprobation du régime politique lorsqu'un seuil de mécontentement est dépassé. Les habitants sont susceptibles de manifester ce mécontentement, nous dirons de s'activer. Ils contribuent également à une influence sur les autres habitants du même district. Cette activation de chaque habitant dépendra non seulement de l'activation des autres habitants du même district, mais aussi d'un facteur global du système susceptible de modifier le seuil de réactivité. Ce facteur global provient de l'information que chaque habitant peut observer sur le fonctionnement global du système. En particulier ce facteur dépend du nombre total d'habitants qui ont choisi de s'exiler et de la répression que pourrait émettre le pouvoir sur l'ensemble du système. Le nombre total d'habitants de tous les districts qui expriment leur opposition au régime fournit le niveau global du phénomène de mobilisation de masse à chaque instant.

Les habitants

Les habitants de la RDA constituent les unités de base du système. On les regroupe en différentes classes grâce à des considérations sociologiques. Ils agissent entre eux, et s'influencent mutuellement au sein des districts pour s'activer ou non. Cette activation est considérée comme le fait d'être mécontent du régime et le montrer au cours de manifestations.

La classification des habitants La classification des habitants est réalisée à partir des données obtenues par différentes sources et qui sont recoupées. Les sources concernant les caractéristiques des habitants proviennent des archives officielles de l'ancienne RDA. Ces éléments ont ensuite été analysés avec différents outils statistiques. On observe alors qu'une classification en deux classes sociales est possible en fonction d'un critère qui correspond au niveau de formation des habitants. Ce critère constitue un facteur de discrimination systématique et fondamental qui permet de rendre compte de l'importance relative des manifestations selon les districts. La première classe correspond aux personnes qui sont allées à l'université, ou qui occupent des emplois d'encadrement ou de technicien supérieur. La seconde classe correspond aux ouvriers ou à des personnes ayant fait peu d'études. On observe également un recouplement de cette classification avec des critères de salubrité du logement.

Dans le modèle, le comportement des habitants dépend d'un seuil de réactivité. Ce seuil correspond à la propension à s'activer, c'est-à-dire à manifester, qui ne sera pas la même selon la classe sociale. Les deux classes d'habitants d'un même district se distinguent alors grâce à un seuil de réactivité différent pour chacune d'entre elles. Nous appellerons r_1 la réactivité de la première classe et r_2 celle de la seconde classe. Cette réactivité est la même pour tous les éléments d'une même classe sociale dans un même district, mais elle peut varier d'un district à l'autre.

Les caractéristiques des habitants Chaque habitant est caractérisé par sa classe sociale et donc sa réactivité, ainsi que par son activité. Ce paramètre permet de préciser si l'habitant fait ou non partie des manifestants. Il vaut 1 pour quelqu'un d'actif (et donc qui manifeste) et 0 si la personne est inactive, c'est-à-dire ne manifeste pas. Ces habitants sont également caractérisés par un troisième paramètre lié à leur appartenance ou non à un noyau dur. Dans ce cas on dira que ce sont des militants. On trouve à la fois des militants révolutionnaires qui font partie des groupes contestataires et qui resteront actifs quoi qu'il arrive et des militants réactionnaires qui resteront toujours inactifs.

De plus, à l'intérieur d'un même district, les habitants sont reliés les uns aux autres au sein d'un réseau. Ce réseau correspond aux liens que chaque habitant entretient avec d'autres habitants qui influencent son activation ou sa désactivation. Dans ce modèle tous les habitants ont le même nombre p de voisins. Ce réseau est constitué grâce à un tirage aléatoire. Dans un premier temps, ces voisinages sont fixes et ne changent pas au cours de la dynamique. On parle alors de graphe gelé en opposition à un graphe recuit lorsque les voisinages et non seulement l'activité des habitants changent eux aussi au cours de la dynamique.

En résumé, chaque individu est caractérisé par trois nombres, son seuil de réactivité lié à sa catégorie sociale, son activité et son militantisme. De plus à chaque indi-

vidu est associée une liste de voisins qui ont également une activité et une catégorie sociale.

La dynamique d'activation des habitants La dynamique d'activation des individus est locale. Elle ne dépend que des caractéristiques propres à chaque individu. Dans chaque district, cette dynamique s'opère de manière parallèle et synchrone sur les individus. Nous avons déjà vu dans le chapitre trois l'importance du choix d'une dynamique synchrone ou asynchrone. Comme dans notre modèle la dynamique est synchrone au sein d'un district, on peut à chaque pas de temps compter aisément le nombre d'habitants actifs. De même on peut déduire au niveau de l'ensemble des districts le nombre total d'habitants actifs de façon parallèle et synchrone.

Les individus qui sont militants restent toujours dans le même état d'activité. Pour les individus qui ne sont pas militants, on calcule la moyenne de l'activité de ses voisins. Lorsque cette moyenne est supérieure à son seuil de réactivité, cet habitant va s'activer ou rester actif avec une certaine probabilité. Lorsque cette moyenne est inférieure à son seuil de réactivité, il va se désactiver ou rester inactif avec une certaine probabilité.

Actuellement ces probabilités sont les mêmes quels que soient la catégorie sociale et le district d'appartenance de l'individu.

Les districts

Classification des districts et le cas de Berlin Les différents districts ont été également catégorisés à partir de données provenant des archives de la RDA, avec les outils statistiques classiques en sociologie.

Cette analyse met en avant une première catégorisation grossière en deux grands types de districts, les premiers correspondant grossièrement aux zones urbaines et les seconds aux zones rurales.

Pour Berlin, le problème qui se pose est légèrement différent. En effet, Berlin-est comportait à elle seule onze districts différents. Les données sociologiques peuvent être obtenues pour chacun des districts de Berlin-est. Le problème se pose par contre pour l'obtention des données des manifestations. Il est en effet tout à fait impossible lorsqu'une manifestation importante a lieu dans le centre de Berlin-est, de savoir de quel district berlinois proviennent chacun des participants. Il est donc préférable dans ce cas de considérer Berlin-est comme un seul district gigantesque.

Les caractéristiques des districts Le district est tout d'abord déterminé par son numéro, et le nombre de ses habitants.

Chaque district est caractérisé en fonction de sa classe par la répartition de sa population dans chacune des catégories sociales. Lorsque la proportion de la population du district selon chacune des classes sociales est fixée, on effectue un tirage aléatoire qui permet de fixer la catégorie sociale de chacun des habitants.

Chaque district est également caractérisé par la donnée des seuils de réactivité des différentes classes sociales r_1 et r_2 . Ces seuils de réactivité sont les mêmes pour chaque habitant d'une même classe sociale, mais ils changeront au cours de la dynamique des districts. De plus, à chaque pas de temps, on peut calculer pour chaque district, le nombre d'actifs.

La dynamique des districts et les facteurs géographiques. Un phénomène de réseau est alors introduit au niveau des districts. Actuellement la dynamique qui intervient entre les districts ne modifie que les seuils de réactivité. On pourrait envisager que cette dynamique modifie également le nombre d'habitants de chaque district, par exemple en retirant du district, à chaque pas de temps les habitants qui ont opté pour l'exil. Les seuils de réactivité des catégories sociales d'un district sont modifiées par deux éléments, d'une part le facteur global que nous verrons un peu plus loin et d'autre part par l'activité observée dans les autres districts de la RDA. plus précisément en ce qui concerne ce deuxième aspect, les seuils de réactivité du district sont modifiés en fonction de la variation de l'activité totale des autres districts pondérée par l'inverse de la distance à ces districts. Ceci permet d'introduire un effet de diffusion indirecte de l'activité des autres districts. De plus la pondération par l'inverse de la distance permet que cette diffusion dépende de la proximité géographique des districts.

Le facteur global

Un facteur global va également influencer les paramètres des districts et donc de manière indirecte l'activité des habitants en modifiant les seuils de réactivité des différents habitants.

Ses composants : l'exil, la répression et la libéralisation du régime Ce facteur global est constitué par trois éléments déterminés et mesurés par le spécialiste en sciences politiques. Le premier élément est l'exil qui est obtenu comme la variation du nombre d'individus qui ont quitté le territoire de la RDA entre deux pas de temps successifs. Le second élément est la répression qui est donnée par le nombre d'arrestations et d'emprisonnements politiques. Cette donnée est obtenue grâce aux archives de Radio Free Europe, à Munich. Le troisième élément est la levée des sanctions, qui est constituée, entre autres par le nombre de prisonniers politiques relâchés par le régime entre deux pas de temps.

Toutes ces données sont censées être connues par chaque individu par exemple par l'intermédiaire de la télévision ouest-allemande qui était reçue dans tout le pays.

Sa détermination par agrégation Les trois éléments que nous avons définis déterminent ensemble un facteur global qui correspond à une sorte de mécontentement global lié aux informations connues de tous. Cependant l'influence relative de chacun de ces éléments est difficile à déterminer. Dans un premier temps, nous considérerons une agrégation de ces éléments sous forme d'une combinaison linéaire, les différents coefficients en jeu étant fournis grâce à une analyse des données politiques.

Son effet sur les districts Ce facteur global a ensuite une influence sur les seuils de réactivité de chaque catégorie sociale et dans chaque district. Lorsque le facteur global augmente, cela entraîne une diminution des seuils de réactivité des catégories sociales.

2.2.2. Une version plus formelle du modèle

Notation

Pour tout $p \in \mathbb{N}$, on note \mathbb{N}_p l'ensemble $\{0, 1, \dots, p - 1, p\}$.

Soit $n \in \mathbb{N}$, n correspond au nombre de districts.

Soit $E = (E_1, E_2, \dots, E_i, \dots, E_n)$, avec les E_i , n ensembles finis, chaque E_i correspond à un district.

Description des E_i

\mathfrak{R}_i , une relation binaire sur E_i :

- 1) cette relation est *symétrique* : $\forall x, y \in E_i, x \mathfrak{R}_i y \iff y \mathfrak{R}_i x$
- 2) cette relation est *non réflexive* : $\forall x \in E_i, \neg(x \mathfrak{R}_i x)$

Cette relation binaire permet de définir les voisins de x . Pour tout $x \in E_i$, on peut définir la classe de x dans E_i , c'est-à-dire l'ensemble des voisins de x , $C_x \in (E_i)$. Soit $C_x = \{y \in E_i / x \mathfrak{R}_i y\}$.

On fait de plus l'hypothèse H_1 que toutes les classes C_x ont le même cardinal, $\exists p \in \mathbb{N}^*, \forall i, 1 \leq i \leq n, \forall x \in E_i, Card(C_x) = p$

Remarque : \mathfrak{R}_i n'est pas une relation transitive (sauf dans des cas dégénérés).

On définit une **fonction qui correspond à la classe sociale** de x , avec q classes sociales possibles, $q \in \mathbb{N}^*$:

$$\begin{aligned} f : \bigcup_{i=1}^n E_i &\rightarrow \mathbb{N}_q \\ x &\rightarrow f(x) \end{aligned}$$

On définit une **fonction qui correspond au militantisme** de x :

$$m : \bigcup_{i=1}^n E_i \rightarrow \{0, 1\}$$

$$x \mapsto m(x)$$

Description de E

\mathfrak{R} , une relation binaire sur E :

1) cette relation est *symétrique* : $\forall E_i, E_j \in E, E_i \mathfrak{R} E_j \iff E_j \mathfrak{R} E_i$

2) cette relation est *non réflexive* : $\forall E_i \in E, \neg(E_i \mathfrak{R} E_i)$

Cette relation binaire permet de définir les districts voisins.

On définit la **réactivité des classes sociales** :

$$g : E \rightarrow [0; 1]^q$$

$$E_i \rightarrow (r_i^1, \dots, r_i^q)$$

On définit la **classe des districts** :

$$h : E \rightarrow \mathbb{N}_{q'}$$

$$E_i \rightarrow h(E_i)$$

De même on définit les **voisinages C_i de chaque district E_i** :

Soit $C_i = \{E_j \in E / E_i \mathfrak{R} E_j\}$

$$d : C_i \rightarrow \mathbb{N}$$

$$E_j \rightarrow d(E_i, E_j)$$

Éléments dépendants de t

Soit $t \in T \subset \mathbb{N}$. On définit le **champ extérieur global $U(t)$** comme une combinaison linéaire (agrégation des termes globaux externes) avec $\alpha, \beta, \gamma \in \mathbb{R}^+$:

$e_1(t) \in \mathbb{R}^+$: *l'exil*, variable connue au cours du temps et fournie par les données.

$e_2(t) \in \mathbb{R}^+$: *la répression*, c'est-à-dire les sanctions, variable connue au cours du temps et fournie par les données.

$e_3(t) \in \mathbb{R}^+$: *la levée des sanctions*, variable connue au cours du temps et fournie par les données.

$$U(t) = \alpha e_1(t) + \beta e_2(t) + \gamma e_3(t).$$

On définit pour chaque individu son activité :

$$a : \bigcup_{i=1}^n E_i \times T \rightarrow \{0, 1\}$$

$$(x, t) \rightarrow a(x, t)$$

Dynamiques

Dynamique de l'activité Pour tout $x \in E_i$, on définit l'**activité de son voisinage** comme :

$$B_x(t) = \frac{\sum_{y \in C_x} a(y, t)}{Card(C_x)}$$

La dynamique de l'activité de tout individu x peut s'écrire en fonction de la réactivité de sa classe sociale dans le district :

$$\begin{cases} m(x) = 1 \implies a(x, t+1) = a(x, t) \\ m(x) = 0 \implies \begin{cases} P(a(x, t+1) = 1 / (a(x, t) = 0 \wedge B_x(t) \leq g_{f(x)}(E_i) = r_i^{f(x)}(t))) = p_0 \\ P(a(x, t+1) = 1 / (a(x, t) = 0 \wedge B_x(t) \geq g_{f(x)}(E_i) = r_i^{f(x)}(t))) = p_1 \\ P(a(x, t+1) = 1 / (a(x, t) = 1 \wedge B_x(t) \leq g_{f(x)}(E_i) = r_i^{f(x)}(t))) = p_2 \\ P(a(x, t+1) = 1 / (a(x, t) = 1 \wedge B_x(t) \geq g_{f(x)}(E_i) = r_i^{f(x)}(t))) = p_3 \end{cases} \end{cases}$$

Dynamique des réactivités La dynamique des réactivités peut s'écrire en fonction du champ global :

$$\forall i \in IN_q, \forall t \in T^*, \\ r_i(t) = r_i(t-1) + K_i(U(t) - U(t-1)) + K'_i \left(\sum_{E_j \in C_i} \frac{1}{d(E_i, E_j)} (H_{E_j}(t) - H_{E_j}(t-1)) \right)$$

Initialisation du modèle

Le graphe est initialisé comme un **graphe aléatoire**. On choisit au départ une proportion de militants assez faible (0, 1). On choisit également une proportion d'actifs assez faible (0, 1). On définit les proportions sociales pour les districts :

$$PROP = \{(prop_1, \dots, prop_q) \in [0; 1]^q / \sum_{i=1}^q prop_i = 1\} \\ s : IN_{q'} \rightarrow PROP \\ j \rightarrow prop$$

On choisit des valeurs pour les probabilités de retournement ; on prend pour p_0 une petite probabilité (0, 1), pour p_1 une probabilité moyenne (0, 5), et pour p_2 et p_3 de grandes probabilités (0, 9)

Mesure globale sur le système

On définit l'énergie totale du district comme :

$$H_{E_i}(t) = \sum_{x \in E_i} a(x, t)$$

On définit l'énergie totale du système comme :

$$H(t) = \sum_{i=1}^n H_{E_i}(t)$$

On aurait pu définir l'énergie du district comme :

$$H_{E_i}(t) = \sum_{x \in E_i} B_x(t)$$

2.3. La simulation informatique et son interprétation

Afin de poursuivre la modélisation telle que nous l'avons présentée sur la figure 2 de ce chapitre, nous avons programmé une simulation du modèle précédent. Celle-ci a été effectuée en Visual Smalltalk⁴, un environnement de programmation objet qui convient particulièrement bien pour ce type de modèles.

L'architecture du modèle à deux niveaux (figure 8) met en jeu une collaboration entre districts. Afin d'analyser la dynamique du processus de mobilisation de masse dans chaque district, nous avons étudié dans un premier temps la simulation de ce modèle pour un unique district. Il nous a semblé en effet plus pertinent de commencer par travailler sur un seul district afin de bien comprendre comment se comportait la simulation. Cette étude détaillée est de toute façon nécessaire afin d'élaborer plus précisément la dynamique entre districts.

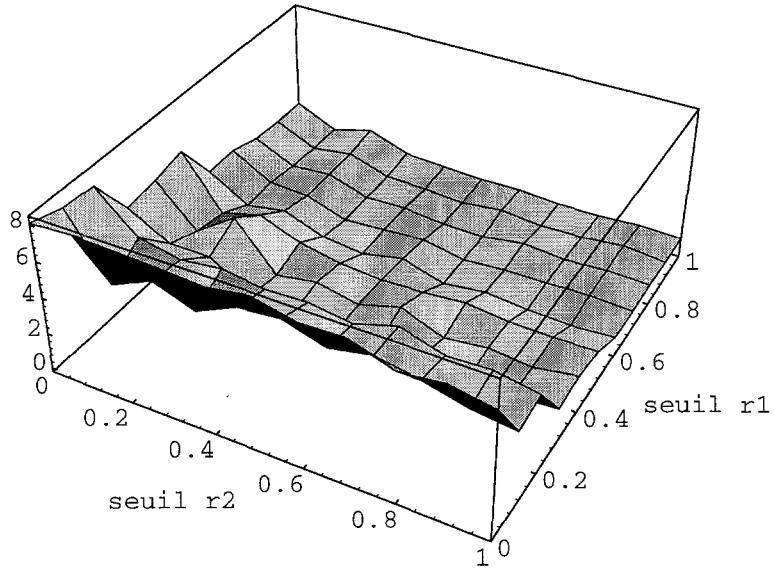
2.3.1. Description de la simulation pour un district

La simulation pour un district nécessite d'abord de fixer le nombre d'habitants du district. Les nombres possibles pour la simulation sont hélas petits du fait de la limitation de la puissance des machines. Nous précisons dans la section suivante comment on peut tout de même espérer obtenir des informations sur le comportement de districts comprenant un grand nombre d'habitants à partir de simulations comportant peu d'habitants. Les simulations ont été faites avec 9, 99 et 999 habitants.

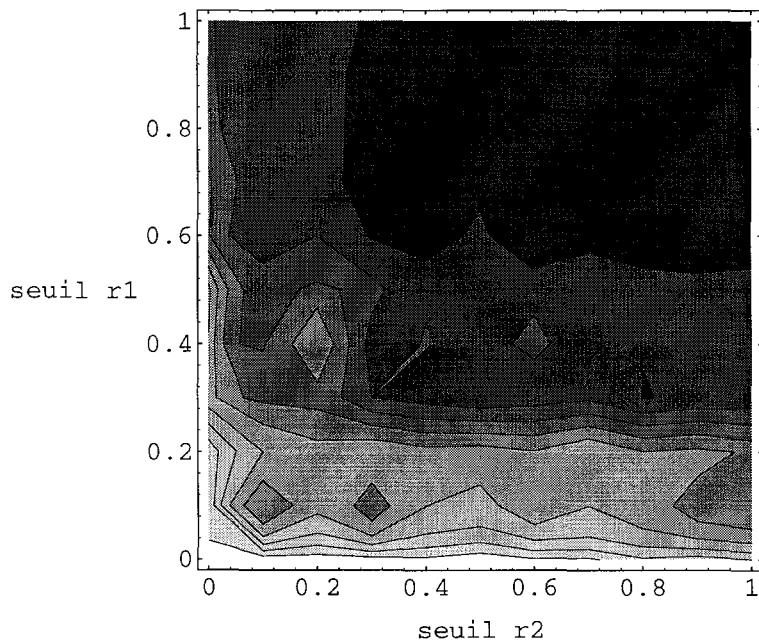
Lorsque l'on a fixé le nombre d'habitants, il faut ensuite fixer la proportion sociale du district. Celle-ci dépend, comme nous l'avons vu, de la classe du district. Cependant, dans nos simulations, nous avons étudié une unique classe de district et nous avons donc fixé une seule proportion sociale. Cette proportion sociale est de 80% pour la première classe sociale et de 20% pour la seconde.

De même nous nous sommes restreints à une unique proportion initiale, pour les militants, dont la valeur est de 10%. Parmi ces militants nous avons fixé à 10% le

⁴Visual Smalltalk 3.0.1 est une version de Smalltalk pour PC sous Windows 3.11.



Graphique 9a : Représentation d'une “carte d’activité” d’un district. Elle permet de visualiser selon l’axe des z la moyenne de la moyenne de l’activité d’un district pour des valeurs des seuils de réactivité variant de 0 à 1. Le district comprend ici 9 habitants, 4 voisins par habitant et 10 itérations par essai. La carte a été construite avec 121 points (11x11). La moyenne en chaque point a été calculée sur 100 essais.



Graphique 9b : Même “carte d’activité” d’un district que sur le **Graphique 9a**. On a choisi ici une représentation en lignes de niveaux comme pour les cartes de géographie (dites cartes d’état-major). Il s’agit donc d’une vue verticale “du dessus” du graphique précédent. Les zones claires correspondent aux zones pour lesquelles la moyenne de la moyenne de l’activité est élevée alors que les zones sombres rendent compte de celles pour lesquelles cette valeur est faible.

FIGURE 9 : Deux représentations différentes d'une même “carte d'activité”.

nombre de militants actifs (noyau dur contestataire, soit 1% des habitants) et donc à 90% les militants inactifs (noyau dur communiste, soit 9% des habitants).

Nous avons ensuite fixé le nombre de voisins. Les simulations ont été réalisées avec trois, quatre et dix voisins. On définit ensuite par un tirage aléatoire un graphe de connexion des habitants au sein du district. Le codage des habitants a été optimisé. Cependant, comme le graphe est aléatoire, quand on code un habitant il faut aussi retenir le numéro de tous ses voisins. Le codage a été fait en mémorisant un entier⁵ par habitant, les premiers chiffres correspondant aux numéros des voisins et le dernier codant à la fois la classe sociale, le militantisme et l'activité de l'habitant. L'obligation de mémoriser les voisins fait toute la différence entre la simulation de modèles où l'on prend un graphe, comme c'est le cas ici ou pour les réseaux de neurones, et la simulation de modèles où l'on choisit un réseau régulier (réseau triangulaire, réseau carré⁶). Pour les réseaux réguliers il est inutile de mémoriser les numéros des voisins car on peut les recalculer dès que cela est nécessaire. Mais dans ce cas on est en fait obligé de considérer un élément externe au réseau sous la forme d'une fonction qui calcule les numéros des voisins. On perd alors l'aspect "local" du modèle puisque la simulation impose l'utilisation d'un "module" de calcul externe aux éléments du réseau. Ceci explique que les simulations de réseaux réguliers peuvent être faites avec un nombre important d'éléments. Au contraire nos simulations, comme celles faites pour les réseaux de neurones, ne peuvent que difficilement dépasser le millier d'éléments. Il est évidemment très gênant de simuler un modèle avec une dizaine, une centaine, voire un millier d'éléments, alors qu'on le justifie par une théorie qui nécessite la limite thermodynamique (voir le chapitre 3). Ce problème se pose évidemment dans les mêmes termes pour les modèles de réseaux de neurones dont les propriétés "émergentielles" s'appuient sur les mêmes fondements théoriques⁷. On trouve ici une des raisons de fond qui nous ont amené à envisager une complexité au-delà de la complexité/quantité (voir le chapitre 5), et en particulier une autre simplification de la complexité que celle qui impose de définir un système comportant un nombre infini d'éléments (voir les paragraphes 10 et 11 de la section 2 du chapitre 5).

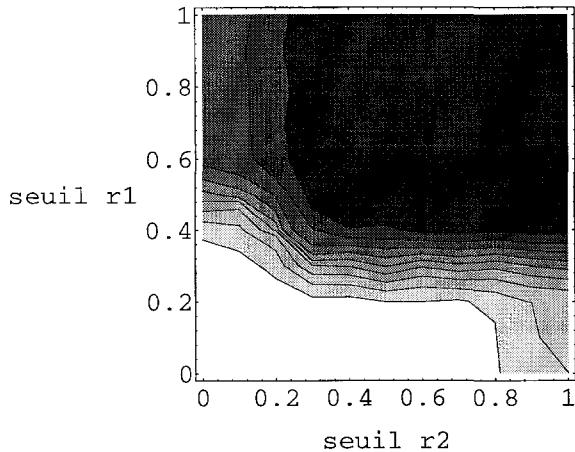
Avant de pouvoir mettre en jeu les dynamiques, il reste encore à préciser les valeurs des seuils de réactivité qui sont comprises entre 0 et 1.

On peut alors mettre en œuvre la dynamique de l'activité des habitants au

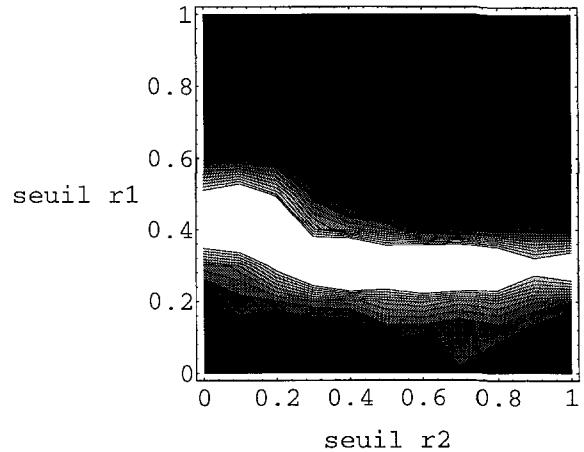
⁵Nous avons d'abord vérifié que Visual Smalltalk gérait les entiers de manière optimale et en particulier les "grands entiers".

⁶Voir la figure 1 du chapitre 3.

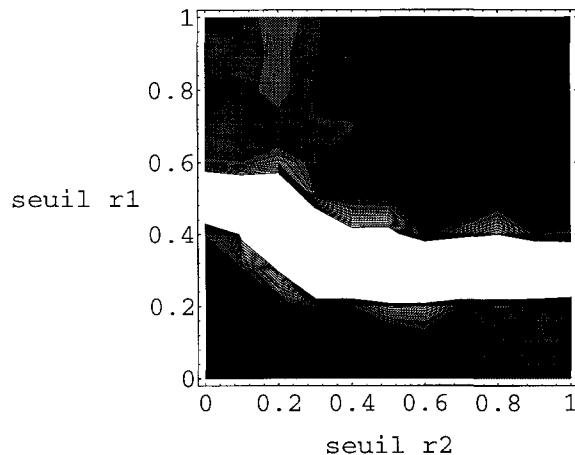
⁷Sinon on se retrouve devant des modèles qui expliquent uniquement les capacités de traitement de l'information, mais ne disent rien des propriétés "émergentielles" que l'on prête au cerveau. Pour une étude plus approfondie des limites d'une théorie qui explique uniquement la capacité de traitement de l'information sans offrir d'autre argumentation pour expliquer la notion de sujet, on pourra consulter la thèse de Patrice Prez [Prez, 1997] intitulée *De la vision à l'acte de voir : un parcours technique et épistémologique dans les sciences cognitives*.



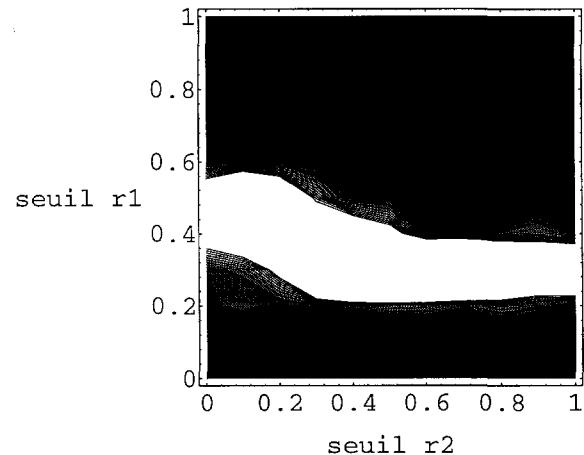
Graphique 10a : Moyenne de la moyenne de l'activité. On distingue nettement les deux phases (phase de forte activité en blanc et phase de faible activité en noir) et la zone de transition entre ces phases.



Graphique 10b : Moyenne de la variance de l'activité pour le même district. La zone en blanc correspond aux fortes valeurs de la moyenne de la variance et à la zone de transition du **Graphique 10a**. Les deux zones en noir correspondent aux deux phases.



Graphique 10c : Variance de la moyenne de l'activité pour le même district. La zone en blanc correspond aux fortes valeurs et à la zone de transition du **Graphique 10a**. Les deux zones en noir correspondent aux deux phases.



Graphique 10d : Variance de la variance de l'activité pour le même district. La zone en blanc correspond aux fortes valeurs et à la zone de transition du **Graphique 10a**. Les deux zones en noir correspondent aux deux phases.

FIGURE 10 : Mise en évidence de la “transition entre deux phases” pour une simulation d'un district comprenant 99 habitants, 10 voisins et 10 itérations à partir de la moyenne de la moyenne de l'activité, de la moyenne de la variance, de la variance de la moyenne et de la variance de la variance. Les quatre cartes ont été construites avec 121 points (11x11). Les coordonnées des différents points ont été établies à partir de 100 essais. La zone de transition entre les phases observée sur le graphique 10a pour la moyenne de la moyenne est confirmée comme étant la zone des plus fortes variations de l'activité grâce aux graphiques 10b, 10c et 10d.

Chapitre 4 -- Figure 10

sein du district. On fixe le nombre d'itérations, c'est-à-dire d'applications de cette dynamique de manière synchrone sur l'ensemble des habitants du district. Nous avons choisi différentes valeurs du nombre d'itérations entre 10 et 1000. Lorsque ce nombre d'itérations a été réalisé, l'essai est terminé et on peut définir un nouveau graphe aléatoire pour effectuer un nouvel essai.

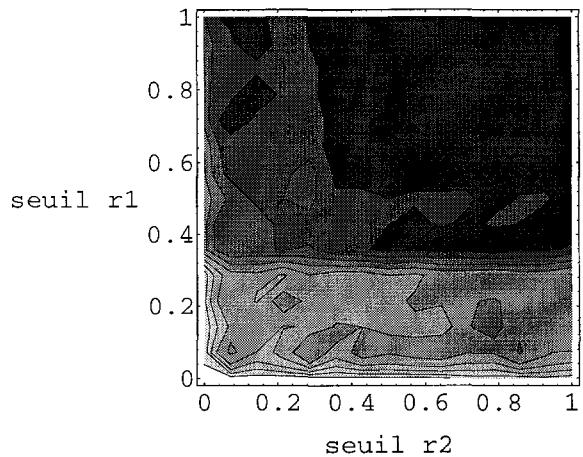
Au cours de l'essai, on calcule l'activité du district après chaque itération. Ceci permet de calculer la moyenne de l'activité à la fin de l'essai. On calcule de même la variance de l'activité à la fin de l'essai. Ceci nous permet de savoir si l'activité a subi de fortes variations ou non au cours de l'essai.

On réalise ensuite plusieurs essais à la suite, pour savoir si ce qui a été observé sur un essai se répète et comment cela se répète. Pour cela on calcule la moyenne de la moyenne de l'activité sur un essai, la variance de cette moyenne, ainsi que la moyenne de la variance et la variance de la variance. Tous ces indicateurs vont nous permettre d'étudier le comportement de la simulation sur un essai en fonction des différents paramètres qui nous intéressent.

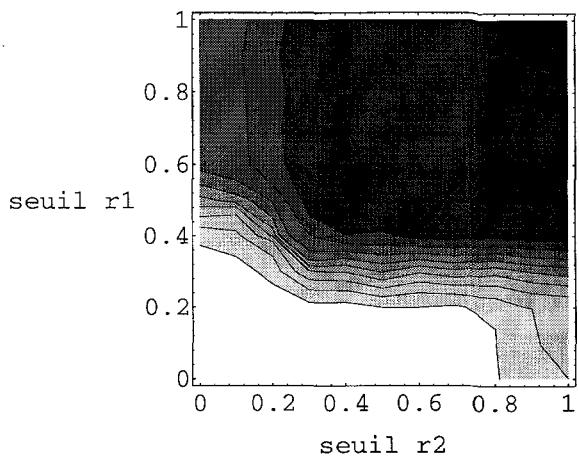
Pour analyser plus en détail cette simulation, il faut effectuer ces essais successifs pour différentes valeurs des seuils de réactivité de chacune des classes sociales. Plus précisément, nous avons fait des simulations avec des essais pour des valeurs des seuils de réactivité variant de 0 à 1. Comme on a deux classes sociales et donc deux seuils de réactivité, ils font office d'axes de coordonnées. On peut ainsi construire des "cartes d'activité" (voir figure 9), dont la valeur selon le troisième axe est donnée soit par la moyenne de la moyenne de l'activité, soit par la variance de la moyenne, soit par la moyenne de la variance, soit encore par la variance de la variance. Comme pour les cartes de géographie, nous avons privilégié une représentation en courbes de niveaux, la hauteur des lignes de niveaux correspondant à la valuation selon le troisième axe (figure 9).

2.3.2. Existence des différentes phases

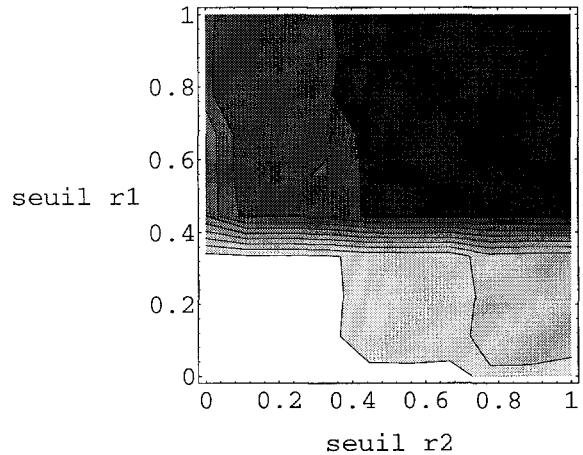
La première question qui se pose est de savoir si la simulation permet ou non l'observation de différentes phases et d'une transition entre ces phases. Dans notre cas, les différentes phases que la dynamique peut produire sont essentiellement deux types opposés et extrêmes de comportements pour le système, d'une part un état du système où presque tous les éléments sont actifs et d'autre part un état du système où presque aucun élément n'est actif. Les premiers résultats des simulations sont concluants et montrent que pour les valeurs des paramètres qui ont été choisies, on observe bien une "transition de phase". On observe nettement les deux phases sur les cartes d'activité de la moyenne de la moyenne de l'activité (voir graphique 10a). Chacune d'entre elles correspond à une zone de couleur uniforme soit noire, pour la phase de faible activité, soit blanche pour la phase de forte activité.



Graphique 11a : Moyenne de la moyenne de l'activité pour la simulation d'un district comportant 9 habitants, 3 voisins et 10 itérations.



Graphique 11b : Moyenne de la moyenne de l'activité pour la simulation d'un district comportant 99 habitants, 10 voisins et 10 itérations.



Graphique 11c : Moyenne de la moyenne de l'activité pour la simulation d'un district comportant 99 habitants, 3 voisins et 10 itérations.

FIGURE 11 : Comparaison des “cartes d’activité” obtenues lorsque l’on modifie le nombre d’habitants et le nombre de voisins. Un argument théorique laisse penser que le facteur déterminant serait de la forme $E(\ln(m/p)/\ln(p-1)) + 2$, avec m le nombre d’habitants du district et p le nombre de voisins. Cela justifierait une proximité plus grande, entre la zone de transition du graphique 11b avec 99 habitants et 10 voisins et celle du graphique 11a avec 9 habitants et 3 voisins, qu’entre le graphique 11c avec 99 habitants et 3 voisins et le graphique 11a. On observe effectivement une zone de transition marquée en forme de S renversé pour les graphiques 11a et 11b, mais le résultat n’est pas suffisamment net pour que l’on puisse en tirer des conclusions radicales.

Après avoir observé les deux phases distinctes, il est intéressant d'étudier plus en détail la zone de transition entre ces phases. Cette zone intermédiaire entre les phases est remarquable par le fait que les lignes de niveaux y sont très serrées, la transition entre phases étant brutale. L'étude de cette zone de transition est plus aisée sur la carte de la moyenne de la variance de l'activité (graphique 10b), sur celle de la variance de la moyenne de l'activité (graphique 10c) ou encore sur celle de la variance de la variance de l'activité (graphique 10d). Sur ces trois graphiques 10b, c, d, les zones des fortes valeurs représentées en blanc correspondent également aux zones des fortes variations de l'activité qui dénotent la zone de transition. On remarque que pour ces trois graphiques la zone de transition est identique et qu'elle correspond bien au domaine du graphique 10a pour lequel on a des lignes de niveaux très resserrées.

2.3.3. Variations des transitions en fonction des paramètres du modèle

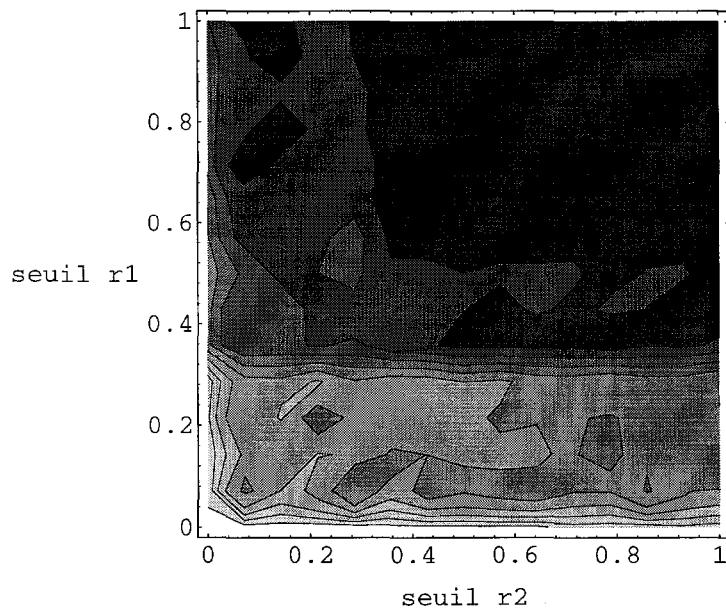
Estimation des seuils possibles pour les phases

Grâce aux simulations, on peut en fait être beaucoup plus précis sur la délimitation des domaines spécifiques des différentes phases. En effet, les transitions semblent concentrées autour de seuils. Des paliers sont particulièrement marqués sur le graphique 9b ou sur le graphique 11c. Plus généralement, lorsque l'on fait une simulation avec trois voisins, on a deux valeurs importantes qui jouent un rôle de seuil pour les seuils de réactivité. Ces valeurs sont $1/3$ et $2/3$. En effet si le seuil de réactivité est inférieur à cette première valeur de $1/3$, l'habitant pourra s'activer, avec une probabilité importante, s'il possède un unique voisin actif. Entre les deux valeurs $1/3$ et $2/3$ un habitant a besoin de deux voisins actifs pour s'activer lui-même. Au-dessus d'une valeur de $2/3$ tous les voisins d'un habitant doivent être actifs pour que cet habitant s'active⁸. Pour les mêmes raisons on s'attend également à observer des paliers pour quatre voisins aux valeurs $1/4$, $1/2$ et $3/4$. Pour dix voisins ces paliers se situeraient aux valeurs $1/10$, $1/5$, $3/10$, $2/5$, $1/2$, $3/5$, $7/10$, $4/5$, et $9/10$. Cependant ces paliers n'apparaissent clairement ni pour quatre voisins ni pour dix voisins (graphique 11b). En fait ces paliers, s'ils jouent un rôle, ne sont cependant pas tous de la même importance. On peut s'en féliciter car des paliers tous identiques seraient incompatibles avec un phénomène de transition de phase.

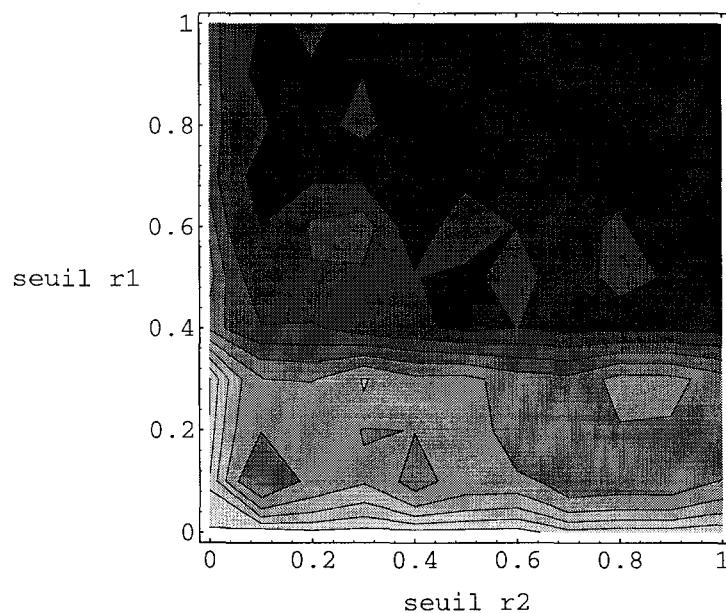
Le rôle du nombre des voisins

On cherche à étudier le comportement de la simulation après renormalisation des voisinages, c'est-à-dire les variations des résultats obtenus dues à une modification du

⁸En fait la dynamique est un peu plus subtile dans la mesure où l'activation n'est pas automatique mais a lieu avec une probabilité importante (voir les probabilités p_0 , p_1 , p_2 , p_3 dans la description plus formelle du modèle de la section précédente).



Graphique 12a : Moyenne de la moyenne de l'activité pour la simulation d'un district comportant 9 habitants, 3 voisins et 10 itérations.



Graphique 12b : Moyenne de la moyenne de l'activité pour la simulation d'un district comportant 9 habitants, 3 voisins et 100 itérations.

FIGURE 12 : Comparaison des “cartes d’activité” obtenues lorsque l’on modifie le nombre d’itérations pour un même district. Le domaine de chaque phase conserve une allure très similaire. On observe cependant une plus grande homogénéité des domaines correspondant à chacune des phases. De plus cette augmentation semble stabiliser la dynamique.

nombre de voisins et du nombre total d'habitants dans le district. Pour cette étude et par souci de simplicité, on pourrait être tenté de supposer que le comportement de la simulation reste le même lorsque le nombre de voisins ne change pas et que l'on augmente le nombre d'habitants. Cette hypothèse est vraisemblable dans la mesure où l'on est ici en présence d'un graphe aléatoire. Cependant la simulation nous permet de montrer que cette hypothèse n'est pas fondée, les graphiques 11a et 11c ne mettant pas en évidence les mêmes domaines de transition. Dans le premier cas on a 9 habitants et trois voisins alors que dans le deuxième on a 99 habitants et également 3 voisins. Par contre la simulation avec 9 habitants et trois voisins semble proche de celle à 99 habitants et 10 voisins (graphique 11b).

En fait il semble bien que le nombre de voisins joue le rôle d'un facteur de dilution. Par exemple, si on choisit de fixer les voisinages à trois voisins pour une population de 9 habitants, chaque habitant se situe à une distance maximale (en probabilité, dans la mesure où le graphe est aléatoire) de trois liens par rapport à n'importe quel autre habitant du district. Si on prend le même nombre de voisins dans une population de 99 habitants, la distance maximale entre deux habitants sera de sept. En fait on peut plus généralement considérer que la distance maximale sur un graphe aléatoire à p voisins est la même que celle que l'on obtiendrait⁹ sur un arbre à p branches, dont chaque sous-branche aurait elle-même $p - 1$ sous-branches, dont chacune aurait de nouveau $p - 1$ sous-branches, etc. Pour un tel arbre le nombre d'éléments est égal à $p \times (p - 1)^k$ et la distance maximale entre deux éléments est égale à $k + 1$. On peut en déduire selon ces hypothèses que si l'on prend un graphe aléatoire à m habitants et p voisins, la distance maximale entre deux habitants est égale à $E\left(\frac{\ln\left(\frac{m}{p}\right)}{\ln(p-1)}\right) + 2$, avec $E(x)$ la partie entière de x . Si l'on souhaite obtenir une distance maximale identique pour une population de 9 habitants et trois voisins et pour une population de 99 habitants, il faut choisir un graphe avec des voisinages comportant 10 voisins. On pourra comparer les graphiques 11a et 11b qui semblent confirmer la pertinence de cette distance maximale. Il est donc possible que cette valeur de $E\left(\frac{\ln\left(\frac{m}{p}\right)}{\ln(p-1)}\right) + 2$, pour un district comportant m habitants et p voisins, permette de classifier le comportement des districts.

Le rôle du nombre d'itérations

Lorsque l'on change le nombre d'itérations sans modifier les autres paramètres, les courbes de niveaux semblent très similaires. Dans les exemples reproduits sur les graphiques 12a et 12b nous avons modifié le nombre d'itérations de 10 à 100. Ceci ne change ni le comportement global du système ni les domaines des différentes phases. Les domaines semblent plus homogènes lorsque le nombre d'itérations augmente.

⁹On notera qu'une fois de plus pour avoir un résultat mathématique exact on est obligé de faire l'analyse mathématique d'un modèle approché et non du modèle initial.

Le système réagit comme si l'augmentation du nombre d'itérations ne modifiait pas la nature de la phase mais la confirmait. En somme, il semble que l'on obtienne une plus grande stabilisation du système et de ses différents domaines lorsque le nombre d'itérations augmente. On observe également qu'un nombre somme toute assez faible d'itérations permet déjà de se faire une idée correcte du comportement du système.

2.3.4. Quelques remarques supplémentaires sur les cartes d'activité

Sur l'ensemble des cartes on peut observer une asymétrie des mesures de l'activité par rapport à la première bissectrice. Cette asymétrie est due à la proportion des classes sociales. Cette proportion elle-même n'est pas symétrique puisqu'elle est de 80% pour la première classe et 20% pour la seconde. Pour obtenir des cartes symétriques il aurait fallu que la proportion sociale soit elle-même symétrique, c'est-à-dire ici de 50% pour la première classe et de 50% pour la seconde. Ceci montre que dans la simulation l'appartenance à une classe sociale joue bien un rôle important sur le comportement qualitatif du système dans son ensemble.

Le dernier point à noter est l'utilisation qui sera faite ultérieurement de ces cartes. Elles ont été établies pour un seul district en fonction des seuils de réactivité de chacune des classes sociales. Lorsque la deuxième dynamique intervient entre les districts, ou du fait du champ global, les seuils de réactivité changent. Ceci revient à une modification de l'abscisse et de l'ordonnée sur la carte. La seconde dynamique sera alors un déplacement sur cette surface. On peut envisager, pour cette dynamique de modification des seuils de réactivité, une échelle de temps différente de celle choisie pour la dynamique de l'activité au sein des districts. Seules importent en fait les valeurs relatives de ces deux échelles de temps. On peut déjà faire des simulations de ces deux dynamiques, mais seule la validation permettra de préciser les valeurs relatives pertinentes entre ces deux échelles de temps.

Ceci nous amène à la fin des résultats obtenus actuellement pour la simulation. Les résultats obtenus jusque là grâce à la simulation peuvent sembler triviaux, mais il n'en est rien. En effet ils permettent d'affirmer sans conteste l'existence de transitions de phase. Si cela n'avait pas été le cas on aurait été irrémédiablement amené à abandonner la structure du modèle et à remettre en cause par exemple le choix d'un graphe aléatoire gelé. De plus ces résultats nous permettent d'attester que le système est sensible non seulement au nombre de voisins mais encore au nombre total d'habitants dans le district. Ceci permet de mettre en évidence l'inévitable problème de la comparaison du comportement de deux districts qui diffèrent par leur taille. Les simulations nous ont également permis d'établir que plus on augmentait le nombre d'itérations et plus on stabilisait l'ensemble du système.

3. POUR ALLER PLUS LOIN DANS LA PRATIQUE DE LA MODÉLISATION

Cette dernière section se présente sous la forme d'une conclusion. Elle a deux objectifs. Le premier est de préciser comment il est possible de poursuivre le travail déjà réalisé. Le deuxième est d'insister sur l'intérêt de ce travail dans le cadre des sciences cognitives.

Nous avons insisté dans la première section de ce chapitre sur le fait que la modélisation était une activité. Ceci signifie en particulier que l'avancement de la modélisation dans chacun des domaines distingués ne peut pas être considéré comme définitif tant qu'il ne l'est pas pour les autres domaines. Les modifications qui affectent un des domaines de la modélisation affectent également les autres domaines. Prenons comme exemple la classification des 226 districts qui dépendra plus particulièrement du comportement de ses habitants pendant la période qui nous intéresse (nombre de groupes de contestataires, nombre de manifestants, nombre de manifestations, nombre de personnes qui choisissent l'exil, etc.) et également de sa composition sociale (salubrité des habitations, niveau d'éducation, ruralité du district, etc.). Une première analyse de données permet d'obtenir plusieurs classifications avec pour chacune d'entre elles un nombre de classes différent selon l'importance relative attribuée à chacun des facteurs qui interviennent. Les résultats obtenus pour la simulation avec un district, par exemple, vont influencer le choix d'un nombre de districts et d'un nombre de classes de districts à envisager pour la simulation avec plusieurs districts. Or déterminer ce nombre de districts et de classes de districts va avoir également des répercussions sur l'analyse de données elle-même et en particulier sur la classification des 226 districts en un nombre de classes déterminé. Ainsi l'avancement dans un des domaines de la modélisation est étroitement lié à l'avancement dans les autres domaines et le travail de modélisation consiste aussi en une gestion de ces interactions.

La deuxième section présente l'état actuel de la modélisation de la mobilisation de masse en RDA en 1989. On ne peut donc pas y déceler facilement les pistes qui ont été écartées ni les interactions qui ont eu lieu entre les domaines. La retranscription de cette modélisation est présentée sous forme écrite, donc linéaire, et permet difficilement d'observer les interactions permanentes et nécessaires qui ont eu lieu entre les différents domaines de la modélisation.

Puisque cette modélisation n'est pas achevée à l'heure actuelle, il convient de

présenter brièvement l'ensemble des travaux qui doivent encore être réalisés dans chacun des domaines qui nous concernent, ainsi que les conséquences éventuelles que ces travaux peuvent avoir sur les autres domaines.

3.1. Travaux à réaliser

Il convient tout d'abord d'insister sur les travaux à entreprendre pour poursuivre la modélisation.

3.1.1. Analyse du modèle

La première chose à faire serait sans doute d'établir l'approximation en champ moyen du modèle précédent. Ceci permettrait d'obtenir un nouveau modèle approché. Comme nous l'avons déjà précisé, ce travail consiste en fait à faire une analyse mathématique d'un modèle approché. Cette approximation permet d'obtenir les valeurs théoriques des seuils où les transitions éventuelles apparaissent. On pourra ensuite comparer ces valeurs théoriques des seuils aux valeurs numériques obtenues grâce à la simulation du modèle exact.

L'analyse mathématique peut également être réalisée en comparant le modèle proposé au modèle du votant (*voter model*). Ce modèle est présenté dans [Liggett, 1985]. Il a déjà été étudié pour lui-même comme exemple de processus markovien [Malyshev *et al.*, 1995]. On connaît son comportement. L'ensemble des éléments laisse présager de fortes et profondes ressemblances avec le modèle que nous avons proposé pour la mobilisation de masse en RDA.

3.1.2. Analyse de la simulation

Un gros travail reste encore à faire sur la simulation. La première chose consiste à caractériser *la nature de la transition*. Ceci est possible en calculant à partir de la simulation les valeurs des exposants critiques. Bien entendu les exposants obtenus dans ce cadre presupposent que la transition est caractérisée par des lois puissances, comme c'est toujours le cas lorsque l'on calcule des exposants critiques à partir de simulations. Comme nous l'avons déjà souligné, cette attitude peut présenter des risques puisque certaines transitions sont par exemple caractérisées par des lois logarithmiques et non des lois puissances.

Il faut noter que nous avons choisi volontairement un modèle avec un graphe aléatoire afin de faciliter la caractérisation de la transition à partir de la simulation. En effet, le choix d'un graphe aléatoire permet d'assurer une isotropie spatiale en moyenne sur un grand nombre de graphes. Ainsi peut-on caractériser la nature de la transition observée grâce à une simulation avec un grand nombre d'essais qui compense le nombre petit, d'éléments du système.

Le travail suivant sur la simulation consiste à *comparer la valeur des exposants critiques en fonction du nombre d'habitants*. L'objectif est d'essayer d'obtenir une loi de dépendance de l'exposant critique en fonction du nombre d'habitants¹. Cette loi de dépendance peut être obtenue pour les faibles valeurs du nombre d'habitants. En effet, pour ces petites valeurs, connaissant déjà partiellement la loi de dépendance, on peut prédire la valeur attendue de l'exposant critique pour une valeur fixée, avant d'effectuer la simulation. On effectue ensuite la simulation et on compare la valeur obtenue par la simulation à celle prédictive théoriquement. *Ce protocole permet de conforter la loi obtenue avec la simulation pour les petites valeurs.* On peut ensuite envisager de prolonger cette loi de dépendance pour les plus grandes valeurs. Construire une loi de dépendance peut être particulièrement intéressant, car cela permet d'obtenir des informations théoriques sur des simulations que l'on ne peut pas effectuer du fait de la puissance limitée des machines. En effet, on ne peut pas proposer de simulations pour des nombres d'habitants trop importants et en particulier pour des nombres d'habitants proches du nombre réel d'habitants en RDA en 1989. Dans ce cas cette loi de dépendance permettrait d'obtenir une valeur estimée de l'exposant critique sans même effectuer la simulation. La loi de dépendance permet alors de prédire les valeurs des exposants critiques alors que l'on ne peut les obtenir à l'aide des simulations.

Un travail supplémentaire doit être envisagé dans le développement même de la simulation. Comme nous l'avons déjà écrit, la simulation a été conçue pour que coopèrent plusieurs districts, mais n'a été effectivement réalisée que pour un seul district. Il faudrait donc simuler un ensemble de districts en interrelations.

3.1.3. “Validation” du modèle

Plus globalement se pose le problème d'une confrontation² plus fine entre le modèle, la simulation et les données obtenues et analysées par Wolf-Dieter Eberwein. On ne peut évidemment pas parler de validation au sens propre. On peut au mieux envisager une confortation du modèle et de son intérêt.

Dans ce cadre la difficulté principale consiste à spécifier et préciser le choix réalisé en collaboration avec Wolf-Dieter Eberwein de la base de comparaison entre les simulations et les données. La difficulté porte en fait sur le choix des bonnes mesures

¹Il est possible que le paramètre pertinent ne soit pas le nombre d'habitants, mais la proportion entre le nombre de voisins et le nombre d'habitants, car c'est elle qui détermine la distance maximale entre deux habitants au sein d'un même district.

²Wolf-Dieter Eberwein et moi-même avons organisé le 23 novembre 1995 une réunion avec comme sujet la “validation” de ce modèle et plus généralement la confrontation des aspects formels proposés aux données accessibles. Cette réunion a eu lieu à la Maison Suger et a été rendue possible grâce au soutien de la Maison des Sciences de l'Homme. Une douzaine de spécialistes de sciences politiques, de la modélisation en sciences et plus généralement de la pluridisciplinarité ont participé à cette journée. Nous avons tenu compte ici des suggestions qui avaient été faites à cette occasion.

qui vont permettre cette comparaison. Pour le modélisateur ceci signifie notamment qu'il faut proposer une mesure adéquate sur la simulation. De nombreuses mesures sont possibles, c'est pourquoi l'étude extensive de la simulation que nous avons réalisée était un préalable nécessaire. Il fallait d'abord connaître parfaitement le fonctionnement de la simulation avant de pouvoir envisager une confrontation fine avec les données sur la RDA. De son côté, Wolf-Dieter Eberwein a également effectué de nombreux tests statistiques très variés dont l'objectif était de mieux comprendre les données récoltées. Le but était de proposer diverses classifications des districts, des catégories sociales et des événements observés, selon le test choisi. Ceci permet également d'envisager divers découpages temporels du phénomène. Ainsi nous avons analysé respectivement et chacun de notre côté la simulation et les données sur le phénomène, afin de proposer des mesures sur la simulation et sur ces données qui soient compatibles. Le but recherché est de confronter de cette manière les deux types de données.

Pour être plus précis il s'agit ici de déterminer le nombre de catégories sociales observées selon les types de districts. On cherche également à obtenir le nombre de types de districts pertinents pour ce phénomène. On doit aussi choisir quels districts vont être pris comme référence. Enfin une échelle de temps doit être déterminée, par exemple la semaine, sur laquelle on s'appuie pour déterminer le nombre et l'ampleur des événements observés.

Les problèmes de validation que nous venons d'évoquer sont essentiellement liés à la sémantique que l'on attribue au modèle et aux liens que l'on peut établir entre la sémantique du modèle et celle du phénomène. On peut prendre un exemple simple avec la notion d'habitant. En ce qui concerne la mobilisation de masse, elle ne pose évidemment aucune difficulté puisque l'habitant est en fait le citoyen allemand lui-même. Dans le cadre du modèle, par contre, ce que nous désignons par habitant est évidemment très différent. Nous ne pouvons pas dire que nous avons "modélisé" un habitant. Nous avons simplement défini un élément formel possédant certaines caractéristiques et susceptible d'avoir certains "comportements" que nous associons à celui que peut avoir un habitant de la RDA au sein d'une population, dans le cadre du phénomène qui nous intéresse. Mais doit-on associer l'habitant formel à un habitant pour le phénomène, en particulier lorsque nous devons les dénombrer, c'est-à-dire effectuer une mesure ? Ne vaut-il pas mieux associer un habitant formel à un groupe d'habitants réels ? Mais quelle doit être alors la taille de ce groupe ? Doit-on associer un habitant formel à dix, cent ou mille habitants réels ? Si on associe un habitant formel à un groupe d'habitants réels, quel doit être alors le sens des liens entre habitants ? Lorsque l'on interprétait les habitants formels comme des habitants réels, on pouvait dire que le lien entre ces habitants formels était une relation de connaissance. On pouvait même proposer ensuite un modèle "recuit", c'est-à-dire pour lequel les liens auraient changé au cours du temps, par exemple à chaque pas de temps. Dans ce cas les liens auraient été les rencontres entre les

habitants effectuées en un pas de temps, c'est-à-dire par exemple en une journée. Mais quel sens doit-on désormais attribuer aux liens si les habitants formels sont en fait des groupes d'habitants réels ? Doit-on considérer que ces liens sont en fait une moyenne des liens entre les habitants ? Faut-il au contraire définir une notion de lien abstraite entre groupes dont le pendant "réel" reste à définir ? On voit bien ici que les difficultés de la validation sont étroitement liées au choix de deux sémantiques, pour le phénomène et pour le modèle, qui soient compatibles. Ces deux sémantiques sont évidemment en partie définies et réinterprétées en même temps, ce qui impose une collaboration étroite et une remise en cause permanente et commune en fonction des différents éléments que l'on peut obtenir dans chaque domaine de la modélisation.

On voit bien ici que les difficultés et les problématiques posées sont d'abord méthodologiques avant d'être techniques. Il faut d'abord être capable d'analyser et de mettre en relation deux "sémantiques" différentes avant de se concentrer uniquement sur l'une d'entre elles en pouvant affirmer scientifiquement qu'elle a bien un lien étroit avec l'autre et qu'elle peut être réinterprétée dans les termes de l'autre. L'enfermement dans un choix technique particulier est toujours possible et pourrait constituer une "échappatoire", mais ce ne pourrait être une solution au problème de la modélisation qui consiste à mettre en relation étroite deux domaines initialement bien séparés. Ainsi les véritables difficultés de la modélisation sont avant tout méthodologiques avant d'être techniques. Modéliser consiste alors non seulement à résoudre des équations en se concentrant sur un domaine mais aussi à effectuer des choix et à réfléchir sur ce qui les motive et les justifie scientifiquement. La réflexion méthodologique est alors partie intégrante de la modélisation. Considérer que l'une ne relève pas de la science mais de la métaphysique c'est assurer à l'autre le même sort.

3.2. Apports aux sciences cognitives

Dans cette section nous souhaitons dégager l'intérêt que ce travail peut avoir pour les sciences cognitives. On pourrait affirmer tout d'abord que les phénomènes de sciences politiques relèvent également des sciences cognitives. Dans cette sous-section nous supposerons qu'il n'en est rien et que les sciences cognitives se limitent à l'étude du cerveau/esprit et de son fonctionnement. Nous avons déjà écrit que les modèles proposés pour l'étude de la mobilisation de masse et ceux utilisés pour étudier les réseaux de neurones sont identiques. Si l'on se contente d'étudier des modèles, il est donc inutile de distinguer sciences politiques et neurosciences cognitives. Les phénomènes n'ont alors aucune importance. Il en est tout autrement dans le contexte de la modélisation. Dans ce cadre on ne se contente pas d'étudier des modèles formels, mais comme nous l'avons vu dans la première section de ce chapitre on étudie également les phénomènes et on les confronte à ces modèles. La

distinction entre sciences politiques et neurosciences cognitives reprend alors son sens. On peut donc de nouveau s'interroger sur ce que la modélisation et la pratique de la modélisation en sciences politiques peuvent nous apporter comme connaissance sur la modélisation et la pratique de la modélisation en neurosciences cognitives et plus généralement en sciences cognitives. Aussi, pour mettre en évidence cet intérêt, nous montrerons cette proximité en nous référant systématiquement aux réseaux de neurones réels ou formels comme exemples de phénomène et de modèle étudiés en sciences cognitives.

3.2.1. Proximité des modèles de réseaux neuronaux

Il faut rappeler tout d'abord, comme nous l'avons déjà fait à plusieurs reprises, que le modèle proposé pour l'étude de la mobilisation de masse en RDA est très proche dans sa structure des modèles de réseaux neuronaux. C'est essentiellement l'interprétation que l'on peut faire de ces modèles selon le phénomène étudié qui diffère. On peut déduire de cette constatation que tous les éléments d'étude obtenus sur le modèle lui-même ou sur les simulations peuvent être réinterprétés dans le contexte des réseaux de neurones. On ne peut bien entendu plus parler de modélisation dans ces conditions, sauf à reprendre l'ensemble des interrelations entre le modèle, la simulation et des données précises effectivement obtenues pour des réseaux de neurones réels.

Le choix d'un graphe aléatoire, la possibilité de développer des réseaux recuits, l'utilisation de l'approximation en champ moyen, l'obtention des exposants critiques, la définition d'un réseau à deux étages avec des dynamiques pour l'un et pour l'autre, sont tous des éléments d'étude du modèle ou de la simulation qui peuvent être directement appliqués aux réseaux de neurones formels.

Nous avons également, dans ce contexte, posé une question qui nous semble essentielle dans l'étude même des réseaux de neurones formels et qui, à notre connaissance, n'a jamais été posée. *Il s'agit des modifications du comportement global du réseau de neurones en fonction du nombre des éléments qui le composent.* Nous avons abordé cette question pour notre modèle dans le cadre des sciences politiques. Puisque nous avons posé cette question dans le cadre du modèle et de la simulation elle se pose dans les mêmes termes pour les réseaux de neurones formels. Lorsque l'on propose l'étude d'un réseau de neurones, on se donne un nombre de neurones, généralement petit, par exemple cent, et on étudie alors le comportement global du réseau³. Certains travaux plus sophistiqués qui proposent des dynamiques sur le graphe introduisent ainsi de nouveaux neurones et modifient donc le nombre de ces neurones dans le réseau. Cette modification du nombre de neurones est

³Puisque les modèles sont très similaires, on comprend que le nombre d'éléments que l'on peut mettre en jeu soit approximativement le même pour les simulations de réseaux de neurones ou notre modèle de la mobilisation de masse.

alors effectuée pour accroître les performances du réseau. *Notre optique est tout à fait différente. Nous ne souhaitons pas améliorer les performances du réseau mais étudier les modifications du comportement global du réseau en fonction du nombre de neurones.* Cette question est évidemment fondamentale. En effet si on ne se pose pas la question de l'évolution du comportement global du système en fonction du nombre de ses éléments, cette étude n'a plus aucun sens. Sans cela, comment peut-on encore oser parler de réseau de neurones à l'échelle de l'homme quand on n'a que quelques centaines de neurones ou de comportement politique à l'échelle d'un État quand on n'a que quelques centaines d'habitants ? Le biologiste qui a réalisé des expériences sur le cerveau, et aussi le politologue, s'en offusqueraient avec raison.

Dans ce contexte de nouvelles questions se posent aux sciences cognitives. Pour pouvoir évoquer la notion de limite thermodynamique, il ne faut plus seulement étudier le comportement du système quand le nombre d'éléments est très grand. Il faut d'abord étudier les modifications du comportement en fonction du nombre d'éléments, puis *si on le souhaite et si cela est possible*, par un "passage à la limite", essayer d'établir ce qu'il en serait asymptotiquement.

3.2.2. Difficultés pour valider les modèles en sciences cognitives

Il faut noter également que les difficultés méthodologiques rencontrées ne sont pas propres à la modélisation particulière réalisée. Elles sont liées en partie au type de modèle proposé. Nous les avons abordées dans le cadre d'un phénomène de sciences politiques, mais elles sont tout à fait identiques en sciences cognitives. *Ainsi le problème de l'établissement d'une interprétation commune au phénomène et au modèle se pose également en sciences cognitives*, si on ne souhaite pas se cantonner dans une analyse sommaire du phénomène ni dans une conception naïve des modèles. On peut se demander par exemple quelle correspondance il faut choisir entre les durées mesurées sur les expériences et les durées programmées dans les simulations. Nous avons déjà évoqué la difficulté consistant à choisir également une correspondance entre le nombre de neurones observés et le nombre de neurones programmés. On peut alors se demander si la taille (le nombre d'éléments) et le temps (la durée de certains événements) du système sont liés l'un à l'autre. Si c'est le cas la correspondance entre le système observé et le modèle ou la simulation doit être définie simultanément pour le nombre d'éléments et pour la durée. Ces questions sont évidemment très importantes si l'on souhaite "valider" des modèles pour pouvoir les interpréter correctement en retour.

Nous avons déjà signalé que les données, quoique partielles, sont d'un accès relativement aisé pour le phénomène de sciences politiques qui nous intéresse. Ceci n'est bien entendu pas en contradiction avec le fait que ces données sont ensuite difficiles à traiter et à interpréter. Les données biologiques sur les réseaux de neurones sont également très nombreuses. Cependant, elles sont généralement obtenues

en effectuant des mesures moyennes sur un ensemble de neurones. Quand les mesures sont réalisées sur des neurones individuels, on ne connaît alors pas les connexions effectives entre ces neurones, c'est-à-dire les circuits neuronaux précis sur lesquels on effectue les mesures. En conséquence on peut penser que les données biologiques qui permettraient de valider les modèles de neurones formels sont sans doute extrêmement difficiles à obtenir. Elles nécessiteraient des mesures très fines du comportement de chaque neurone en connaissant sa position relative dans le réseau, ce qui est rarement possible à l'heure actuelle. Ainsi, dans le cadre des sciences cognitives, de nouvelles difficultés viennent encore s'ajouter à celles que nous avons rencontrées dans la validation du modèle de sciences politiques.

Il nous faut également insister sur une autre difficulté constatée sur notre modèle et que l'on retrouve pour les modèles de réseaux de neurones. Ces modèles, comme le précise le cadre de Farmer, sont constitués d'un graphe. Or on ne peut généralement pas associer de dimension d'espace pour les graphes, sauf localement. Ceci signifie en particulier que pour ces graphes on n'a pas l'isotropie spatiale qui permet souvent d'obtenir des résultats théoriques sur le comportement des systèmes. Comme on a un nombre fini d'éléments, on peut tout de même les plonger dans un espace continu de dimension suffisamment grande. Ceci a été fait cependant sans tenir compte des modifications des connexions du graphe au cours du temps. Ainsi, même si on peut plonger à chaque instant le graphe dans un espace continu, la dimension de cet espace devra *a priori* changer au cours du temps.

En résumé les difficultés méthodologiques rencontrées sont importantes. Elles se retrouvent toutes en sciences cognitives mais sont sans doute bien moins grandes que celles qui sont liées à la "validation" des réseaux de neurones formels comme modèles de systèmes nerveux. En particulier il semble que la "validation" de tels modèles nécessite de porter son attention sur les détails neuroanatomiques et neurofonctionnels qui, à l'heure actuelle, ne sont pas toujours accessibles à la mesure. Cependant, même quand la modélisation n'est pas achevée et quand les difficultés ne sont pas toutes franchies, l'expérience nous a montré que la modélisation est une occasion d'enrichir les différents domaines qui collaborent. *En particulier, la modélisation, même partielle, grâce à un dialogue incessant entre domaines, permet d'approfondir et de préciser les concepts utilisés dans chacun de ces domaines.* Cette collaboration étroite permet, par exemple, que des questions en provenance d'un domaine soient posées au sein d'un autre domaine alors même qu'elles ne s'y étaient pas posées naturellement.

RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 4

- [Bak *et al.*, 1992] BAK P., FLYVBJERG H. ET LAUTRUP B. Coevolution in a rugged fitness landscape. *Physical review A*. 1992, vol. 46, no 10, p. 6724–6730.
- [Blattert *et al.*, 1995] BLATTERT B., RINK D. ET RUCHT D. Von den Oppositionsgruppen der DDR zu den neuen sozialen Bewegungen in Ostdeutschland. *Politische Vierteljahresschrift*. 1995, no 3, p. 397–422.
- [Casti, 1996] CASTI J. *The outer limits: in search of the “unknowable” in science*. Santa Fe Institute, Santa Fe, 1996. Rapport technique 96-01-001.
- [Eberwein *et al.*, 1990] EBERWEIN W.-D., CUSACK T., JOHNSON C., LÜCHAUER A., GÖBEL G., GÖBEL J. ET J. S. *Der Drang nach Westen : zur Analyse der Volkskammerwahl*. 1990. FIB Papers, Forschungsgruppe Internationale Beziehungen, Wissenschaftszentrum Berlin, P 90-301, Juli.
- [Eberwein *et al.*, 1991] EBERWEIN W.-D., JOHNSON C., STANGEL J. ET GÄRTNER K. *Vom Aufstand der Massen zum Ende der DDR*. 1991. Veröffentlichungen der Forschungsgruppe Internationale Beziehungen, Wissenschaftszentrum für Sozialforschung Berlin, P 91-308.
- [Eberwein et Saurel, 1995] EBERWEIN W.-D. ET SAUREL P. *The dynamics of collective behavior: modeling mass mobilization in the GDR*. Panel “Formal theory” at the Second pan-european conference in international relations, Fondation des sciences politiques, Paris, 13-16 septembre 1995.
- [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*. 1994, no 349, p. 29–42.
- [Hilbert, 1971] HILBERT D. *Les fondements de la géométrie*. Première édition Dunod. Dunod, Paris, 1971. 311 p. Traduction de Grundlagen der Geometrie. Première édition allemande Teubner, 1899.
- [Hirschman, 1993] HIRSCHMAN A. Exit, voice, and the fate of the german democratic republic: an essay in conceptual history. *World politics*. 1993, vol. 45, p. 173–202.

- [Köhler, 1990] KÖHLER A. Ist die übersiedlerwelle noch zu stoppen. *Deutschland-Archiv*. 1990, no 3, p. 425–430.
- [Kuran, 1989] KURAN T. Sparks and prairie fires: a theory of unanticipated political revolution. *Public choice*. 1989, vol. 61, p. 41–74.
- [Kuran, 1991a] KURAN T. The east european revolution of 1989: is it surprising that we were surprised? *The american economic review*. 1991, vol. 81, no 2, p. 121–125.
- [Kuran, 1991b] KURAN T. Now out of never: the element of surprise in the east european revolution of 1989. *World politics*. 1991, vol. 44, p. 7–48.
- [Liggett, 1985] LIGGETT T. *Interacting particle systems*. Springer, New York, 1985. 488 p. Volume 276 de la collection Grundlehren der mathematischen Wissenschaften. ISBN 0-387-96069-4 3-540-96069-4.
- [Lohmann, 1994] LOHMAN S. Dynamic of informational cascades: the monday demonstrations in Leipzig, East Germany, 1989-91. *World politics*. 1994, vol. 47, no 1, p. 42–101.
- [Malyshev et al., 1995] MALYSHEV V., MANITA A., PETROVA E. ET SCACCIATELLI E. Hydrodynamics of the weakly perturbed voter model. *Markov processes and related fields*. 1995, vol. 1, no 1, p. 3–56.
- [Opp, 1991] OPP K.-D. DDR'89. Zu den Ursachen einer spontanen Revolution. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. 1991, vol. 43, no 2, p. 302–321.
- [Opp et Gern, 1993] OPP K.-D. ET GERN C. Dissident groups, personal networks, and spontaneous cooperation: the east german revolution of 1989. *American sociological review*. 1993, vol. 58, p. 659–680.
- [Popper, 1978] POPPER K. *La logique de la découverte scientifique*. Reproduction de la première édition française parue en 1973. Payot, Paris, 1978. 484 p. Collection Bibliothèque scientifique. Traduit de The logic of scientific discovery. Première édition anglaise Hutchinson, 1959. Première édition allemande 1934. ISBN 2-228-11391-3.
- [Prez, 1997] PREZ P. *De la vision à l'acte de voir : un parcours technique et épistémologique dans les sciences cognitives*. Thèse de doctorat, université de Paris-sud, Orsay, 1997.
- [Rink, 1991] RINK D. Soziale Bewegungen in der DDR: die Entwicklungen bis Mai 1990. Dans *Neue Soziale Bewegungen in der Bundesrepublik Deutschland*, 2, Roth R. et Rucht D., éditeurs, p. 54–70. Bundeszentrale für politische Bildung, Bonn, 1991.

- [Rosen, 1987] ROSEN R. On the scope of syntactics in mathematics and science: the machine metaphor. Dans *Real brains, artificial minds*, Casti J. et A. K., éditeurs, p. 1–23. North-Holland, New York, 1987. ISBN 0-444-01155-2.
- [Sinaceur, 1994] SINACEUR H. *Jean Cavaillès : philosophie mathématique*. Presses Universitaires de France, Paris, 1994.
- [Ulrich, 1990] ULRICH R. *Die Übersiedlerbewegung in die Bundesrepublik und das Ende der DDR*. Wissenschaftszentrum Berlin für Sozialforschung, 1990. Rapport technique 90302.

CHAPITRE 5

**La complexité associée à la cognition
nécessite la matérialité :
une thèse pour une complexité
au-delà de la complexité/quantité**

We have absolutely no past experience with systems of this degree of complexity.

John VON NEUMANN
The general and logical theory of automata, 1948¹

INTRODUCTION

La complexité, comme nous l'avons vu, s'appuie aujourd'hui sur quatre points : le **grand nombre** d'éléments, les **interactions** entre ces éléments, les **dynamiques** de ces éléments ainsi que la **redondance** entre ces éléments. Comme nous avons essayé de le montrer jusqu'à présent, les concepts des sciences de la complexité s'appuient tous sur des propriétés liées à la quantité. En effet, le grand nombre permet d'approximer les interactions entre les éléments par une "interaction moyenne". Les dynamiques de ces éléments sont en fait les mêmes pour tous les éléments du système. La redondance, quant à elle, prend différentes formes, soit une redondance spatiale, c'est-à-dire une périodicité dans l'espace des voisinages, soit une redondance temporelle qui prend la forme de répétitions lorsque la dynamique de l'automate se déroule. La redondance sous ses différentes formes restreint l'ensemble formel effectivement étudié mais permet de rendre compatibles deux exigences contradictoires, celle de l'infini pour obtenir des propriétés théoriques et celle de la finitude qui permet la mise en œuvre pratique.

Cette situation générale n'est pas propre aux sciences de la complexité. Dans les différentes sciences et en particulier en sciences cognitives les principaux modèles s'appuient sur ces hypothèses du grand nombre et de la redondance. La question

¹Références précises [von Neumann, 1951].

qui se pose alors est de savoir si ces hypothèses qui sont fondatrices des principaux modèles actuels sont suffisantes pour modéliser le cerveau. Un des objectifs des sciences cognitives est de comprendre le fonctionnement du cerveau. Cet objectif peut-il être atteint sans comprendre le cerveau dans sa biologie ? Peut-on ramener les éléments pertinents pour notre propos, et issus de la biologie, à la seule quantité ? Les hypothèses sous-jacentes dans les modèles actuels sont-elles compatibles avec nos connaissances neurobiologiques sur le cerveau ? Nous ne le pensons pas, comme nous essaierons de le montrer brièvement dans ce chapitre. Pour tenir compte de ces éléments, il faut proposer des modèles s'appuyant sur les concepts de la quantité, mais sur d'autres concepts également. Il faut même aller plus loin. Comme nous le montrerons, *c'est la modélisation elle-même qui doit prendre une autre forme pour tenir compte de ces nouveaux éléments.*

Dans une première section, nous considérerons implicitement que la compréhension scientifique de la cognition est étroitement et exclusivement liée à une modélisation convenable du cerveau. Dans ce contexte, nous mettrons en évidence différents aspects de la complexité cérébrale généralement délaissés dans les modèles issus de la complexité/quantité : **la diversité, la mixité de structure logique, l'hétéarchie.** Ces caractéristiques sont essentielles pour décrire les comportements biologiques. Ils sont certainement des éléments nécessaires d'une modélisation appropriée du cerveau pour rendre compte de la cognition.

Dans une seconde section nous montrerons quelles modifications conceptuelles cela entraîne. Ces éléments nous amènent à formuler une nouvelle thèse : **la complexité associée à la cognition nécessite la matérialité des éléments en jeu.** Cette thèse a de multiples conséquences. On peut la considérer comme une thèse alternative à celle symboliste de Herbert Simon. Si on la prend au sérieux, elle impose par exemple, dans le cadre du connexionisme, de ne peut plus concevoir les processus émergents comme construits sur des objets. Le processus² devient premier. *Les processus doivent être élaborés à partir d'autres processus.* Les processus ne doivent pas être conçus statiquement comme le sont les symboles ou les jeux de symboles. Cette thèse nous amène également à dépasser la conception admise dans la première section, selon laquelle le cerveau détenait l'exclusivité des facultés cognitives. *Nous sommes amenés à concevoir la cognition dans un cadre qui ne se restreint pas aux cellules nerveuses.* La modélisation elle-même ne peut alors plus être pratiquée de la même manière. Un nouveau terme entre en jeu dans l'étude d'un phénomène, la *reconstruction matérielle du phénomène étudié avec le même matériau.* On ne peut alors plus parler de modélisation, mais plutôt de **mise en parangon.** Le terme de parangon est ici à mettre en balance avec celui de modèle. Modèle et parangon ont tous deux valeur d'exemple et de domaine qui synthétise une forme de connais-

²Nous attribuons ici au terme de processus le sens que nous lui avons donné dans le chapitre 2. Un processus est donc l'abstraction d'un phénomène dynamique. Il ne faudrait donc pas confondre le terme avec la notion technique de système dynamique.

sance. Comme depuis le début de cette thèse, le terme de modèle se rapporte à la notion de système formel. Au contraire, nous réservons le terme de parangon pour renvoyer à une construction matérielle à valeur exemplaire. *Le parangon est alors au centre de la mise en parangon, comme le modèle est au cœur de la modélisation.* Méthodologiquement, la différence entre modélisation et mise en parangon devient fondamentale.

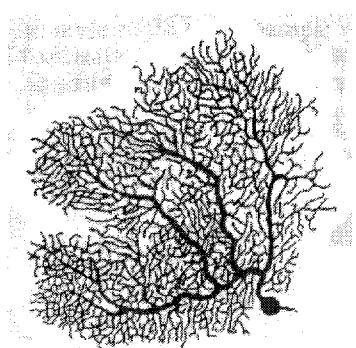
1. LA COMPLEXITÉ AU-DELÀ DE LA QUANTITÉ

Dans cette section nous montrerons que les descriptions biologiques du cerveau nous amènent à introduire des notions nouvelles et inconnues des modèles de la complexité/quantité. Ces notions sont **la diversité, la mixité des structures logiques et l'hétéarchie**.

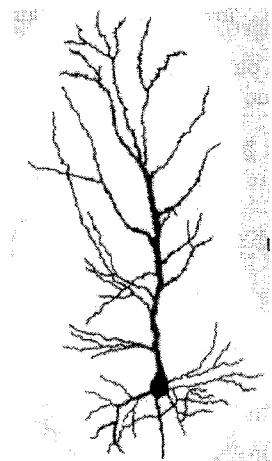
1.1. La diversité des éléments cérébraux

“Le cerveau comprend une variété de cellules beaucoup plus grande que celle de tout autre organe”[Rosenzweig et Leiman, 1991][p. 42]. Choisir d’insister sur la diversité des éléments cérébraux, comme nous le faisons, pour parler de la cognition, consiste déjà à accepter une hypothèse généralement admise, à savoir que la pensée, mais aussi la vision et les autres fonctions “cognitives”, se déroulent essentiellement “dans la tête”¹. Certains arguments biologiques percutants peuvent être invoqués dans ce sens, comme la présence de la barrière hématocérale. *Il faut se garder cependant d'une conception trop radicale et d'une coupure trop nette liée à cette “barrière”.* Comme l’expliquent Pierre Buser et Michel Imbert, “on sait précisément, depuis longtemps déjà, que certaines zones du cerveau n’ont que peu ou pas de véritable barrière hématoencéphalique” [Buser et Imbert, 1993][p. 66]. On distingue désormais deux types de barrage, “celui entre le sang et le tissu cérébral (barrière hématocérale) et celui des plexus choroïdes, interposé entre le sang et le liquide céphalorachidien (barrière hématoLCR) ; on discute encore de savoir si ces deux barrières présentent une perméabilité identique” [Buser et Imbert, 1993][p. 66-67]. Il est intéressant de noter également comment se déroule le franchissement éventuel de cette barrière : “ou bien la substance présente une propriété de liposolubilité, ou bien elle est reconnue par un des systèmes de transport actif présents au niveau des cellules du barrage” [Buser et Imbert, 1993][p. 67]. Ce franchissement correspond à deux mécanismes très différents dont on ne pourrait sans doute pas rendre compte avec des modèles présentant les mêmes hypothèses. Il s’agit là d’un problème de

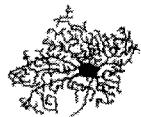
¹Nous reprenons ici l’expression de McCulloch lors de sa conférence de 1948 au Hixon Symposium dont le titre était “Why the mind is in the head” [McCulloch, 1951]. En fait plutôt que de la tête, on parle ici du cerveau, voire du néocortex.



Cellule de Purkinje



Cellule pyramidale



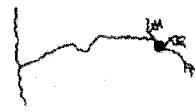
Cellule du noyau de l'olive



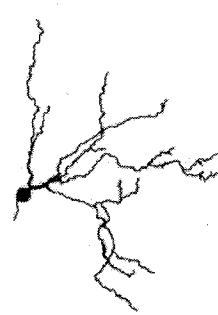
Petite cellule de la substance gélatineuse



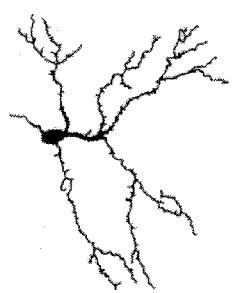
Noyau en fuseau de la substance gélatineuse



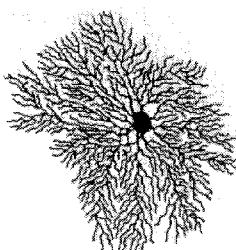
Cellule granulaire



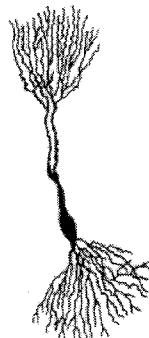
Cellule ovoïde



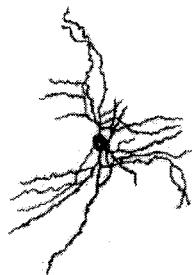
Grande cellule du noyau trigéminé



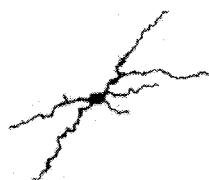
Cellule du noyau thalamique



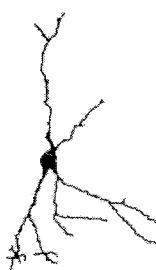
Cellule en double pyramide



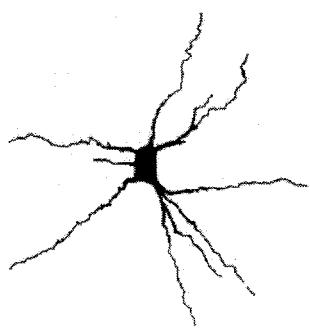
Neurone du putamen à noyau lenticulaire



Petite cellule de la formation réticulaire



Cellule du globus pallidus



Grande cellule de la formation réticulaire

FIGURE 1 : Bestiaire montrant quelques aspects de la diversité des cellules nerveuses et en particulier des formes des arborisations dendritiques (d'après Pour la Science N°181, 1992, p33)

Chapitre 5 -- Figure 1

modélisation de mécanismes qui fonctionnent selon des logiques différentes. Nous reviendrons plus loin sur ce type de difficulté.

1.1.1. La diversité des cellules nerveuses

On trouve plusieurs classes de cellules dans le cerveau, les cellules nerveuses ou neurones et des cellules non nerveuses comme les cellules gliales. Dans les modèles actuels qui doivent nous permettre d'améliorer notre connaissance des fonctions cérébrales, on attribue aux cellules nerveuses un rôle exclusif. Les modélisateurs ont considéré dans une première approximation qu'on pouvait se contenter d'étudier des modèles de réseaux comportant uniquement des neurones. C'est dans cette optique que les modèles de "verre de spin", par exemple, sont censés rendre compte de l'apprentissage. "Le paysage d'énergie d'un verre de spin, avec ses multiples vallées, fournit alors une merveilleuse métaphore pour un réseau de neurones fonctionnant comme mémoire" [Mézard et Toulouse, 1991][p. 621]. Il convient cependant d'aller plus loin que la métaphore, si merveilleuse soit-elle, et de se demander si ces hypothèses permettent également une véritable modélisation susceptible de nous apporter des explications, voire une compréhension de certains aspects du fonctionnement cérébral.

Le ou les neurones et leur classification

Nous voulons faire part ici de notre étonnement devant la diversité des neurones réels qui se trouvent effectivement dans le cerveau humain. Selon Jean-Didier Vincent "ils sont au nombre d'environ 10^{12} dans le cerveau humain, représentés par un millier de types différents"² [Vincent, 1988][p. 69]. On s'est souvent étonné et émerveillé devant la quantité, le "nombre astronomique" de ces neurones. C'est d'ailleurs ce grand nombre qui justifie l'utilisation des verres de spin comme modèle des réseaux neuronaux. Jean-Didier Vincent, cependant, insiste également sur le grand nombre de variétés de neurones différents.

Différentes classifications de ces neurones ont été proposées. Pierre Buser et Michel Imbert, par exemple, choisissent de présenter une classification générale et traditionnelle qui comporte trois grandes classes [Buser et Imbert, 1993][p. 34]. La première classe comprend des neurones à axone long dits du "type I de Golgi". La seconde est constituée des neurones bipolaires longs qui sont donc un cas particulier des neurones à axone long. La troisième classe enfin comprend les neurones à axone court, dits du "type II de Golgi".

Cette première classification ne permet cependant pas de se faire une idée de la diversité extraordinaire des cellules nerveuses (figure 1). *Selon Sereno, le nombre de types différents de neurones du néocortex d'un même individu se situerait entre*

² C'est nous qui soulignons.

50 et 500 [Sereno, 1988]. La diversité des neurones provient de la variété de leurs tailles, de leurs formes et de leurs activités chimiques.

Sejnowski présente une description relativement détaillée qui permet de mettre en évidence les difficultés de la classification des neurones. "Il y a des types très différents de neurones, et différentes parties du cerveau ont des neurones avec des propriétés spécialisées. Il y a cinq types généraux de neurones dans la rétine³, par exemple, qui ont chacun une morphologie, une connectivité, des propriétés physiologiques et une origine embryologique bien distinctes. Ces dernières années, de plus, des différences physiologiques et chimiques ont été découvertes à l'intérieur de ces classes. Par exemple, 23 types différents de cellules ganglionnaires (dont les axones se projettent jusqu'au cerveau à travers le nerf optique) ont été identifiés ainsi que 22 types différents de cellules amacrines (qui permettent des interactions latérales et une différenciation temporelle). Il y a sept types généraux de neurones dans le cervelet et à peu près douze types généraux dans le néocortex, avec de nombreux sous-types distinguables par leurs propriétés chimiques, comme les neurotransmetteurs qu'ils contiennent. La définition d'un type neuronal est quelque peu arbitraire, car les appréciations sont généralement réalisées sur la base de subtiles différences morphologiques qui peuvent être graduelles plutôt que catégoriques. Plus on découvre de marqueurs chimiques, cependant, *plus il devient clair que la diversité des neurones à l'intérieur du cortex cérébral a été largement sous-estimée*⁴"[Churchland et Sejnowski, 1992][p. 42-43]. Comme le met en évidence cette citation, la classification des neurones est difficile à réaliser.

Les critères de différenciation sont de trois types : des critères morphologiques, physiologiques et chimiques. Deux aspects de la classification expliquent les difficultés rencontrées. Le premier aspect est lié au fait que *certaines critères de discrimination choisis sont qualitatifs et graduels et peuvent être d'application délicate*. Les critères morphologiques sont généralement de ce type. Il devient alors difficile d'attribuer à chaque neurone observé la classe qui lui est associée. Le deuxième aspect est lié à la combinaison des différents critères qui sont en jeu. *Le nombre de classes observées augmente beaucoup plus vite que le nombre de critères ajoutés*. En effet, chaque critère additionnel va se combiner avec l'ensemble des valeurs possibles des critères précédents. La multiplication du nombre de critères nouveaux, du fait de l'observation de nouveaux transmetteurs chimiques, par exemple, va entraîner une augmentation très importante du nombre total de classes de neurones. On peut alors parler de *difficulté combinatoire* de la classification.

Quelles contraintes nouvelles et minimales ceci impose-t-il sur nos modèles ? Si l'on souhaite tenir compte de cette classification des neurones biologiques, il faut qu'elle se retrouve dans la description des neurones formels. Il faut alors que les

³Ces cinq types sont les bâtonnets et les cônes, les cellules horizontales, les cellules bipolaires, les cellules amacrines, et les cellules ganglionnaires [Hubel, 1988][p. 52].

⁴*C'est nous qui soulignons.*

modèles de neurones formels étudiés comportent des paramètres que l'on puisse associer à chacun des critères de la classification. Il faut donc des paramètres, pour les différents critères morphologiques, physiologiques et chimiques. Ceci entraîne évidemment une importante complication des modèles à étudier par rapport aux modèles actuels. Il est à noter en particulier que les modèles actuels de réseaux de neurones qui tiennent compte de la forme du neurone individuel sont rarissimes. On trouve au mieux des descriptions du neurone individuel qui ramènent la forme du neurone individuel à une forme standard "équivalente". Les données biologiques incitent cependant à tenir compte de la forme précise du neurone individuel pour préciser la fonction du neurone. "Cette vision de l'évolution laisse supposer que la forme de la cellule nerveuse est une caractéristique déterminante⁵ permettant de définir les attributs fonctionnels d'une cellule nerveuse donnée" [Rosenzweig et Leiman, 1991] [p. 52].

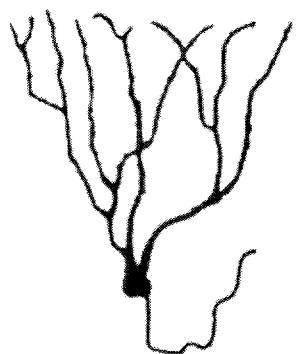
On pourrait même être plus exigeant et souhaiter que certaines difficultés associées à la classification des neurones observés se retrouvent lorsque l'on cherche à classifier rétroactivement les neurones formels. La difficulté que nous avions appelée combinatoire se retrouve déjà si, comme nous l'avons proposé dans le paragraphe précédent, on tient compte d'un nombre beaucoup plus important de paramètres. La difficulté liée au fait que certains critères ne peuvent être évalués que qualitativement et non quantitativement est, quant à elle, beaucoup plus difficile à retranscrire dans un modèle formel. Les critères morphologiques, par exemple, sont difficiles à intégrer dans un unique paramètre. En effet, les descriptions associées, par exemple "être rond", "de forme elliptique", "allongé", etc., peuvent difficilement être ordonnées sur une même échelle. Sans cet ordre, on ne peut pas espérer associer une unique valeur numérique à ce critère. On peut essayer d'associer un symbole à chacune des situations rencontrées. Mais il est alors difficile de retranscrire dans des relations entre les symboles les relations que nous faisons naturellement entre les différentes situations qui ont été symbolisées. En un mot, nous inférons des relations "topologiques" entre les situations rencontrées, et nous ne savons pas comment les retranscrire en des relations "topologiques" opératoires entre les symboles.

Ce que la comparaison entre les animaux nous apprend sur le fonctionnement des cellules nerveuses

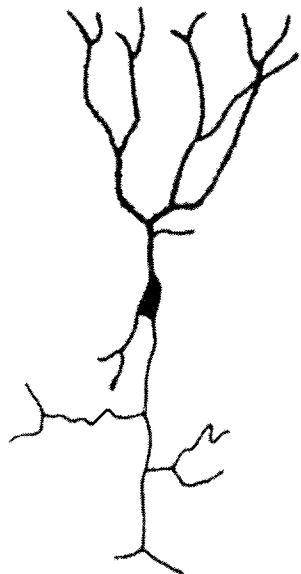
La comparaison entre les animaux⁶ apporte de nombreux éléments permettant d'appréhender le fonctionnement des cellules nerveuses. L'élément mis en avant dans cette comparaison est bien souvent le nombre et la répartition des dendrites. "L'arborisation dendritique est l'un des traits évolutifs les plus frappants des architectures

⁵C'est nous qui soulignons.

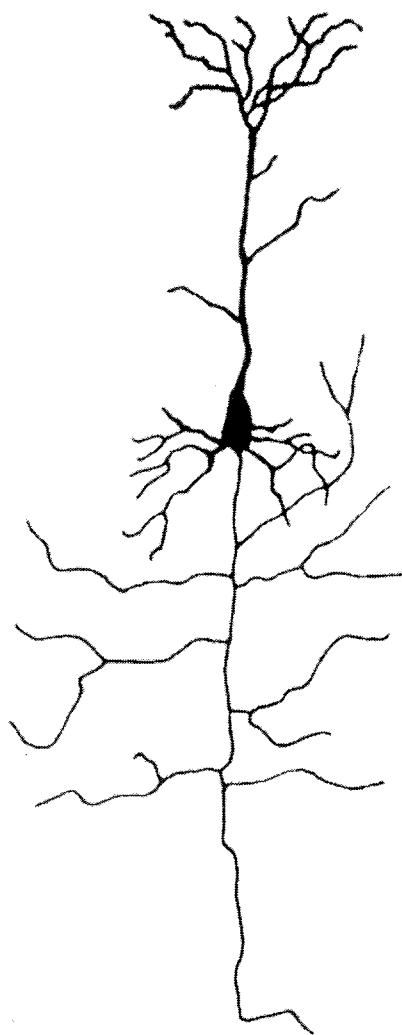
⁶Cette comparaison est souvent elle-même interprétée comme une reconstitution de l'"évolution", et on parle parfois, dans ce contexte, d'**évolution comparée**. C'est dans ces termes qu'il faut comprendre la comparaison entre animaux réalisée et commentée dans [O.F.T.A., 1991].



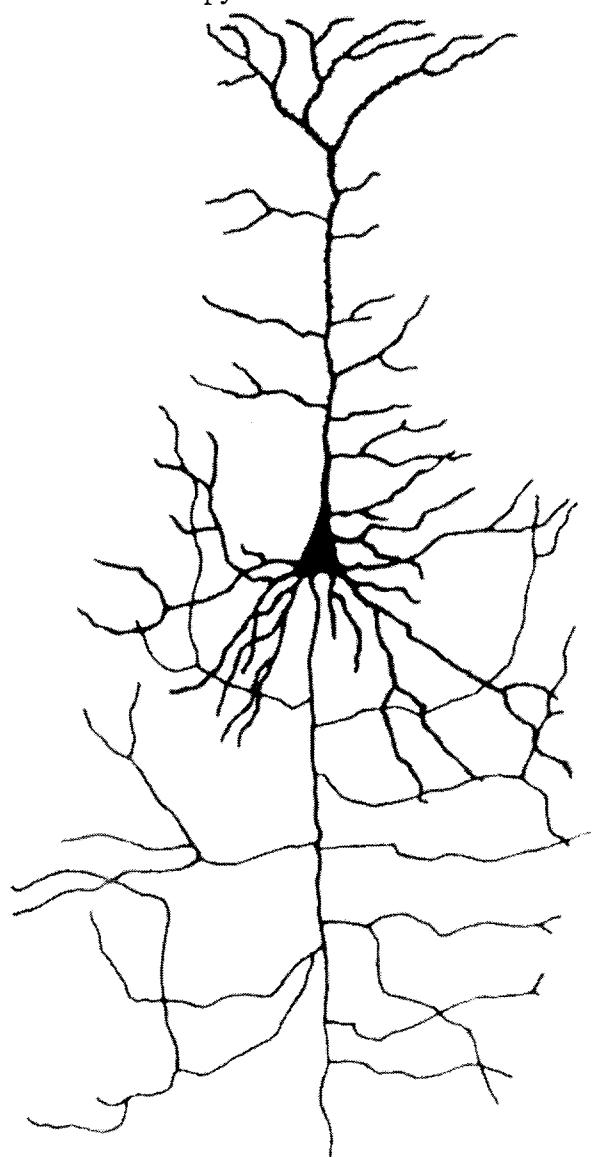
La cellule pyramidale chez la grenouille



La cellule pyramidale chez le lézard



La cellule pyramidale chez le rat



La cellule pyramidale chez l'homme

FIGURE 2 : Une illustration des modifications des cellules nerveuses et en particulier des arborisations dendritiques au cours de l'évolution (d'après [AVA91][p36]). La cellule pyramidale chez la grenouille, le lézard, le rat et l'homme.

Chapitre 5 -- Figure 2

de cerveaux (figure 2). Alors que la taille des corps cellulaires varie peu au cours de l'évolution, c'est-à-dire lorsque l'on s'élève dans l'échelle des vertébrés (phylogénèse), celle des arborisations dendritiques augmente et leur forme se complique. Par exemple on constate que le volume du cerveau humain est beaucoup plus important que celui de la grenouille, qu'il contient un beaucoup plus grand nombre de neurones et donc d'interconnexions. Mais une observation plus fine montre aussi la variété des formes des arborescences dendritiques, et leur complexité structurale croît de manière fantastique (figure 1). *La différence entre le cerveau de la grenouille et celui de l'homme ne réside pas seulement dans le nombre d'éléments simples, mais surtout dans l'extraordinaire diversité de la forme des dendrites*"[O.F.T.A., 1991][p. 35]. Ces éléments laissent penser que la quantité, c'est-à-dire le nombre des cellules nerveuses en jeu, n'est pas le facteur unique d'une explication du fonctionnement cérébral. "Il est désormais possible d'élaborer un modèle de la circulation des courants dans une arborisation dendritique, en tenant compte de sa géométrie tridimensionnelle. On constate que, même dans les conditions les plus simples, cette circulation est beaucoup plus complexe qu'on ne l'imaginait. On découvre que chaque dendrite du neurone véhicule les courants synaptiques selon sa propre géométrie"[O.F.T.A., 1991] [p. 39].

En fait, c'est la conception même selon laquelle on peut considérer les neurones comme des unités élémentaires *simples* pour expliquer le fonctionnement cérébral qui doit sans doute être remise en cause. "Les recherches sur les propriétés des neurones montrent également que ce sont des mécanismes processuels beaucoup plus complexes qu'on ne l'avait imaginé auparavant. Par exemple *les dendrites des neurones sont eux-mêmes extrêmement spécialisés et certaines parties peuvent sans doute agir comme des unités de fonctionnement indépendantes*"⁷[Churchland et Sejnowski, 1992] [p. 44]. Il est sans doute nécessaire de considérer le neurone comme l'unité d'explication élémentaire. Il est par contre certainement abusif, quand on veut rendre compte du fonctionnement cérébral, de modéliser cette unité élémentaire comme si elle était simple. La modélisation nécessite des simplifications. Cependant, malgré le grand nombre de neurones présents dans le cerveau, il semble qu'une simplification extrême du neurone élémentaire ne permet pas de rendre compte du fonctionnement global du système nerveux. **La tentation du modélisateur est de croire que le grand nombre d'éléments permet de compenser la simplification nécessaire de chaque élément.** Il semble bien qu'en ce qui concerne le système nerveux cette approche ait montré actuellement certaines de ses limites. La simplification est certes nécessaire, il ne faut cependant pas qu'elle soit abusive.

Quelles contraintes nouvelles et minimales ceci impose-t-il sur nos modèles ? Si l'on souhaite tenir compte des différences observées entre les neurones biologiques des différents animaux, il faut que cette distinction se retrouve dans la description des neurones formels. Ceci signifie en particulier que l'on ne peut pas utiliser le même

⁷ *C'est nous qui soulignons.*

neurone formel pour chaque animal. Si l'on souhaite utiliser un unique modèle de neurone formel, dont on modifera certains paramètres pour chaque animal, il faut en particulier préciser l'impact précis de l'arborisation dendritique dans le modèle choisi et la manière dont la modification de cette arborisation entraîne la modification de certains paramètres du modèle.

Peut-être faut-il même que dans le modèle du neurone la géométrie du neurone soit différente en fonction de l'espèce animale, si, comme le suggère le texte cité précédemment, “chaque dendrite du neurone véhicule les courants synaptiques selon sa propre géométrie”[O.F.T.A., 1991][p. 39].

Il faudrait que le modèle du neurone formel soit encore plus sophistiqué, si l'on souhaite retrouver certaines propriétés fonctionnelles qui seraient propres aux dendrites. En effet si, comme l'affirment Churchland et Sejnowski, “certaines parties [des dendrites des neurones] peuvent sans doute agir comme des unités de fonctionnement indépendantes”[Churchland et Sejnowski, 1992][p. 44], alors ceci doit influencer de manière importante le mode même de fonctionnement du neurone. Si on conçoit formellement le neurone biologique comme une boîte noire, et si l'on pense que c'est le mode de fonctionnement du neurone qui est explicatif, au sein d'un réseau, du mode de fonctionnement d'un ensemble de neurones formels, alors il faut retranscrire cet aspect processuel du neurone biologique dans la dynamique de nos modèles du neurone formel.

Ce que le développement individuel nous apprend sur le fonctionnement des cellules nerveuses

Les “modèles” de réseaux de neurones sont essentiellement conçus actuellement pour résoudre des problèmes de catégorisation, de classification, etc. En toute généralité les modèles de réseaux permettent de résoudre les problèmes qui mettent en jeu les propriétés de plasticité du système nerveux. Pourquoi ces modèles ne sont-ils alors pas étudiés et utilisés de façon particulière, précise et spécifique, pour étudier de manière constructive le développement des systèmes nerveux ? En effet, “la plasticité la plus spectaculaire pour le système nerveux se déroule pendant la phase développementale. L'apprentissage qui nous fascine tant et nous mystifie lorsque nous contemplons l'acquisition de connaissance générale et épisodique, les capacités motrice et cognitive, le langage et la lecture, palit en comparaison avec l'éénigme stupéfiante qu'est le développement. Des centaines de types différents de cellules doivent être produits, avec la morphologie et la physiologie caractéristique de leur classe”[Churchland et Sejnowski, 1992][p. 307].

Des auteurs comme Quartz et Sejnowski [Quartz et Sejnowski, 1996] considèrent désormais que les modèles formels des réseaux de neurones doivent tenir compte de ces propriétés développementales, de manière constructive, ce qui n'est pas le cas à l'heure actuelle. Tenir compte dans un modèle de la croissance dendritique, et donc

avoir une attitude constructiviste, est *a priori* quelque chose de très différent d'une attitude de type sélectionniste qui consiste à éliminer un ensemble de neurones. Ce type d'approche ne peut pas être réduit immédiatement, purement et simplement à une approche en terme d'apprentissage du réseau. Les éléments en jeu sont différents et nécessitent une modélisation claire et à part entière.

On peut donner des exemples qui sont associés à la différenciation cellulaire. L'observation du développement des cellules de Purkinje dans le cervelet, par exemple, permet de mettre en évidence la différenciation de la forme de certaines cellules nerveuses au cours du développement individuel. "Toute région donnée du système nerveux parvenu à maturité contient un ensemble de cellules nerveuses formé de deux ou plusieurs types de cellules. Par exemple, le cortex cérebelleur contient des cellules de Purkinje et de la microglie. Toutefois, les cellules en migration vers une région sont des neuroblastes qui, au début, sont tous semblables. Ainsi, un neuroblaste donné a donc la capacité de se transformer en l'un des différents types de cellules nerveuses" [Rosenzweig et Leiman, 1991][p. 106]. Les modèles connexionnistes actuels ne permettent pas, en l'état, de tenir compte de ces données. En effet ces modèles négligent complètement la microglie. Ils ne peuvent donc pas expliquer comment une même cellule (le neuroblaste) peut soit donner une cellule nerveuse dont on tient compte dans le modèle, soit une cellule qui est délaissée par le modélisateur. Les modèles connexionnistes de type "sélectionniste" ne permettent pas, nous semble-t-il, de présenter cet aspect de la différenciation cellulaire.

Un autre aspect de la différenciation est très difficile à expliquer dans le cadre des réseaux connexionnistes actuels. Il s'agit du mécanisme de différenciation de la forme des cellules nerveuses. "L'adoption d'une forme caractéristique par un neurone dépend en partie de facteurs déterminants au sein de la cellule individuelle et en partie d'influences provenant des cellules environnantes. Certains éléments d'une cellule donnée croissent de façon typique quel que soit leur environnement. D'autres constituants semblent réagir aux caractéristiques de l'environnement dans le cerveau, telle la présence d'autres cellules" [Rosenzweig et Leiman, 1991][p. 106]. La forme des cellules nerveuses n'étant pas véritablement prise en compte dans les modèles connexionnistes, sa dépendance en fonction des cellules nerveuses qui l'environnent l'est bien moins encore.

Quelles contraintes nouvelles et minimales ceci impose-t-il sur nos modèles ? Si l'on souhaite tenir compte du développement individuel, il faut que cet aspect de la plasticité neuronale se retrouve dans la description des neurones formels. Ceci est possible, par exemple, si on a un modèle qui tient compte explicitement des dendrites et de la géométrie des arborisations dendritiques. Il "suffit" alors de définir, au sein du modèle, des dynamiques de modification et de développement de l'arborisation dendritique de chaque neurone formel en relation avec les autres neurones formels. Ces dynamiques devront éventuellement ne pas être identiques à tout instant et dépendre par exemple de l'arborisation déjà existante. On connaît en physique des

modèles dont les dynamiques sont dépendantes des constructions déjà effectives, ce sont les modèles de diffusion limitée par agrégation (DLA) [Saurel, 1992][p. 29].

On doit être plus exigeant encore si l'on pense que la répartition générale et la différenciation des cellules jouent un rôle sur les relations actuelles entre neurones et modifient la géométrie, l'organisation et la structure globale du cerveau. Dans ce cas il faut aussi que le modèle comporte des éléments susceptibles d'expliquer les relations spatiales, géométriques et morphologiques entre neurones ainsi que l'historique de ces relations.

“Quatre processus cellulaires finement contrôlés sont responsables des changements anatomiques importants qui se produisent dans le système nerveux au cours de la vie embryonnaire et fœtale” [Rosenzweig et Leiman, 1991][p. 102]. Ces processus sont *la prolifération cellulaire*, *la migration cellulaire*, *la différenciation cellulaire* et *la mort cellulaire*. Pour être satisfaisants les modèles de réseaux neuronaux proposés devraient être issus d'un unique cadre explicatif. Ce cadre devrait permettre de modéliser ces différents phénomènes avec éventuellement une modélisation particulière pour chacun de ces processus, en insistant alors sur certains aspects et en laissant les autres de côté. L'essentiel ici est de proposer un tel cadre explicatif unique permettant ces modélisations. La contrainte consiste en particulier à définir ce cadre en déterminant ce qui est commun aux différentes cellules nerveuses, c'est-à-dire :

- i) ce qu'ils partageaient avant de se différencier. Mais cet aspect ne correspond pas et ne doit pas être confondu avec les modèles actuels du neurone. En effet, ces modèles sont eux-mêmes déterminés à partir des “éléments communs” aux cellules nerveuses *après la différenciation cellulaire*.
- ii) ce qu'ils développent de la même manière après leur différenciation. Cet élément semble assez difficile à déterminer.

1.1.2. La diversité des cellules du cerveau et du néocortex

D'autres cellules sont présentes dans le cerveau et en constituent une part très importante. On peut distinguer les cellules gliales et les cellules épendymaires. Leur rôle n'est pas toujours bien connu. Nous ne parlerons ici que des cellules gliales. Les modélisateurs ont été poussés jusqu'à présent à négliger ces cellules lorsqu'ils voulaient rendre compte du fonctionnement du cerveau. On comprend pourquoi les modélisateurs ont laissé de côté les cellules gliales quand on lit : “des traités consacrés au cerveau ignorent souvent les cellules gliales” [Vincent, 1988][p. 71].

On résume parfois le comportement des cellules du cerveau en disant que *les neurones ont une activité électrique et transmettent l'influx nerveux, alors que les cellules gliales ont des fonctions trophiques*. Les deux types de cellules n'ont pas

des comportements de même nature, l'un étant électrique et l'autre de simple support. Les cellules gliales ont alors comme rôle principal “[le] support et [le] maintien du tissu nerveux et [le] remplissage, par prolifération, des régions où ont été lésés les neurones” [Buser et Imbert, 1993][p. 64]. Classiquement, ces deux actions sont décrites comme bien séparées et le fonctionnement de l'une n'a pas d'influence sur le mode de fonctionnement de l'autre. Cette simplification qui interviendra alors dans la modélisation est-elle acceptable et scientifiquement féconde ?

Les cellules gliales sont regroupées en deux ensembles, les **macroglies** et les **microglies**. Les macroglies comportent les **astrocytes** et les **oligodendrocytes**. Les cellules gliales sont capables de se diviser pendant toute la vie du sujet, contrairement aux cellules nerveuses qui, une fois leur état différencié acquis, sont incapables de nouvelle division⁸. De plus les cellules gliales ne comportent qu'un seul type de prolongement. Les relations entre cellules gliales n'ont donc pas lieu par l'intermédiaire de synapses et de neurotransmetteurs, mais par différents mécanismes liés à la proximité spatiale des deux cellules (*zones d'adhérence* et “*gap junctions*”).

On associe désormais aux astrocytes d'autres rôles que les simples fonctions trophiques. Elles servent “de guide matériel et de support lors de la croissance ou de la migration neuronale⁹ (par exemple, la migration des cellules granulaires du cortex cérébelleux de la couche moléculaire vers la couche profonde au cours de l'ontogénèse)” [Buser et Imbert, 1993][p. 64]. Elles servent aussi et “*surtout, de régulateur du fonctionnement synaptique*¹⁰”. À cet égard, il est essentiel de souligner les contacts étroits fréquemment établis aux niveaux des articulations synaptiques, par des prolongements astrocytaires, sur les éléments, pré- et surtout post-synaptiques” [Buser et Imbert, 1993][p. 64].

L'hypothèse essentielle qui justifie le fait de ne pas tenir compte des cellules gliales dans les modélisations du cerveau consiste à dire que le fonctionnement de l'activité électrique et celui de soutien de cette activité sont bien séparés et que le fonctionnement de l'une n'a pas d'influence sur le mode de fonctionnement de l'autre. Cette hypothèse peut-elle encore être soutenue compte tenu du rôle

⁸Ceci n'est pas en contradiction avec l'observation de phénomènes de récupération des fonctionnalités sensorielles. Un tel phénomène a été observé pour les cellules nerveuses à vocation sensorielle après altération de la muqueuse olfactive [Holley et McLeod, 1977].

“L'opinion la plus largement partagée veut que tous les plus gros neurones que le cerveau adulte possèdera soient déjà présents à la naissance. Toutefois, il existe quelques régions autour des ventricules cérébraux, zones dites sous-ventriculaires, où la division mitotique des précurseurs de cellules nerveuses est encore évidente après la naissance. Plusieurs régions du cerveau des rats, y compris le bulbe olfactif et l'hippocampe, semblent acquérir de petits neurones additionnels provenant de ces zones. De fait, Graziadei et Monti-Graziadei (1978) ont prétendu que le remplacement des cellules nerveuses de l'organe olfactif terminal se poursuivait durant toute la vie” [Rosenzweig et Leiman, 1991][p. 111].

⁹Il s'agit de nouveau de la migration cellulaire.

¹⁰C'est nous qui soulignons.

désormais attribué à ces éléments ? Ces cellules sont en fait en contact étroit avec les synapses qui sont les zones principales de transmission de l'information électrique d'une cellule nerveuse à une autre par l'intermédiaire des neurotransmetteurs. Si les cellules gliales interviennent à cet endroit, comme le fait comprendre le texte de Buser et Imbert, c'est toute la communication entre neurones, c'est-à-dire tout le mode de transmission des informations, qui est concerné par ces cellules. Elles participent donc à la régulation de la transmission des informations¹¹. On voit mal alors comment des modèles pourraient être fondés biologiquement sans en rendre compte. On peut citer deux exemples de phénomènes bien attestés qui peuvent être interprétés dans le sens d'une dépendance du fonctionnement des cellules nerveuses en fonction de celui des cellules non nerveuses. Le premier exemple est lié à l'observation des lésions cérébrales. “[...] Certaines catégories de cellules gliales, notamment les astrocytes, réagissent aux traumatismes cérébraux en changeant de dimensions par gonflement : c'est le processus d'**œdème** qui perturbe les fonctions des neurones¹² et est responsable de plusieurs symptômes associés aux lésions cérébrales” [Rosenzweig et Leiman, 1991][p. 50]. Ainsi, sauf à délaisser systématiquement l'explication de ces lésions dans nos modèles du fonctionnement cérébral, on est amené à affirmer que ces modèles doivent non seulement tenir compte de l'activité des cellules nerveuses mais aussi des cellules non nerveuses qui sont susceptibles de modifier cette activité, même si elles n'y participent pas directement. Les modèles proposés à l'heure actuelle ne vont pas dans ce sens. Le deuxième exemple est à associer à la myélinisation. “La formation d'une gaine autour des axones (processus nommé **myélinisation**) modifie considérablement la vitesse de transmission des messages dans les axones. Ce changement devrait avoir un impact considérable sur le comportement puisqu'il affecte profondément le déroulement temporel des événements dans le système nerveux¹³” [Rosenzweig et Leiman, 1991][p. 109]¹⁴.

Depuis peu, le rôle des cellules gliales est très étudié et il semble que ces cellules ont en fait, également, une activité électrique potentielle, comme les neurones. Kettenman et Ransom ont observé que “les astrocytes des mammifères en culture sont étroitement reliés, couplés électriquement les uns aux autres” [Kettenman et Ransom, 1988]. De plus “des études qualitatives ont montré que les astrocytes de culture forment un *syncytium* étroitement couplé électriquement”

¹¹Ces éléments peuvent ne pas participer à l'activité électrique et pourtant modifier le fonctionnement de cette activité électrique.

¹²C'est nous qui soulignons.

¹³C'est nous qui soulignons.

¹⁴Le texte se poursuit ainsi : “Il existe malheureusement peu d'études qui utilisent en même temps des techniques biologiques modernes et des techniques behavioristes pour vérifier ainsi les relations entre les attributs biologiques du système nerveux et les changements dans le comportement. Par conséquent, la démonstration des rapports entre changements de comportement et myélinisation reste toujours à faire” [Rosenzweig et Leiman, 1991][p. 109].

[Kettenman et Ransom, 1988]¹⁵. Contrairement aux neurones dont l'activité électrique est à courte portée, on attribue aux cellules gliales une propriété de communication à longue distance entre astrocytes couplés. Cette propriété n'a à l'heure actuelle jamais été démontrée directement *in vivo*, pour des raisons liées à des difficultés techniques, selon Kettenman et Ransom.

Si ces éléments se confirment, les cellules gliales pourraient également avoir une activité de type électrique, et les modèles de communication entre cellules gliales seraient comparables à ceux proposés entre les neurones. Il faudrait alors étudier en finesse les relations entre le fonctionnement des neurones et le fonctionnement des astrocytes, c'est-à-dire analyser comment le fonctionnement des uns modifie le mode de fonctionnement des autres, et réciproquement. Nous verrons plus loin que le mode d'interaction entre des systèmes suivant des fonctionnements différents est fondamental pour pouvoir proposer une conception du fonctionnement de leur ensemble.

Les pituicytes donnent un autre exemple des difficultés auxquelles se trouve confronté le modélisateur lorsqu'il veut rendre compte correctement du fonctionnement du système nerveux indépendamment des cellules gliales. “*Les pituicytes* représentent une catégorie particulière de cellules gliales ; ces cellules spécialisées se situent dans la neurohypophyse. *Leurs processus cytoplasmiques s'entremêlent avec les terminaisons nerveuses de fibres issues de l'hypothalamus*¹⁶” [Buser et Imbert, 1993][p. 64]. Il est alors difficile de dissocier, même fonctionnellement, l'activité électrique des cellules nerveuses et le fonctionnement des pituicytes. Comment ne pas penser que la modélisation de l'activité de cette zone cérébrale doit tenir compte d'une forme de couplage entre cellules nerveuses et pituicytes ? En est-il autrement pour le reste du système nerveux ? Il faut peut-être tenir compte, dans notre modélisation des fonctions cérébrales, à la fois des cellules nerveuses et des cellules non-nerveuses comme les cellules gliales.

Quelles contraintes nouvelles et minimales ceci impose-t-il sur nos modèles ? Si l'on souhaite tenir compte de cette diversité, il faut qu'elle se retrouve dans la description des neurones formels et des différentes cellules dont on va tenir compte. Même si le modèle construit tient compte uniquement des cellules nerveuses et non des autres cellules du cerveau, *il faut que nous soyons capables d'expliciter ce qui distingue notre modèle d'un neurone d'un modèle des autres cellules cérébrales*. Il faut que les modèles des cellules nerveuses soient suffisamment sophistiqués pour pouvoir être distingués des modèles des autres cellules cérébrales. En résumé, il faut être capable de proposer un modèle du neurone et des modèles pour les autres cellules cérébrales, et de plus il faut qu'en examinant uniquement ces modèles on soit capable de distinguer le modèle du neurone des modèles des autres cellules cérébrales.

¹⁵Cité dans [Cunningham et Waxman, 1991].

¹⁶*C'est nous qui soulignons.*

On pourrait même penser qu'il faut aller plus loin. Avec les éléments précédents, on est capable de proposer un modèle comportant à la fois des éléments formels correspondants aux neurones et des éléments formels associés aux autres cellules cérébrales. On peut alors être plus exigeant et souhaiter être capable de distinguer et de séparer nettement, au sein de ce modèle, le mode de fonctionnement de “type électrique” associé au fonctionnement du système nerveux et les autres activités qui seraient fonctionnellement indépendantes de l’activité électrique. *Cette exigence correspondrait à la nécessité de retrouver au sein du modèle l’indépendance fonctionnelle entre cellules nerveuses et cellules cérébrales non nerveuses, qui guide les observations de l’activité cérébrale.*

Peut-être doit-on exiger plus encore. En effet, les cellules gliales servent de “régulateur du fonctionnement synaptique” [Buser et Imbert, 1993]. Comme nous l'avons déjà souligné, l'activité des cellules nerveuses peut être dépendante du mode de fonctionnement des cellules non nerveuses, sans pour autant que ces cellules non nerveuses participent directement à l'activité électrique. Un modèle devrait alors tenir compte à la fois des cellules nerveuses et des cellules non nerveuses et permettre d'observer une activité électrique à laquelle ne participent pas les cellules non nerveuses. Mais ce modèle devrait aussi témoigner d'une dépendance du fonctionnement de l'activité électrique en fonction du comportement de certaines cellules non nerveuses. Ainsi ce modèle devrait permettre un enchevêtrement entre l'activité des cellules nerveuses et celle des cellules non nerveuses, de sorte qu'ait été explicitée la manière dont le mode de fonctionnement de l'une peut influencer celui de l'autre.

1.1.3. La diversité moléculaire

Nous avons déjà vu que la diversité neuronale était en partie justifiée par Sejnowski sur des critères qui portent sur le type de neurotransmetteurs activés par les différents neurones. Nous voulons insister ici sur le fait que la diversité observée dans le cerveau n'est pas seulement une diversité cellulaire, mais aussi une diversité moléculaire. Cette diversité moléculaire repose directement, en ce qui concerne les cellules nerveuses, sur *la diversité des neurotransmetteurs*, qui eux-mêmes sont susceptibles de modifier le fonctionnement effectif de l'activité électrique.

Il ne nous revient pas, en tant que modélisateur, de montrer ici la diversité des neurotransmetteurs. On la trouvera présentée dans tout manuel de biochimie du système nerveux. Elle est en partie évoquée et décrite chez J.-D. Vincent [Vincent, 1988], par exemple, ou dans la conférence intitulée *La chimie des communications cérébrales* et prononcée par J.-P. Changeux le 27 janvier 1994 au Collège de France¹⁷.

¹⁷Nous avons déjà cité cette conférence dans le chapitre 3, section 1. Elle est disponible sous la forme d'une cassette vidéo [Changeux, 1994].

Il faut noter de plus qu'un enchevêtrement est potentiellement présent entre l'“activité électrique” et l’“activité moléculaire”. “D'ailleurs la signalisation chimique existe toujours dans les organismes à systèmes nerveux complexes ; ces êtres vivants sont dotés d'un système endocrinien dont *les messagers hormonaux agissent en synergie avec les signaux nerveux*¹⁸” [Rosenzweig et Leiman, 1991][p. 96]. Ainsi l'existence d'une rupture proposée par le modélisateur entre l'activité nerveuse et l'activité hormonale est loin d'être évidente. Cette séparation n'est peut-être pas pertinente lorsque l'on souhaite rendre compte des capacités fonctionnelles cérébrales dans leur ensemble. On peut ainsi se demander si certains phénomènes moléculaires, et en particulier les phénomènes hormonaux, ne doivent pas également être considérés comme faisant partie intégrante de la cognition, du fait de la nécessité de leur présence dans l'organisme pour développer des capacités cognitives.

Quelles contraintes nouvelles et minimales ceci impose-t-il sur nos modèles ? Pour tenir compte de la diversité moléculaire, si cela est nécessaire, il faut d'abord préciser explicitement comment les différentes molécules en jeu interviennent dans le mode de fonctionnement du neurone formel. Il faut donc avoir tout d'abord un modèle suffisamment sophistiqué du neurone formel pour envisager comment tenir compte ou non de certains aspects de la diversité moléculaire. Nous n'avons actuellement aucune idée précise sur la question, des modèles plus sophistiqués du neurone formel nous paraissant un préalable nécessaire à toute discussion sur ce sujet.

1.2. Rejet de l'étude d'une unique structure logique

1.2.1. Éléments pour rejeter une modélisation “trop fine” du neurone

On peut se demander dans quelle mesure toutes les complications envisagées précédemment sont bien nécessaires. S'il en était ainsi, serions-nous encore capables d'étudier ces modèles et auraient-ils encore une quelconque utilité ? Seraient-ils à même de nous apporter des connaissances supplémentaires sur le mode de fonctionnement des activités cérébrales ?

La modélisation nécessite “la” bonne simplification, comme nous l'avons déjà répété à plusieurs reprises. On peut alors se demander dans quelle mesure il est bien nécessaire de tenir compte de la diversité des éléments cérébraux dans nos modèles. On peut citer Farmer [Farmer, 1990], qui propose quelques arguments pour rejeter une modélisation trop détaillée des neurones individuels. “Il y a plusieurs raisons qui nous poussent à laisser de côté les contraintes de la modélisation des neurones réels :

- i) nous ne comprenons pas le comportement des neurones réels,

¹⁸ C'est nous qui soulignons.

- ii) même si on les comprenait, il serait computationnellement inopérant d'implémenter le comportement complet des neurones réels,
- iii) il est peu vraisemblable que nous ayons besoin de la complexité complète des neurones réels pour résoudre des problèmes d'apprentissage de machine,
- iv) en expérimentant différentes approches des modèles simplifiés de neurones, on peut espérer extraire les principes de base avec lesquels ils opèrent et découvrir quelles propriétés sont vraiment essentielles pour l'apprentissage" [Farmer, 1990][p. 158].

Le premier argument est évidemment sans intérêt, car nous pouvons comprendre demain ce que nous ne comprenons pas aujourd'hui et que nous n'avons pas démontré comme intrinsèquement incompréhensible. L'argument est par contre dangereux. Il montre comment le modélisateur peut se "défausser" sur un autre, en l'occurrence le biologiste, des difficultés qu'il n'arrive pas à surmonter. Il s'agit ici d'une manière étonnante de concevoir la pluridisciplinarité. Ceci est d'autant plus vrai que, comme nous l'avons montré dans le chapitre 4, *le propre de la modélisation est de permettre de formuler de nouvelles questions aux différents partenaires qui coopèrent*.

Le second argument est une autre manière, souvent pratiquée, de se "défausser" sur une autre discipline, en l'occurrence la technologie informatique. L'argument consiste à dire, implicitement, que les machines actuelles ne sont pas suffisamment puissantes et que, finalement, les modèles actuels seront très efficaces sur les machines de demain¹⁹. On peut dire, sans préjuger de ce que permettront la science et la technologie de demain, qu'il s'agit là d'une "attitude de prophète" plutôt que de scientifique. Une fois de plus il s'agit de masquer que, sans doute, si on n'arrive pas aux résultats escomptés, c'est finalement parce qu'on s'y prend mal.

Le troisième argument est celui, fondamental, qui justifie toute modélisation. Rendre compte d'un phénomène avec des modèles est évidemment inutile si on ne croit pas, *a priori*, que tous les éléments présents ne sont pas nécessaires dans leur identité propre. La question qui se pose alors est de savoir quels sont les éléments effectivement nécessaires au modélisateur par rapport au but affiché. Par ailleurs le troisième argument de Farmer nous montre bien que ce qui l'intéresse est le problème de l'apprentissage des machines. En fait l'objectif des sciences cognitives²⁰ est quant à lui beaucoup plus ambitieux puisqu'il s'agit, qu'on le dise ou non officiellement,

¹⁹Il est à noter que cet argument n'est jamais accompagné, contrairement à ce que l'on trouve chez von Neumann, de spécifications précises des outils technologiques qui seraient nécessaires.

²⁰L'objectif "limité" affiché par Farmer ne correspond d'ailleurs pas non plus aux objectifs "ambitieux" de l'Artificial Life, courant de recherche dont Farmer fait partie clairement et officiellement. On peut même certainement considérer Farmer comme le théoricien, la "tête pensante" de ce mouvement. Nous avons déjà présenté et commenté les objectifs d'Artificial Life et la méthodologie qui leur est associée dans la section 3 du chapitre 3.

de comprendre sur une base biologique le fonctionnement global du cerveau/esprit, c'est-à-dire de la conscience, de la pensée et des fonctionnalités qui peuvent leur être associées.

Le quatrième argument porte, comme le troisième, sur la conception nécessaire au modélisateur selon laquelle certaines simplifications permettent de comprendre certains comportements. On peut essayer de suivre la foi²¹ de Farmer et considérer que des modèles simplifiés et tous identiques de neurones permettront d'aboutir à une compréhension de l'apprentissage. Il faut cependant une foi plus forte encore pour croire que ces modèles simplifiés et identiques de neurones permettront d'expliquer la conscience scientifiquement et sur une base biologique.

1.2.2. De la thèse de l'identité logique à l'isolation normative

Selon nous, l'argument le plus fort qui justifie une conception des modèles du système nerveux s'appuyant sur des modèles de neurones tous identiques²² est celui de l'identité logique²³. L'identité qui est en jeu ici n'est évidemment pas une identité individuelle, chaque neurone étant "irremplaçable" à l'identique. Deux cellules pyramidales ne sont jamais exactement identiques, puisqu'elles n'ont jamais exactement le "même" passé²⁴. Quand on parle d'identité logique on ne nie pas non plus le fait attesté par la neurobiologie selon lequel une cellule pyramidale n'a pas exactement les mêmes caractéristiques "fines" qu'une cellule de Purkinje. La **thèse de l'identité logique** consiste à affirmer que l'on peut rendre compte du *fonctionnement* global de tous les systèmes de cellules nerveuses dans un même cadre logique, par exemple connexionniste. La thèse de l'identité logique s'appuie sur une présentation simplifiée des résultats de la neurobiologie selon lesquels le propre de l'activité cérébrale est celle du système nerveux dont les éléments caractéristiques se résument en une *activité électrique* entre cellules nerveuses. La thèse de l'identité logique s'appuie alors essentiellement sur l'unicité logique du fonctionnement d'une activité électrique. Cet argument ne nie pas l'existence de variations qui sont localement observables. Mais on considère que ces variations peuvent être intégrées dans la modification de la valeur de quelques paramètres. Ces variations n'affectent donc pas la structure logique dans laquelle on veut en rendre compte. Les dynamiques

²¹Nous utilisons volontairement ce terme, mais sans considération péjorative. Il s'agit bien ici d'une croyance, d'une prise de position personnelle qui constitue un acte d'engagement individuel. Il est nécessaire de considérer cette croyance sous cet angle, tant qu'elle n'a pas montré ses preuves grâce à un faisceau d'arguments assis sur une base scientifique et communément partagée. "Aussi intense soit-il, un sentiment de conviction ne peut jamais justifier un énoncé" [Popper, 1978][p. 43].

²²Les neurones modélisés peuvent être éventuellement tous identiques, sans être simples pour autant. Ce que nous discutons ici est seulement le fait de modéliser les cellules nerveuses toutes à l'identique.

²³Nous appellerons désormais cet argument la **thèse de l'identité logique**.

²⁴Il faudrait sinon qu'elles mettent en œuvre les mêmes éléments de matière. Cette contrainte étant d'ailleurs elle-même nécessaire mais non suffisante.

peuvent être légèrement différentes d'un système à l'autre, mais elles resteront du même type²⁵. L'argument de l'identité logique ne minimise pas la diversité observée mais considère que cette variabilité n'affecte pas la structure logique des modèles utilisés, eu égard aux objectifs de la modélisation. Ainsi, selon la thèse de l'identité logique, la diversité est "apparente". On peut la ramener à une unique structure commune. La diversité se retrouve alors dans les variations des paramètres.

En fait la thèse de l'identité logique va plus loin. Elle considère implicitement qu'une unique structure logique, correspondant à l'activité électrique du système nerveux, permet de rendre compte du fonctionnement cérébral dans son ensemble. Cette unique structure logique, celle des petits réseaux de neurones formels, pouvant se trouver à différents niveaux et en différentes zones du cerveau modélisé.

La thèse de l'identité logique s'appuie en résumé sur le constat que l'on peut grossièrement expliquer le fonctionnement de tous les groupes de cellules nerveuses dans un unique cadre explicatif. Une analogie, un "isomorphisme" est ainsi établi entre le "monde réel" et le modèle connexionniste²⁶.

De cet "isomorphisme" issu de la thèse de l'identité logique, on déduit qu'il est possible de se contenter de travailler sur l'unique structure logique à laquelle chacun des espaces considérés est isomorphe. Dans la suite de cette section nous admettrons la validité de la thèse de l'identité logique. Dans ce contexte, lorsque l'on souhaite étudier différents espaces, nous appellerons **isolation normative** *le fait de travailler uniquement sur la structure logique à laquelle chacun des espaces dont on veut parler est considéré comme isomorphe*.

On peut interpréter l'isolation normative dans le cadre des figures présentées dans la section 1 du chapitre précédent. L'attitude consiste alors à fixer une structure logique particulière, un modèle, et à travailler exclusivement sur lui. On se contente alors de deux domaines, le modèle et la simulation auxquels on peut éventuellement ajouter l'analyse de modèle et l'analyse de simulation. En aucun cas on ne peut parler de modélisation au sens propre comme nous l'avons définie dans le chapitre précédent.

Le but des paragraphes qui suivent est de montrer que même si on admet la thèse de l'identité logique, on ne peut pas se contenter de travailler sur une unique

²⁵On pourra, par exemple, toujours en rendre compte sous la forme d'équations aux dérivées partielles. On trouve clairement cette attitude chez Stephen Grossberg, comme le montre Patrice Prez [Prez, 1997].

²⁶Rappelons cependant que la notion d'isomorphisme en mathématiques est une relation entre deux ensembles de même nature (espaces vectoriels normés, espaces topologiques, etc.). Le modélisateur connaît le type formel des modèles connexionnistes. Il ne peut cependant rien affirmer sur la nature, en tant qu'espace "abstrait", du monde "réel" et matériel des cellules nerveuses, sans prendre une position ontologique, ce qui ne relève pas de ses attributions. Le monde réel et le modèle connexionniste n'ont donc pas *a priori* la même nature symbolique. Dans le cadre de la thèse de l'identité logique, la notion d'isomorphisme est donc à employer avec des guillemets.

structure logique. Nous allons donc montrer que l’isolation normative, au sens où nous l’avons définie, ne peut pas être exercée sans d’extrêmes précautions. Nous voulons insister sur le fait que la difficulté n’est pas liée à l’existence ou non d’un isomorphisme, mais plutôt au fait que l’on ait ou non à expliciter l’isomorphisme qui est en jeu.

Comme nous allons le montrer, la thèse de l’identité logique qui devait initialement nous permettre de supprimer *la diversité des structures considérées nous amène en fait à considérer une nouvelle diversité potentielle, celle des isomorphismes entre les structures considérées et la structure logique prise comme référence*. La thèse de l’identité logique a bien supprimé la diversité initiale, mais elle en a peut-être introduit une nouvelle.

1.2.3. Un exemple formel des difficultés que présente l’isolation normative

Comme nous l’avons déjà souligné, la notion d’isomorphisme ne peut pas être employée au sens propre dans le contexte de la thèse de l’identité logique. Les difficultés que nous voulons mettre en évidence ne sont pas du tout liées à cet aspect des choses. Aussi, dans ce paragraphe, nous présentons ces difficultés dans un contexte purement formel, celui des espaces de Hilbert.

Nous nous plaçons ici dans le cadre purement formel des espaces de Hilbert de dimension infinie et séparables. **On peut démontrer mathématiquement que dans ce cadre la thèse de l’identité logique est vérifiée.** Il est à noter également que ces espaces sont d’une extrême diversité. On peut citer l’espace l^2 des suites de nombres réels de carré sommable, l’espace $L^2(\mathbb{R})$ des fonctions de carré sommable sur \mathbb{R} , l’espace de Sobolev $H^1(\mathbb{R})$ des fonctions de $L^2(\mathbb{R})$ dont toutes les dérivées partielles sont aussi dans $L^2(\mathbb{R})$. Ces trois espaces sont isomorphes, ce qui est en conformité avec la thèse de l’identité logique. On peut décider de choisir comme espace de référence l’espace $L^2(\mathbb{R})$. On est bien ici dans le contexte souhaité lorsque l’on fait l’hypothèse de l’identité logique. L’isolation normative consiste alors à se contenter d’étudier l’espace $L^2(\mathbb{R})$, les autres espaces considérés pouvant s’y ramener.

On obtient alors un isomorphisme φ_1 qui permet de passer de l’espace l^2 à l’espace de référence $L^2(\mathbb{R})$. On peut aussi trouver un isomorphisme φ_2 qui permet de passer de l’espace $H^1(\mathbb{R})$ à l’espace de référence $L^2(\mathbb{R})$. On a cependant $\varphi_1 \neq \varphi_2$. Il n’est donc pas possible de passer par une même transformation des espaces l^2 et $H^1(\mathbb{R})$ à l’espace $L^2(\mathbb{R})$. On peut en conclure que l’existence d’un isomorphisme pour chaque espace de Hilbert vers $L^2(\mathbb{R})$ ne nous assure pas pour autant que le même isomorphisme s’applique pour passer de tout espace de Hilbert vers $L^2(\mathbb{R})$. La difficulté est liée au fait que l’on veut désormais passer “en même temps” des différents espaces de Hilbert vers $L^2(\mathbb{R})$. Pour aller plus loin, il faut étudier plus

finement les relations particulières entre ces espaces de Hilbert lorsque l'on veut les “transporter ensemble” vers $L^2(\mathbb{R})$.

L’isolation normative est donc possible si on est capable d’exhiber les différents isomorphismes qui permettent de basculer des espaces considérés vers l’espace de référence. Que se passe-t-il si on souhaite ensuite étudier les deux espaces de Hilbert non pas séparément, mais en conservant les relations qu’ils ont entre eux ? Le fait de préciser l’isomorphisme qui permet pour chaque espace considéré de passer à l’espace de référence ne nous suffit pas pour savoir comment ces espaces *considérés en même temps* peuvent être transformés ou non en l’espace de référence. Ceci dépend des relations particulières qui peuvent exister entre ces espaces. Considérons par exemple $L^2(\mathbb{R})$ et $H^1(\mathbb{R})$, et essayons de les transporter par un isomorphisme φ_3 sur $L^2(\mathbb{R})$. L’isomorphisme φ_3 envoie alors nécessairement $L^2(\mathbb{R})$ sur $L^2(\mathbb{R})$, l’identité de $L^2(\mathbb{R})$ étant un candidat potentiel d’un tel isomorphisme. Mais comme $H^1(\mathbb{R})$ est un sous-espace strict de $L^2(\mathbb{R})$, alors l’image de $H^1(\mathbb{R})$ par φ_3 , $\varphi_3(H^1(\mathbb{R}))$ est un sous-espace strict de $\varphi_3(L^2(\mathbb{R})) = L^2(\mathbb{R})$. Donc $\varphi_3(H^1(\mathbb{R})) \neq L^2(\mathbb{R})$. On ne peut donc pas trouver d’isomorphisme φ_3 qui envoie en même temps $L^2(\mathbb{R})$ et $H^1(\mathbb{R})$ sur $L^2(\mathbb{R})$. Ainsi, même si $L^2(\mathbb{R})$ est une structure de référence pour chaque espace précédent pris isolément, il ne peut pas en être de même lorsque l’on souhaite avoir une approche globale qui considère ces différents espaces en même temps. En l’occurrence la relation entre les deux espaces considérés est une relation d’inclusion stricte.

Concrètement, on peut en conclure que, si des “isomorphismes” sont concevables entre le monde matériel et les modèles connexionnistes, ce ne sont pas forcément les mêmes qui sont en jeu pour chaque système de cellules nerveuses matérielles. Même si on soutient la thèse de l’identité logique, il est nécessaire d’exhiber les “isomorphismes” qui permettent de basculer de ces cellules matérielles vers l’espace de référence. Ainsi, même dans le cas où on considère une unique structure logique, la diversité des cellules matérielles nécessite d’exhiber les isomorphismes qui permettent de passer à la structure de référence. *Le risque est alors grand que la diversité des cellules matérielles étudiées, qui a été réduite grâce à la thèse de l’identité logique, se transforme en une diversité des isomorphismes étudiés.*

On peut souhaiter de plus être capable de rendre compte de ces différentes cellules ou de ces différents espaces, en même temps, ce qui est nécessaire si on veut entamer une approche globale de la cognition. Il faut alors être capable d’exhiber un isomorphisme qui permette de passer en même temps de l’ensemble des espaces considérés vers l’espace de référence. *L’isolation normative, en oubliant d’étudier les relations entre les espaces considérés initialement, n’est alors pas suffisante.* Pour étudier cet aspect, il est au minimum nécessaire d’exhiber les isomorphismes qui permettent de passer à la structure de référence, ce qui oblige à aller plus loin que l’isolation normative. Concrètement cela signifie que l’on ne peut pas utiliser les structures formelles, par exemple de type connexionniste, sans avoir

en même temps une démarche modélisatrice qui tienne compte explicitement des ensembles de cellules nerveuses dont on veut effectivement rendre compte. On entre alors dans une démarche de modélisation à proprement parler qui tente d'expliquer comment a été réalisé le passage des éléments dont on veut rendre compte à la structure formelle qui constituera le modèle.

1.2.4. Les difficultés de l'isolation normative associées à un échange de quantificateurs

La thèse de l'identité logique est certainement liée en partie à la spécialisation des disciplines. Cette spécialisation engage vivement à regarder chaque expérience, chaque élément étudié, de manière autonome. On en vient alors à formuler un résultat qui semble général : *pour toute expérience E sur les cellules nerveuses, il existe une transformation T, une façon d'observer l'expérience, telle que l'on peut rendre compte des cellules nerveuses dans le cadre connexionniste de Farmer F*. On peut traduire cela sous la forme $\forall E \exists T$ tel que $T(E) \in F$. Il s'agit là de la thèse de l'identité logique. Une recherche de globalisation entraîne, quant à elle, une nouvelle question qui interroge la générnicité des résultats obtenus dans chaque cas particulier. Désormais on ne demande plus si pour chaque exemple il est possible de trouver une abstraction possédant une certaine forme logique, mais plutôt si la transformation obtenue peut correspondre à chacun des exemples susceptibles d'être étudiés. On s'appuie sur l'énoncé : *il existe une transformation T telle que, pour toute expérience E sur les cellules nerveuses, on peut rendre compte des cellules nerveuses dans le cadre connexionniste de Farmer F*. La traduction logique de cet énoncé devient alors $\exists T \forall E$ tel que $T(E) \in F$. On peut alors se contenter d'étudier la structure F . Pour chaque expérience E , il suffit alors d'appliquer la même transformation T pour se ramener à la structure F . C'est cet énoncé qui est nécessaire pour que l'isolation normative soit justifiée.

Il est à noter tout d'abord que les deux énoncés logiques ne sont pas identiques. En effet, dans le premier énoncé, la transformation particulière peut dépendre de l'expérience considérée. Au contraire dans le second cas la transformation doit être définie avant toute expérience et ne doit pas dépendre des expériences. Les deux énoncés n'ont donc rien à voir l'un avec l'autre. L'isolation normative pratiquée actuellement consiste cependant à raisonner comme si on avait le deuxième énoncé alors que seul le premier est justifié par les expériences et peut être généralisé par la thèse de l'identité logique.

Il nous semble donc nécessaire de rejeter *a priori* l'isolation normative qui consiste à étudier une unique structure logique. **On peut conserver la thèse de l'identité logique²⁷, mais il faut prendre la précaution d'expliquer com-**

²⁷Des arguments auraient pu être avancés contre la thèse de l'identité logique elle-même. On peut se demander par exemple pourquoi des éléments matériellement si

ment on fait pour passer de chaque structure particulière étudiée à la structure que l'on a choisie comme référence. *Dans ce sens, la diversité des structures considérées n'a pas vraiment disparu avec la thèse de l'identité logique.* La diversité des structures se retrouve désormais *a priori* dans la diversité des isomorphismes. Ceci ne pose aucune difficulté particulière si on étudie chaque élément individuellement. Ce n'est pas le cas cependant si on doit étudier ces structures en même temps.

J'engage vivement le lecteur à relire le paragraphe précédent sur la diversité des éléments cérébraux à la lumière de ce que nous venons de montrer sur l'éventuelle nécessité de tenir compte, en même temps, des diverses structures logiques, en précisant dans chaque cas les "isomorphismes" particuliers qui sont en jeu avec la structure de référence.

Quelles contraintes nouvelles et minimales ceci impose-t-il sur nos modèles ? La contrainte principale est que, finalement, une isolation normative quant au fonctionnement du neurone formel, c'est-à-dire le choix d'une structure logique unique, ne résout pas toutes les difficultés. **Il convient également de préciser comment on passe de chaque neurone biologique à ce neurone formel spécifique**, afin de pouvoir établir par la suite les liens entre ces manières de passer d'un neurone biologique au modèle du neurone formel.

Si l'objet est d'étudier le cerveau dans son ensemble, ou certaines zones de taille importante du cerveau, relativement à la taille d'un neurone, les contraintes précédentes sont nécessaires. En effet, dans ce cadre, on est obligé de proposer une unique transformation permettant de passer de cet ensemble de grande taille au modèle formel. Cette transformation doit alors être explicite, en précisant les échelles en jeu, etc. Tout ceci peut et doit ensuite être appliqué à chacun des éléments de l'ensemble étudié.

Ainsi une véritable modélisation est nécessaire comme va-et-vient permanent, en particulier entre des observations précises et fines des éléments cérébraux et le modèle. Cette modélisation nécessite également l'explicitation des critères permettant le passage de ces observations au modèle et pouvant être appliqués à l'ensemble des éléments dont le modèle doit tenir compte.

divers sont présents pour un comportement finalement identique ? On peut se demander également pourquoi des éléments "logiquement identiques" se différencient et se spécialisent au cours de l'ontogénèse pour finalement avoir un comportement identique. Comment peut-on interpréter ces arguments au regard du rasoir d'Ockham ? Ces arguments vont certainement également à l'encontre de l'hypothèse de l'identité logique.

1.3. Mixité et hétérarchie des structures logiques

1.3.1. Diversité des structures logiques

Il est bien connu que la mécanique quantique et la mécanique classique obéissent à des causalités différentes. Éventuellement, il peut “exister” d’autres types de causalités dont on pourrait rendre compte avec des formalismes à construire et qui obéiraient éventuellement à des logiques différentes²⁸. On peut concevoir alors que chaque “type” de phénomène nécessite une logique particulière qui lui est adaptée. C'est d'ailleurs dans cet esprit qu'il faut comprendre le travail de von Neumann sur une logique probabiliste correspondant aux automates susceptibles d'intégrer des “erreurs” ou des “imperfections” [von Neumann, 1956].

Les philosophes parlent de causalités différentes auxquelles les logiciens font correspondre des logiques différentes. En tant que modélisateur, c'est-à-dire également de manipulateur de systèmes formels, nous garderons un point de vue et un vocabulaire correspondant aux systèmes formels. Aussi nous parlerons de structures logiques différentes, c'est-à-dire de systèmes formels différents construits sur des logiques différentes. Dans notre sens, il est donc possible de définir des structures logiques obéissant à des logiques non réductibles les unes aux autres. Cette conception ne pose, au premier abord, aucune difficulté particulière. On peut citer comme exemple les modèles en mécanique quantique, qui sont tout simplement construits comme des structures logiques différentes de celles qui sont adaptées par exemple à la mécanique newtonienne.

1.3.2. Mixité des structures logiques

Jusqu'à présent les différentes disciplines scientifiques ont utilisé systématiquement une unique structure logique. Ceci signifie que les modèles proposés jusqu'à présent obéissaient chacun à une seule logique. *Ceci était possible dans ces différentes disciplines parce que les problèmes posés, tels qu'ils étaient découpés, permettaient de le faire.* Ceci était implicite dans les modèles de la mécanique newtonienne. Cela a nécessité la création d'une logique propre pour la mécanique quantique. En Sciences Cognitives, jusqu'à présent, on a fait comme si une telle approche était possible. Il nous semble cependant que cette attitude n'est pas possible lorsque l'on souhaite aborder l'étude globale des fonctionnalités du cerveau²⁹. La difficulté n'est pas du tout de nature ontologique ou métaphysique, mais simplement liée aux contraintes de la modélisation.

²⁸Il faut noter à ce propos les efforts de von Neumann pour définir une logique de la mécanique quantique qui ne se réduirait pas à la logique classique [von Neumann et Birkhoff, 1936].

²⁹On trouve de nombreuses remarques de ce type, très détaillées, chez von Neumann, qui estime que l'on doit considérer le cerveau comme un système mixte, avec, selon ses termes, des éléments analogiques et aussi d'autres digitaux [von Neumann, 1951][von Neumann, 1966][von Neumann, 1992].

Pour modéliser, il faut “découpler” une zone, un ensemble de phénomènes, du reste du monde. “La composition d'un modèle simple pour une boîte fermée presuppose que certaines variables sont seulement faiblement couplées avec le reste de celles qui relèvent du système” [Wiener et Rosenblueth, 1945][p. 319]. En l'occurrence, nous supposerons que ce que nous cherchons à comprendre est le fonctionnement de l'activité cérébrale, et que cette compréhension ne nécessite que la modélisation du fonctionnement des ensembles de neurones. L'exemple des pituicytes vient alors nous rappeler que le découplage ou le faible couplage entre les neurones et les autres cellules du cerveau ne correspond pas toujours aux observations biologiques. “[Les] processus cytoplasmiques [des pituicytes] s'entremêlent avec les terminaisons nerveuses de fibres issues de l'hypothalamus³⁰” [Buser et Imbert, 1993][p. 64].

Pour modéliser de tels aspects de la cognition, il est nécessaire de “tenir” en même temps et de faire coopérer des structures logiques différentes correspondant en fait à des éléments matériels différents. Cette imbrication des structures logiques est nécessaire pour maintenir une certaine unité présupposée du cerveau et même plus fréquemment du néocortex pour qu'ils soient capables de cognition.

Pour modéliser, des simplifications sont nécessaires. On peut cependant envisager plusieurs façons de simplifier un phénomène, ainsi que plusieurs types de simplifications. Dans la deuxième section de ce chapitre, nous envisagerons une autre simplification, tout à fait différente de celles envisagées traditionnellement pour aborder l'étude de la cognition.

1.3.3. Structures logiques organisationnelles d'un système et structures logiques des unités qui le constituent

À chaque structure logique on pourra associer des éléments symboliques constitutifs qui joueront le rôle d'unités élémentaires³¹. Les différentes unités symboliques, qui sont en jeu dans ces modèles comportant différentes structures logiques, sont définies avec des éléments matériels différents. On peut toujours affirmer que la différence liée au type d'élément matériel n'est pas pertinente pour la modélisation. Il suffit alors d'utiliser des symboles du même type pour représenter des éléments matériels différents. Comme ces symboles ont les mêmes types, la différence entre eux porte uniquement sur leur nom. Comment établir une relation entre ces symboles ou ces systèmes symboliques ? Le premier système symbolique est “isomorphe” au premier type d'élément de matière. Le deuxième système symbolique est “isomorphe” au

³⁰L'exemple ici ne porte pas sur le néocortex. Nos connaissances très réduites en biologie et le temps limité pour argumenter cet aspect des choses ne nous a pas permis d'approfondir la question. Rien ne permet cependant de penser que les choses sont différentes sur le fond pour le néocortex.

³¹Pour simplifier les raisonnements suivants, nous faisons ici comme si les structures logiques étaient constituées d'éléments de base, comme pour la causalité des chocs de la mécanique newtonienne qui est celle régissant le fonctionnement des réseaux de neurones formels (voir chapitre 2 section 2).

second type d'élément matériel. Mais quelle est alors la relation entre le premier et le second type d'élément matériel ? Cette relation n'a pas été conservée par les deux isomorphismes précédents³². Il est très difficile actuellement de préciser l'importance relative des types d'éléments matériels particuliers, en relation avec leurs rôles respectifs. Si on suppose que toutes ces entités matérielles particulières sont identifiables à un type symbolique, on ne sait pas comment il faut réaliser cette identification des éléments entre eux.

Considérer que l'on peut expliquer³³ les systèmes vivants en termes de relations et non à partir des propriétés des composants constitue une position³⁴ qui n'est ni triviale, ni la seule envisageable³⁵. Elle n'est pas triviale car elle suppose qu'il est possible de dégager des phénomènes dynamiques une structure logique indépendante de sa réalisation matérielle. *Ceci constitue une hypothèse de travail*. Elle est défendable si on est toujours capable de séparer³⁶ certains facteurs considérés comme organisationnels alors que d'autres seraient liés uniquement à des contraintes de matérialisation. On peut par exemple envisager que certaines propriétés organisationnelles d'un système ne peuvent être expliquées qu'en tenant compte systématiquement de sa description matérielle. La modélisation nécessite le découplage de phénomènes, alors qu'ici il s'agit de proposer une possibilité de découplage entre les propriétés organisationnelles du phénomène observé et les propriétés des éléments matériels qui constituent l'organisation. *Les deux types de découplage n'ont donc rien à voir l'un avec l'autre et le second type de découplage n'a rien à voir avec une quelconque nécessité de la modélisation.*

Pour illustrer cela, on peut envisager, par exemple, que certaines propriétés dynamiques d'un réseau (matériel) de neurones ne sont pas séparables du fait que ce réseau est matériellement observé sur des neurones dont certaines propriétés biologiques semblent difficiles à rapporter en termes purement organisationnels et sont tout de même nécessaires pour pouvoir effectuer les observations qui nous intéressent sur le réseau. Ceci peut être lié, par exemple, au fait que les neurones sont des cellules. Le fait pour une unité de fonctionner comme une cellule n'a rien à voir avec les propriétés organisationnelles du réseau et ne peut éventuellement pas être rapporté en termes de propriétés organisationnelles. Ceci est terriblement ennuyeux pour le modélisateur si on envisage, de plus, que les propriétés du réseau

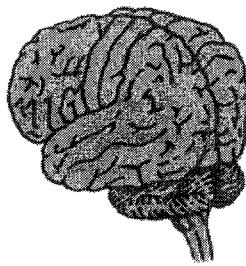
³²Voir le paragraphe précédent sur la thèse de l'identité logique.

³³C'est actuellement sur cette hypothèse que reposent généralement les explications et les travaux en Sciences Cognitives.

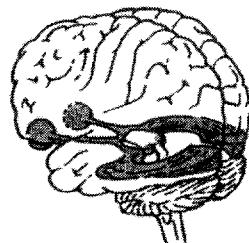
³⁴Voir le chapitre 2 section 1 pour les développements associés à cette position générale prise par les chercheurs en Sciences Cognitives.

³⁵**On verra l'ébauche d'une solution alternative dans la deuxième section de ce chapitre.**

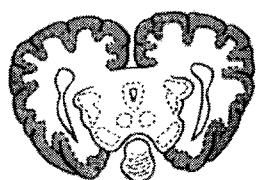
³⁶La notion de séparation n'est jamais absolue. Elle nécessite un observateur qui la réalise et dépend toujours de l'observation du phénomène et des aspects que l'on est prêt ou non à négliger. La possibilité de séparation, de découplage de phénomènes, est le propre de la modélisation, comme nous l'avons rappelé avec la citation de Wiener.



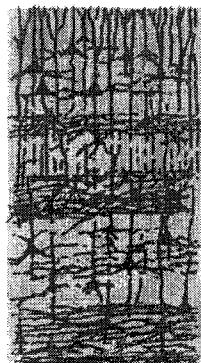
Le cerveau entier
(environ 15 cm)



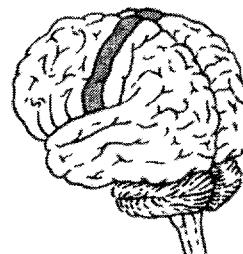
Un système cérébral comportant plusieurs régions cérébrales :
le système visuel



Une région cérébrale : le cortex cérébral
(cortex environ 3 mm d'épaisseur)



Une coupe transversale d'une sous-région cérébrale :
une colonne corticale du cerveau



Une sous-région cérébrale :
le cortex moteur primaire

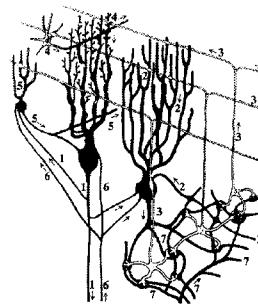
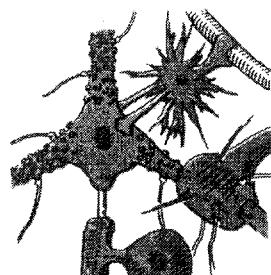
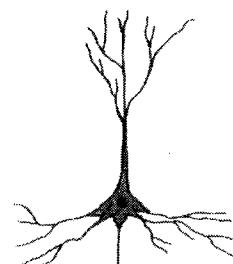


Schéma de câblage du cervelet
1 - Cellule de Purkinje 2 - Cellule de Golgi
3 - Cellule en graine et fibre parallèle
4 - Cellule étoilée 5 - Cellule en panier
6 - Fibre griseante 7 - Fibre mousseuse

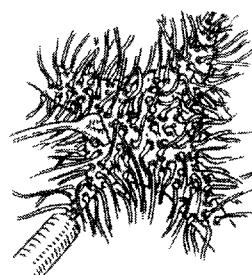
Un circuit local constitué de relations entre cellules
nerveuses : les 7 éléments du circuit cérébelleux fondamental



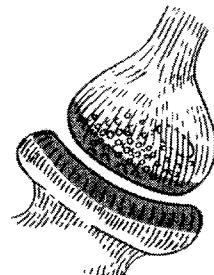
Un environnement local d'un neurone
(gros corps cellulaire 0,1 mm de diamètre)



Une cellule nerveuse :
une cellule pyramidale



Un regroupement synaptique



Une synapse (bouton terminal 1µm de diamètre
et faille synaptique 20nm de large)

FIGURE 3 : Principaux niveaux d'organisation cérébrale, du cerveau entier aux contacts synaptiques. Les différents niveaux se lisent de gauche à droite et de haut en bas. On a indiqué également quelques ordres de grandeurs des tailles (d'après [RL91][p33] et [Ava91][p53] pour le circuit cérébelleux fondamental).

sont sans doute possibles uniquement parce que les neurones sont des cellules et fonctionnent comme telles.

Pour modéliser cet aspect des choses, il faut envisager de construire des modèles comportant au moins deux niveaux. Au premier niveau on retrouve la structure fonctionnelle de chaque neurone individuel. Au second on retrouve les interactions entre ces neurones formels. Il faut bien noter que ceci est radicalement différent de ce qui se pratique actuellement pour l'étude des réseaux de neurones. En effet cette attitude nécessite une modélisation³⁷ qui tienne compte en même temps des deux niveaux. On trouve bien alors une structure logique mixte comportant une première structure logique pour les neurones et une seconde pour leurs interactions. Si les logiques de fonctionnement des neurones et du système ne sont pas identiques³⁸, on peut alors bien parler de structure logique mixte. En fait il s'agit déjà ici d'une structure logique mixte et hiérarchique puisque l'une des sous-structures logiques (celle du réseau) n'est pas au même "niveau" que l'autre (celle des neurones). Nous parlerons alors d'hétérarchie, comme nous allons le voir.

1.3.4. Organisations et hiérarchies cérébrales

Les biologistes considèrent unanimement que le cerveau est organisé. Cette structuration prend de nombreuses formes. L'organisation et la structuration du cerveau se retrouvent en fait à de nombreuses échelles et on parle de "niveaux d'organisation". Ils sont considérés à la fois comme des niveaux anatomiques et des niveaux de traitement, c'est-à-dire des niveaux fonctionnels. On trouvera sur la figure 3 les principaux niveaux considérés en neurobiologie.

1.3.5. Hiérarchie de structures logiques mixtes : l'hétérarchie

Nous avons envisagé précédemment une coopération de structures logiques différentes. La structuration du cerveau nous amène également à envisager une organisation hiérarchisée de structures logiques. Une telle hiérarchie peut être homogène et comporter des structures logiques identiques. Elle peut aussi être hétérogène et comporter des structures logiques distinctes. On définit alors l'**hétérarchie** comme une *structure composée d'une hiérarchie de structures correspondant à des logiques différentes*. Selon nous une modélisation des fonctionnalités cérébrales nécessite

³⁷Il faut bien comprendre le mot au sens fort et non seulement comme l'étude d'un système symbolique qui obéirait juste à quelques contraintes.

³⁸Ceci n'a rien à voir avec les débats sur le réductionnisme qui ne nous intéressent pas du tout. Nous nous plaçons uniquement du point de vue des systèmes symboliques. Notre perspective permet tout aussi bien de défendre ou non une conception réductionniste. Une fois de plus cette question est d'ordre métaphysique et ne concerne pas du tout le modélisateur, les modèles qu'il propose se devant d'être "ontologiquement neutres".

l'hétéarchie, puisqu'elle suppose à la fois une hiérarchie de structures et une mixité des structures logiques considérées.

On ne connaît pas actuellement de modèles dont les structures pourraient être considérées comme hétéarchiques, dans le sens que nous venons de définir.

Pour l'instant nous sommes incapables de préciser quels sont les niveaux pertinents pour évoquer l'hétéarchie. Nous ne pouvons pas désigner, parmi les niveaux distingués sur la hiérarchie de la figure 3, ceux dont il faut effectivement tenir compte et qui correspondent à des structures logiques mixtes.

2. QUE NOUS IMPOSE CETTE NOUVELLE COMPLEXITÉ ?

Nous chercherons ici à énoncer et à envisager certaines des conséquences de la thèse suivante : **la complexité associée à la cognition nécessite la matérialité**. Cette thèse permet d'insister sur le fait que les systèmes symboliques actuels ne permettent pas de tenir compte d'un certain nombre d'aspects qui sont certainement essentiels quand on souhaite aborder la cognition. Il faut la considérer comme une alternative à la thèse symboliste de Herbert Simon¹. Selon nous la thèse précédente doit être un guide, un fil conducteur qui permet d'établir et d'envisager comment on peut dépasser les limites actuelles des études sur la cognition, eu égard aux objectifs de compréhension effectivement affichés.

2.1. Ouvrir l'horizon et concevoir de nouvelles machines symboliques

L'image d'Épinal que les sciences cognitives se font de la machine de Turing consiste à la considérer comme omnisciente. Elle est alors la référence obligée d'une conception symbolique de la cognition. Lorsque l'on veut étudier ou tester un algorithme, il suffit de le faire dans le cadre de la machine de Turing. Elle est considérée comme indépassable. Le modélisateur, quand il veut rendre son modèle opérationnel, le ramène toujours plus ou moins directement à un cadre conceptuel associé à la machine de Turing. *Elle est, selon cette conception, le seul cadre qui permette de rendre opérationnelle une approche symbolique de la cognition.*

Selon nous, une telle conception restreint par trop les capacités effectives des machines symboliques potentielles. La thèse proposée est à l'opposé d'une telle conception. Nous faisons le pari qu'il est toujours possible de concevoir des machines symboliques, de plus en plus compliquées structurellement, qui sont peut-être simulables sur une machine de Turing mais dont le "comportement" observé n'a rien à voir avec le "comportement" simulé qui serait observé sur la machine de Turing².

¹La thèse de Simon consiste à affirmer que les systèmes cognitifs sont construits sur des unités élémentaires qui sont des symboles.

²Par exemple le nombre d'opérations élémentaires en jeu sur la machine de Turing peut être très différent de celui observé sur la nouvelle machine.

Prenons par exemple un “comportement” qui est discernable pour un observateur sur cette nouvelle machine. Ce “comportement” peut cependant être imperceptible lorsqu’il est simulé sur une machine de Turing parce qu’il est trop lent pour que l’observateur réussisse à le distinguer. Il ne suffit pas alors d’étudier la machine de Turing seule quand on veut réaliser des modélisations, il faut la considérer également dans ses relations avec un observateur.

En affirmant la nécessité de la matérialité, nous prétendons que toute machine symbolique sera toujours limitée et insuffisante pour rendre compte de la cognition. Selon nous cette limitation affirmée, en opposition avec l’omnipotence, également affirmée habituellement, de la machine de Turing, peut être un stimulant pour concevoir de nouvelles machines symboliques différentes de celle de Turing, même si elles peuvent éventuellement y être réduites. Ainsi notre thèse permet, selon nous, d’ouvrir l’horizon sur ce que les machines symboliques sont actuellement effectivement capables de faire.

En effet les résultats théoriques actuels concernent plusieurs types de machines dont certaines sont différentes des machines déterministes de Turing, offrant d’autres possibilités. Ces nouveaux résultats ne vont pas à l’encontre de ceux obtenus pour les machines de Turing déterministes, mais ils permettent de proposer de nouvelles classes dans la théorie de la complexité algorithmique. On peut citer, par exemple, les résultats sur les machines de Turing non-déterministes³. Ces machines sont différentes des machines de Turing et on ne sait pas à l’heure actuelle si une réduction⁴ des machines non-déterministes aux machines déterministes est possible.

Un exemple de limitation des théories actuelles est lié au fait que les machines de Turing non-déterministes, elles-mêmes, possèdent une table de changement d’état dont la taille est fixe. Ceci signifie par exemple que le nombre de choix possibles est uniformément borné *a priori*. On peut proposer des machines non-déterministes dont les possibilités de choix augmentent, voire augmentent exponentiellement, lorsque l’on se déplace sur le ruban. À ma connaissance, de telles machines ne sont pas intégrées, à l’heure actuelle, dans les théories de la complexité algorithmique.

Dans ce cadre la thèse selon laquelle la complexité associée à la cognition

³Les machines de Turing non-déterministes possèdent, non pas un, mais plusieurs successeurs dans leur table de changement d’état. Cette table reste elle-même de taille fixe.

⁴On ne sait pas si une simulation d’un problème qui a une certaine complexité sur une machine non-déterministe (on note **NTIME()** cette classe de complexité) restera dans la même classe de complexité sur une machine de Turing universelle (on note **DTIME()** cette classe de complexité). Les chiffres dans la parenthèse de **NTIME()** et **DTIME()** donnent la “durée” du fonctionnement de la machine avant la décision, en fonction de la taille des données. “Signalons [...] que la question de l’égalité des classes **DTIME(cn)** et **NTIME(cn)** est résolue négativement [Paul *et al.*, 1983]. On connaît même un certain nombre de problèmes concrets qui sont dans la classe **NTIME(n log n)** (qui est en un certain sens la version la plus naturelle du temps non déterministe linéaire) et dans aucune classe **DTIME(dn)** [Grandjean, 1990]”[Dehornoy, 1993][p. 124].

nécessite la matérialité peut être un stimulant pour imaginer de nouvelles⁵ machines symboliques dont les caractéristiques ne sont pas intégrées dans les machines symboliques actuelles et qui amèneront l'introduction de nouvelles classes dans la hiérarchie de la complexité algorithmique. On pourra ainsi constituer de nouvelles machines symboliques qui “dépassent” les précédentes, dans le sens où elles permettent de définir de nouvelles classes de complexité, ou dont la complexité linéaire sera, par exemple, une complexité “exponentielle” pour les machines précédentes. Une optique progressiste consiste alors à penser que l'on peut toujours définir de telles nouvelles machines symboliques qui dépassent les machines symboliques précédentes. À la limite, elles convergeraient alors asymptotiquement vers des machines dont la complexité serait aussi proche que l'on veut de celle obtenue dans une conception symbolique de la cognition.

2.2. Une autre définition de la cognition

Cette thèse permet également de proposer une définition large de la cognition. Dans ce contexte, on peut définir **la cognition** comme *une partie des phénomènes qui mettent en jeu un ensemble de cellules matérielles*. Dans cette optique, toute une partie du travail à venir consiste à expliciter, au sein de la matérialité, quels sont les phénomènes qui relèvent de la cognition et ce que sont les autres phénomènes qui mettent en jeu les cellules matérielles.

2.3. Le gouffre de la réduction quantique

Cette définition de la cognition peut sembler beaucoup trop large. Son premier intérêt cependant est de constituer un arrêt brutal à la tentation du “tout cognitif” qui se développe rapidement à l'heure actuelle.

Il faut noter en particulier la multiplication des travaux très sérieux, principalement dans le monde anglo-saxon, dont l'objectif principal est de réaliser une description complète du monde, et en particulier de la conscience, à partir des descriptions, des formalismes et des concepts de la mécanique quantique. L'approche générale consiste à affirmer que la matière est quantique et se décrit donc selon le formalisme de la mécanique quantique. Cette position est généralement partagée par les physiciens. Le second point du raisonnement consiste à rappeler que l'objectif, quand on étudie les capacités cérébrales, est de le faire sur une base matérielle. Puisqu'*in fine* le cerveau n'est que matière, ses capacités doivent être expliquées sur une base matérielle, c'est-à-dire quantique. Dans cette optique, on peut noter, par

⁵ “La fécondité est l'instance devant laquelle tout refus au nom de l'évidence s'avère préjugé : les mathématiques réelles initiales ne sont plus qu'un cas particulier situé au sein des mathématiques nouvelles, expliqué par elles” [Cavaillès, 1938a][p. 46].

exemple, les travaux de Stapp et son ouvrage synthétique [Stapp, 1993], qui sont une tentative de réduction du fonctionnement cérébral dans son ensemble à une activité expliquée uniquement dans les termes de la mécanique quantique.

Notre thèse, dans ce contexte, en définissant la cognition à partir des cellules, instaure une forme d'irréductibilité méthodologique. Ceci n'exclut pas de tenir compte de phénomènes, ou de formalismes associés à la mécanique quantique, qui peuvent intervenir dans le cadre des phénomènes cellulaires. Cependant, en procédant ainsi, on peut garder clairement à l'esprit le fait que les explications doivent être fournies à partir des phénomènes associés aux cellules, et non de phénomènes qui permettraient d'expliquer le fonctionnement de la cellule elle-même.

2.4. Éviter une conception réifiante de nos abstractions sous forme de processus

Il faut envisager la thèse précédente comme une modification de notre conception des processus cognitifs. Désormais la notion de processus introduite dans le chapitre 2 doit bien être comprise comme *l'abstraction d'un phénomène dynamique*, et non simplement comme une modification d'objets statiques. En particulier, associer un symbole ou un ensemble de symboles à un processus ne nous semble pas adéquat et ne peut constituer qu'un dernier recours et une simple contrainte, une nécessité associée à nos outils mathématiques actuels. Il n'est donc pas question ici de réduire les processus à des modifications d'éléments considérés comme statiques (des objets ou des symboles). Désormais, si l'on souhaite “faire émerger” ou “réduire” ces processus, on doit chercher à le faire à partir d'autres processus. *Il faut alors concevoir les processus cognitifs comme construits sur d'autres processus.*

Nous avons déjà dit dans le chapitre 2 que les abstractions peuvent être réalisées par l'observateur soit comme des objets, soit comme des processus. La première conception cherche à réifier les phénomènes et la seconde à leur donner un statut dynamique. Il s'agit d'une position méthodologique. Si on veut suivre la thèse exprimée précédemment, il semble important méthodologiquement d'abstraire tous les phénomènes en jeu comme des processus. En effet, *une attitude d'abstraction qui serait réifiante pourrait facilement être transposée en termes symboliques*. On pourrait alors lui associer une conception de la complexité correspondant à un niveau de la hiérarchie de la complexité algorithmique (symbolique), éventuellement à venir, dans le sens que nous avons proposé dans le paragraphe précédent, *Ouvrir l'horizon et concevoir de nouvelles machines*.

Défendre la thèse que nous avons énoncée permet de s'attacher, quoi qu'il arrive, à des phénomènes (et non à des symboles) et donc à la nécessité de les abstraire de nouveau, à tout moment, sous forme de processus. *Cette thèse, en s'attachant aux phénomènes plutôt qu'aux conceptions abstraites que l'on peut s'en*

faire, permet d'éviter l'élaboration d'un monde abstrait (éventuellement symbolique) qui prendrait progressivement son autonomie. Cette thèse permet donc d'éviter toute dérive qui consisterait à ne s'attacher qu'aux processus et à oublier les phénomènes qui nous ont permis de les construire comme abstractions.

2.5. L'infini en acte

La citation suivante de Jean Cavaillès sur l'infini va nous permettre d'interpréter un certain nombre d'aspects que l'on retrouve pour les phénomènes mettant en jeu des cellules : “d'où la possibilité qu'un ensemble soit infini en acte, c'est-à-dire que nous ne puissions l'embrasser du regard et même que nous ne sachions l'engendrer”[Cavaillès, 1938b][p. 46]. Nous dirons qu'**un phénomène est infini en acte s'il est inaccessible pour les éléments qui le constituent**.

Nous allons montrer, dans un cadre “quasi-formel”, que cette conception de l'infini est opérationnelle. Lorsque nous transposerons ce concept dans un cadre matériel, on ne pourra donc pas considérer qu'il relève de la métaphysique, ou qu'il contribue à développer une conception vitaliste.

Qu'est-ce qu'un infini en acte, et donc, en un certain sens, un fini ? Proposons brièvement⁶ une direction d'interprétation. Considérons les choses d'un point de vue purement formel, au sens de Gödel, c'est-à-dire un sous la forme d'un système “matériel” abstrait, sur le même principe que la construction de la machine de Turing.

Considérons par exemple un ensemble discret et ordonné de symboles comme l'ensemble des entiers relatifs. On suppose que chaque élément de cet ensemble est inscrit, dans l'ordre, sur une case C d'un ruban, avec un seul chiffre par case. Nous appelons variable V_1 une fenêtre transparente mobile qui se déplace en glissant sur le ruban. On peut attribuer comme valeur à cette variable celle de la case C qu'elle recouvre. En conséquence la variable ne peut être incrémentée ou décrémentée que d'une unité à la fois, puisque lors de son déplacement le carré transparent doit passer sur l'une des cases adjacentes. On peut considérer chaque déplacement de la variable comme le bip B_1 d'une horloge. Chaque bip a lieu au moment où la variable et la case sont bien superposés. Au moment du bip B_1 , la variable V_1 a donc la valeur C_{B_1} de la case C_{B_1} qu'elle recouvre. De plus la variable V_1 a la valeur C_{B_1} tant qu'elle n'a pas une nouvelle valeur.

Supposons désormais que, pour la variable V_1 , le bip suivant est fixé, mais pas la case sur laquelle elle va se trouver au moment du bip. La variable peut “décider”

⁶Il ne nous a pas été possible de développer cet aspect de manière complètement formelle dans le temps dont nous disposions pour réaliser cette thèse.

de se retrouver à ce moment soit sur la case de gauche soit sur celle de droite. Nous évacuons toute difficulté qui serait associée à la manière dont serait réalisé ce choix.

On peut maintenant envisager un ensemble de variables qui se déplacent sur les entiers. Chaque variable définit sa propre horloge. On dira que **deux variables se rencontrent** si leurs fenêtres transparentes mobiles sont partiellement superposées. On dira que l'horloge d'une variable V_2 est **plus lente** que celle d'une autre variable V_1 quand la condition suivante est vérifiée : *si les deux variables se rencontrent, alors la variable V_1 peut suivre la variable V_2 .* Cette condition signifie que les changements de position de V_2 sur un nouvel entier peuvent être suivis par un choix de V_1 tel que, suite à son changement de position, les deux variables se rencontrent à nouveau.

Les horloges de deux variables V_1 et V_2 peuvent être **commensurables** entre elles. Nous dirons que deux horloges sont commensurables si l'horloge de la variable V_1 est plus lente que l'horloge de la variable V_2 et si l'horloge de la variable V_2 est plus lente que l'horloge de la variable V_1 .

On peut alors définir un **phénomène pour une variable** comme un élément distinguable pour la variable et auquel elle peut associer une horloge qui lui serait propre. Un phénomène pour la variable V_1 est donc par exemple la donnée sur le même ruban d'une autre variable V_2 avec sa propre horloge⁷. On peut alors chercher à définir la notion de phénomène accessible pour la variable V_1 . Qu'est-ce qui va être accessible pour cette variable V_1 ? Tout "phénomène" dans l'ensemble des entiers dont l'horloge sera plus lente que celle de la variable V_1 . En effet pour tout phénomène de ce type, la variable pourra suivre le phénomène. Au contraire, tout phénomène qui se déroulera sur les entiers, mais comme une exponentielle par rapport à son horloge, restera inaccessible à la variable⁸. Ce phénomène est inaccessible à la variable, même si le phénomène pour la variable et la variable peuvent se rencontrer. En effet la variable, après avoir rencontré le phénomène, ne pourra pas le suivre. Ce phénomène reste fini, mais il est un infini en acte pour la variable.

On peut également imaginer un passage de paramètres de certaines variables⁹ à d'autres variables. Pour une population de variables, il serait alors possible de suivre des phénomènes qu'une seule variable n'aurait pas pu suivre. Pourtant, même pour la population de variables, le phénomène qui se déroulerait comme une exponentielle

⁷À supposer qu'une variable V_1 sait distinguer une autre variable V_2 et qu'elle sait attribuer à V_2 une horloge qui serait bien quelque chose qui a à voir avec l'horloge de V_2 .

⁸En fait il est inutile ici d'avoir au départ un ensemble infini comme l'ensemble des entiers. Considérons un ensemble d'entiers fini qui autoconstruit le successeur de son plus grand élément et le prédécesseur de son plus petit élément. Il suffit alors que cette autoconstruction soit plus rapide que tout déplacement possible d'une variable. Pour la variable, quoi qu'il arrive, l'ensemble sera alors comme infini, mais restera en fait toujours fini.

⁹La variable peut ne pas se réduire à la fenêtre transparente mobile et comporter également une sorte de panier qui pourrait contenir un élément qui serait le paramètre à transmettre.

de l'horloge de référence de la population¹⁰ serait un phénomène fini, mais serait un infini en acte pour la population de variables.

Si on suppose de plus que ce phénomène fini mais infini en acte est constitué à partir d'un ensemble de variables, alors on peut parler de temporalité propre de ce phénomène, puisqu'il est infini en acte pour les éléments qui l'ont constitué¹¹.

Dans la suite, on pourra rapporter la notion d'observateur à celle des variables que nous venons de considérer.

2.6. Le rôle de l'observateur

Dans un premier temps, nous allons suivre un raisonnement en terme de zoom sur les possibilités que les éléments matériels peuvent avoir d'observer leur monde. Ce que nous définissons ici est une hypothèse, mais tellement intuitive qu'elle nécessite d'être précisée dans une discussion sur le rôle de l'observateur.

Nous appellerons **observateur pour le phénomène matériel** P_1 , constitué des éléments matériels E_1 , tout élément matériel constitué d'un ensemble d'éléments matériels E_1 ou d'observateurs des éléments matériels E_1 . Donnons deux exemples, tout d'abord un exemple d'observateur, puis un exemple de non observateur. L'homme, par exemple, est un observateur pour les phénomènes matériels moléculaires. En effet les cellules sont constituées à partir de molécules, donc les cellules sont des observateurs pour les phénomènes moléculaires. L'homme est lui-même constitué de cellules, donc il est constitué d'observateurs pour les phénomènes moléculaires. *L'homme est donc un observateur pour les phénomènes moléculaires.*

Par contre la cellule n'est pas un observateur pour les phénomènes matériels animaux. La cellule n'est pas constituée d'un ensemble d'éléments matériels animaux ni d'observateurs des éléments matériels animaux. D'après notre définition de la notion d'observateur, *la cellule n'est donc pas un observateur pour les animaux*. Les animaux sont constitués de cellules, mais la cellule n'est pas constituée d'éléments matériels cellules, ni d'observateurs pour les cellules.

L'intérêt de cette distinction est de permettre de mettre en évidence le fait que la notion d'observateur n'a rien d'absolu. *Un phénomène matériel n'est pas observable ou non observable par lui-même, il est observable ou non-observable pour un certain type d'observateurs.* Nous allons prendre en particulier le cas de l'observation

¹⁰Nous n'avons pas défini ce qu'était une horloge de référence pour une population de variables et comment éventuellement chaque élément de la population a ou non accès à cette horloge.

¹¹On peut avoir comme image triviale le fait qu'une cellule est constituée de molécules qui vont être l'équivalent de nos variables et en même temps les constituants de la cellule. La cellule étant l'équivalent de notre phénomène fini et infini en acte. La cellule a alors une temporalité propre dans le sens où elle est inaccessible aux molécules qui la constituent.

de phénomènes matériels que les hommes parviennent à abstraire sous forme de symboles.

On appellera **observateur de type humain**, *tout observateur d'un phénomène susceptible de réaliser une abstraction de phénomènes matériels comme symboles quand un observateur humain ferait de même*. Nous ne préjugeons pas de la nature matérielle des observateurs de type humain¹².

Nous souhaitons distinguer ici les symboles des cellules. Selon nous les symboles ne sont observables que pour des observateurs de type humain, alors que les cellules ont des propriétés intrinsèques qui sont indépendantes de toute forme d'observation. Cette distinction entre symboles et cellules est essentielle et explique toute notre approche.

L'unité de la cellule ne nécessite pas d'observateur pour la cellule, contrairement à l'unité du symbole.

Une cellule est cellule comme élément matériel, car elle se définit par elle-même en se perpétuant. La cellule se définit intrinsèquement comme une unité qui ne nécessite pas d'observateur. Au contraire un symbole matériel ne prend son unité comme symbole que parce qu'il est observé par un individu de type humain extérieur et susceptible de considérer l'élément matériel effectif comme un symbole possédant une unité.

2.7. Temporalité

Nous appellerons **temporalité** *la capacité, en perpétuel renouvellement, pour un élément, de se constituer, de se déployer et de se perpétuer¹³ comme infini en acte pour les unités qui le constituent, et ce sans l'intervention d'aucun observateur pour cet élément.*

La cellule est selon nous un exemple d'élément matériel doté de la capacité de temporalité. En effet la cellule est composée de molécules, et elle constitue un infini en acte par rapport à ces molécules. De plus, la cellule se constitue, se déploie et se perpétue sans l'intervention d'un quelconque observateur pour la cellule, et en particulier sans la présence d'un observateur de type humain. Ceci ne signifie pas pour autant que la cellule a cette capacité indépendamment de toute contrainte extérieure, en particulier sur le milieu dans lequel elle est capable de la mettre en œuvre.

¹²On peut d'ailleurs se demander, à ce propos, comment un observateur de type humain peut observer qu'un autre observateur est lui-même de type humain ?

¹³Les trois verbes réflexifs **se constituer, se déployer et se perpétuer** doivent être compris comme des actions toujours inachevées, qui se déroulent.

Nous avons vu que la cellule est un infini en acte, capable de temporalité. Elle peut donc servir de brique élémentaire. Elle est susceptible d'être observée par un observateur de type humain. De plus elle est constitutive de l'être humain dont nous cherchons à analyser les capacités cognitives. *Il est bien difficile de trouver un autre élément matériel qui ait de telles propriétés.* Selon nous, seuls des éléments matériels quantiques pourraient constituer une sérieuse alternative. **Entre le gouffre de la réduction quantique et le choix de la cellule comme brique matérielle élémentaire, sauf à fermer les yeux, il faudra bien choisir.**

Le choix de la cellule comme élément constituant de base se justifie par sa temporalité, dans le sens que nous venons de définir. En effet il est difficile de trouver d'autres phénomènes que l'on considère comme intrinsèquement dotés d'une temporalité sans en venir à l'homme lui-même, ou à un individu pluricellulaire dans son ensemble. Mais si on avait choisi l'homme dans son ensemble comme constituant de base, on se serait tourné vers les Sciences Sociales, plutôt que vers des explications scientifiques des capacités cognitives individuelles assises sur une base biologique.

2.8. Comparaison entre la thèse de Simon et celle proposée

La thèse que nous proposons est une réponse, une thèse alternative à celle symbolique de Simon et Newell.

Les deux points de vue, celui de Simon, comme celui que nous proposons, sont des points de vue constructifs. Ils ne considèrent cependant pas la même "brique" élémentaire. D'un côté la brique élémentaire est abstraite comme un objet, alors que de l'autre elle constitue une brique avant toute abstraction, et son abstraction doit être faite sous forme de processus.

La distinction à faire entre ces deux conceptions est pourtant importante. Dans l'une la brique est un symbole et nécessite alors *a priori* une abstraction, donc l'existence d'un observateur de type humain capable de la réaliser. *Cette position est nécessairement ontologique, si on veut tout de même considérer que l'observateur n'intervient pas pour distinguer les symboles qui constituent la brique.* Pour que la brique soit le symbole, il faut que le symbole possède une unité. Cette unité est fournie par l'observateur de type humain. Si l'on veut que l'observation de ces symboles ne nécessite pas d'observateur de type humain, alors il faut que ce soit le monde lui-même qui soit de nature symbolique. On aboutit alors à l'hypothèse forte sur la nature de la matière, qui a permis la dynamique des sciences cognitives.

Notre conception ne nécessite pas de prendre position sur ce que sont *tous* les états de la matière. En effet affirmer que la cellule, comme élément matériel, possède son unité indépendamment de tout observateur, consiste seulement à prendre position sur *certaines* états de la matière. Du reste cette conception de la cellule est généralement partagée. C'est pour elle-même, dans le tissu de ses relations avec les

autres cellules et avec les molécules, que la cellule prend son identité comme cellule. De ce point de vue, la cellule est une brique qui ne nécessite pas d'observateur de type humain pour se constituer. L'observateur humain peut abstraire cette brique et la considérer comme un symbole, mais ce n'est pas une nécessité. Il peut aussi l'abstraire sous forme de processus comme nous le souhaitons, si l'on veut tenir compte du fait que la cellule construit sa propre unité.

En dehors de ces aspects, la thèse de Simon, si on la prend au sérieux, amène de nouvelles difficultés. En particulier, elle aboutit inéluctablement à des problèmes d'amorçage (boot-strapp). Une des formes du problème de l'amorçage consiste à savoir à quel niveau de description matérielle il faut "ancrer" les symboles. Le problème de l'ancrage des symboles est donc étroitement lié au problème de l'amorçage. En particulier, ne pas trouver de niveau satisfaisant pour ancrer les symboles impose progressivement de redescendre d'un niveau de description de la matière à un niveau plus élémentaire d'une telle description. De proche en proche, on aboutit ainsi à une nécessaire description en termes microscopiques et élémentaires de la matière et on se retrouve de nouveau confronté au gouffre quantique.

Un des intérêts de notre thèse, en proposant de partir directement de la cellule pour étudier la cognition, consiste à éviter dès l'abord de telles difficultés.

2.9. Machines symboliques et machines processuelles

Pour les machines symboliques, les opérations sont données *a priori*. L'observateur peut prédérouler les opérations. À tout instant il peut regarder la machine et savoir quelles sont les opérations observables. Les machines symboliques sont des machines qui manipulent, qui travaillent sur des éléments déjà préabstraits.

On appelle **machines processuelles** *les machines pour lesquelles les opérations ne sont pas données a priori*. L'observateur ne connaît pas les opérations de la machine et à chaque étape, quand il observe, il est obligé de réaliser de nouveau l'opération d'abstraction. De plus cette opération n'a pas nécessairement comme résultat une forme abstraite connue *a priori*.

On peut donner l'exemple suivant de machine processuelle. Prenons une machine cellulaire, c'est-à-dire un ensemble de cellules matérielles constitué par l'observateur. On fixe alors un protocole, c'est-à-dire que l'on fixe l'environnement possible dans lequel l'ensemble de cellules va se retrouver. On ne connaît pas alors les "opérations" éventuelles que pourra réaliser l'ensemble de cellules. On laisse "tourner", c'est-à-dire qu'on laisse le système évoluer. On réalise après un certain temps une abstraction. On recommence. On obtient alors un ensemble de formes abstraites. On peut ensuite regarder comment il est possible d'"algébriser" les relations entre les produits symboliques de ces abstractions.

Les machines symboliques constituent en fait un cadre extrêmement rigide car elles offrent à chaque étape une possibilité d'observation par un observateur de type humain selon des règles qui sont fixées par avance et antérieures à cette observation. Au contraire, pour les machines processuelles, à chaque étape d'observation il faut réaliser une abstraction afin de définir les opérations qui sont en jeu. Ainsi les machines processuelles peuvent-elles fonctionner indépendamment de la présence, même hypothétique, de tout observateur de type humain. Les machines processuelles laissent une part beaucoup plus importante aux phénomènes en jeu, et limitent le rôle attribué à l'observateur dans la notion de machine.

Dans cette optique on peut alors envisager de ne plus étudier la cognition comme une propriété associée exclusivement à des cerveaux ou à des ensembles de cellules nerveuses. *Certains ensembles de cellules non nerveuses peuvent déjà constituer des exemples de machines processuelles.* Il n'est alors plus nécessaire de travailler sur des organismes évolués pour aborder l'étude de la cognition. Ainsi, des organismes pluricellulaires peuvent déjà constituer des machines processuelles permettant une première analyse de la cognition dans le contexte que nous proposons.

On peut aussi envisager d'étudier des machines processuelles comportant des cellules nerveuses et des cellules non nerveuses. Dans ce cadre on peut alors chercher à établir des relations entre le fonctionnement des cellules nerveuses et celui des cellules non nerveuses. On peut également chercher à analyser les relations entre le fonctionnement des différentes cellules et celui de l'organisme dans son ensemble. Dans ce cadre, certaines propriétés des cellules nerveuses qui leur sont spécifiques pourront sans doute être de nouveau mises en évidence. “Les cellules nerveuses partagent plusieurs attributs avec les autres cellules du corps mais elles s'en distinguent au moins sur un point : *seules les cellules nerveuses ont la capacité de produire et de transmettre des signaux à distance.* Au cours de l'évolution des cellules, l'émergence de cette propriété revêt une grande importance pour la compréhension des origines biologiques du système nerveux” [Rosenzweig et Leiman, 1991][p. 50].

2.10. Étudier la cognition sous une forme simplifiée qui préserve sa complexité

Les modélisateurs conviennent que les modèles proposés actuellement sont d'une extrême simplicité, eu égard à la complexité du fonctionnement effectif des systèmes nerveux. Les modélisateurs ont unanimement fait le choix actuellement de choisir comme simplification une description très fruste des cellules nerveuses individuelles. Notre proposition constitue une alternative à de telles simplifications. *En effet, nous proposons une simplification de la cognition tout à fait différente.* On choisit habituellement d'étudier la cognition en analysant une petite partie d'un gros système évolué et très compliqué, dont on remplace chaque élément par un élément simplifié. Nous proposons d'étudier des organismes pluricellulaires matériels

“simples” que l’on essaye d’étudier dans leur ensemble, dans leur enchevêtrement et leur complexité. Cette simplification de la complexité permet selon nous de ne pas discréder la modélisation comme le fait actuellement une attitude consistant à proposer des modèles homogènes, qui réduisent cette complexité à la complexité/quantité.

Les symboles comme symboles, depuis les travaux de Turing, sont définis par leur matérialisation. Ils nécessitent tout de même l’existence d’un observateur de type humain pour les observer. Les processus comme processus doivent être également définis par leur matérialisation, comme nous le proposons par l’intermédiaire des cellules. Les processus matérialisés par l’intermédiaire de la cellule ne nécessitent pas, quant à eux, l’existence d’un observateur de type humain pour se constituer. On peut noter cependant que, contrairement aux possibilités de la machine de Turing pour les symboles, les cellules ne peuvent sans doute pas permettre une matérialisation de tous les processus.

2.11. La méthodologie de la mise en parangon

Dans notre optique, l’objectif, lorsque l’on souhaite étudier la cognition, est de chercher à réaliser une modélisation fine de phénomènes mettant en jeu des ensembles de cellules. La modélisation, comme nous l’avons vu, nécessitait quatre domaines, ceux de la théorie, de l’expérience, du modèle formel et de la simulation informatique. Le point de vue que nous défendons nécessite d’ajouter un nouveau domaine, **la reconstruction du phénomène étudié avec le même matériau**. Nous appelons **parangon** *cette construction matérielle avec le matériau du phénomène étudié, et à valeur exemplaire*. La mise en parangon met en jeu une construction matérielle correspondant au modèle et dont les éléments constitutifs sont des cellules dans le cas des modèles de la cognition, c’est-à-dire possèdent la matérialité même que nous cherchons à étudier.

Au cœur de la modélisation on trouvait le modèle, c’est-à-dire un système formel. Progressivement la simulation se trouvait, à la place du modèle, au centre de la modélisation. **Avec la mise en parangon, c'est la reconstruction matérielle sur le même substrat (c'est-à-dire avec des cellules dans le cas de la cognition) qui constitue le cœur du travail**, . En ce sens, la mise en parangon est extrêmement différente de la modélisation. Le modèle reste un domaine important, mais il n’est plus le centre sur lequel toute l’attention se focalise.

La matérialité, telle que nous venons de la considérer, n’a donc plus rien à voir avec les théories de la “rematérialisation”.

Il est alors nécessaire de développer des outils symboliques qui vont permettre de rendre compte des différents aspects de la complexité que nous avons mis en avant. Ils doivent en particulier étayer la thèse que nous cherchons à défendre ici.

D'un point de vue purement méthodologique, nous avons déjà présenté comment l'étude des machines processuelles doit nous permettre de constituer des "algèbres de processus adéquates".

Nous en resterons là, mais il faut avoir conscience que les difficultés à surmonter pour créer des outils formels adéquats sont certainement liées à la façon dont le temps sera pris en compte.

2.12. Le langage du cerveau

Von Neumann a une conception très large des mathématiques, qui est loin de se réduire à ce que les théories mathématiques peuvent comporter en un temps historique particulier. Pour lui, le langage des mathématiques est tout langage constitué de symboles, c'est-à-dire tout ensemble de relations abstraites entre éléments abstraits. C'est le sens qu'a repris Simon quand il a affirmé la thèse forte selon laquelle la cognition¹⁴ pouvait être identifiée à un langage entre symboles. En ce sens, Simon a perpétué une interprétation de l'héritage de von Neumann.

Le langage considéré dans notre thèse est constitué à partir d'éléments matériels, les cellules. Il n'est pas ici un ensemble de relations entre des éléments qui nécessitent un observateur de type humain pour être définis comme des éléments possédant une unité. Le langage n'est donc pas ici un ensemble de relations entre éléments qui possèdent une unité abstraite, mais un ensemble de relations entre éléments matériels qui définissent leur propre unité. Le langage est ici non pas une abstraction, mais un phénomène matériel particulier, l'ensemble des relations matérielles non abstraites entre cellules.

Simon parle également de matérialité. Cependant, la matérialité qu'il considère n'a rien à voir avec celle que nous envisageons. Elle est toujours soit imprégnée de la présence d'un observateur soit associée à une thèse qui nécessite la confrontation de deux mondes *a priori* distincts et identifiés *a posteriori*, celui des symboles considérés comme tels par des observateurs de type humain et celui des phénomènes.

On peut donc considérer notre thèse comme une tentative d'interprétation et de réhabilitation, différente de celle de Simon, des derniers mots de von Neumann : "il faut étudier le langage du cerveau et non le langage des mathématiques".

¹⁴La position de Herbert Simon va sans doute plus loin en considérant que le monde réel dans son ensemble peut être identifié à un monde de symboles.

RÉFÉRENCES BIBLIOGRAPHIQUES DU CHAPITRE 5

- [Buser et Imbert, 1993] BUSER P. ET IMBERT M. *Neurobiologie I : mécanismes fondamentaux et centres nerveux*. Hermann, Paris, 1993. 631 p. Collection Méthodes. ISBN 2-7056-6195-6.
- [Cavaillès, 1938a] CAVAILLÈS J. *Méthode axiomatique et formalisme*. Première édition. Hermann, Paris, 1938. Volumes 608, 609 et 610 des Actualités scientifiques et industrielles. Trois tomes, I/ Le problème du fondement des mathématiques (83 p.) II/ Axiomatique et système formel (58 p.) III/ La non-contradiction de l'arithmétique (73 p.).
- [Cavaillès, 1938b] CAVAILLÈS J. *Remarques sur la formation de la théorie abstraite des ensembles*. Première édition. Hermann, Paris, 1938. Volumes 606 et 607 des Actualités scientifiques et industrielles. Deux tomes, I/ Préhistoire, la création de Cantor (107 p.), II/ Dedekind, les axiomatisations (Zermelo, Fraenkel, von Neumann) (51 p.).
- [Changeux, 1994] CHANGEUX J.-P. *La chimie des communications cérébrales*. Panda Films EDV40. Arlequin Vidéo, 1994. Collection du Collège de France. Durée 80 mn.
- [Churchland et Sejnowski, 1992] CHURCHLAND P. ET SEJNOWSKI T. *The computational brain*. Première édition. MIT Press, Cambridge (Mass.), 1992. 544 p. ISBN 0-262-03188-4.
- [Cunningham et Waxman, 1991] CUNNINGHAM R. ET WAXMAN A. Astroglial-neural networks, diffusion-enhancement bilayers, and spatio-temporal grouping dynamics. Dans *Proceedings of SPIE '91 in Boston, MA*. 1991.
- [Dehornoy, 1993] DEHORNOY P. *Complexité et décidabilité*. Springer, Paris, 1993. 200 p. Volume 12 de la série Mathématiques et applications. ISBN 2-287-00416-5.
- [Farmer, 1990] FARMER J. D. A rosetta stone for connectionism. *Physica D*. 1990, vol. 42, p. 153–187.
- [Grandjean, 1990] GRANDJEAN E. A nontrivial lower bound for an NP problem on automata. *SIAM journal on computing*. 1990, vol. 19, no 3, p. 438–451.

- [Holley et McLeod, 1977] HOLLEY A. ET MCLEOD P. Transduction et codage des informations olfactives chez les vertébrés. *Journal de physiologie de Paris*. 1977, vol. 73, p. 725–828.
- [Hubel, 1988] HUBEL D. *Eye, brain, and vision*. Scientific American Library, Oxford, 1988. 240 p. Volume 22. ISBN 0-7167-5020-1.
- [Kettenman et Ransom, 1988] KETTENMAN H. ET RANSOM B. Electrical coupling between astrocytes and between oligodendrocytes studied in mammalian cell cultures. *Glia*. 1988, vol. 1, p. 64–73.
- [McCulloch, 1951] MCCULLOCH W. S. Why the mind is in the head. *Cerebral mechanisms in behavior - the Hixon symposium, Pasadena, 1948*. 1951, p. 42–111. L. Jeffress, éditeur. J. Wiley, New York. Reproduit dans les Collected works de Warren McCulloch intitulés *Embodiments of mind*. p. 72-141. MIT Press, Cambridge (Mass.), 1965.
- [Mézard et Toulouse, 1991] MÉZARD M. ET TOULOUSE G. Des verres de spin aux réseaux de neurones. *La recherche*. 1991, vol. 22, no 232, p. 616–623.
- [O.F.T.A., 1991] O.F.T.A. *Les réseaux de neurones*. Masson, Paris, 1991. 212 p. Volume 11 de la collection Arago. ISBN 2-225-82439-8. ISSN 0985-7877.
- [Paul *et al.*, 1983] PAUL W., PIPPENGER N., SZEMEREDI E. ET TROTTER W. On determinism versus nondeterminism and related problems. Dans *24th IEEE Symposium on foundations of computer science, Washington*, p. 429–438. 1983.
- [Popper, 1978] POPPER K. *La logique de la découverte scientifique*. Reproduction de la première édition française parue en 1973. Payot, Paris, 1978. 484 p. Collection Bibliothèque scientifique. Traduit de *The logic of scientific discovery*. Première édition anglaise Hutchinson, 1959. Première édition allemande 1934. ISBN 2-228-11391-3.
- [Prez, 1997] PREZ P. *De la vision à l'acte de voir : un parcours technique et épistémologique dans les sciences cognitives*. Thèse de doctorat, université de Paris-sud, Orsay, 1997.
- [Quartz et Sejnowski, 1996] QUARTZ S. ET SEJNOWSKI T. The neural basis of cognitive development: a constructivist manifesto. Article en ligne sur internet, à paraître dans *BBS*. 1996.
- [Rosenzweig et Leiman, 1991] ROSENZWEIG M. ET LEIMAN A. *Psychophysiologie*. Deuxième édition. InterÉditions, Paris, 1991. 832 p. ISBN 2-7296-0238-0.
- [Saurel, 1992] SAUREL P. *Q-learning et apprentissage par renforcement*. DEA de sciences cognitives, EHESS, Paris, 1992.

- [Sereno, 1988] SERENO M. The visual system. Dans *Organization of neural networks: structures and models*, von Seelen W., Shaw G. et Leinhos U., éditeurs, p. 167–184. VCH, 1988.
- [Stapp, 1993] STAPP H. *Mind, matter, and quantum mechanics*. Springer, Berlin, 1993. 248 p. ISBN 3-540-56289-3.
- [Vincent, 1988] VINCENT J.-D. *Biologie des passions*. O. Jacob, Paris, 1988. 345 p. ISBN 2-02-009-040-6.
- [von Neumann, 1951] VON NEUMANN J. The general and logical theory of automata. *Cerebral mechanisms in behavior - the Hixon symposium, Pasadena, 1948*. 1951, p. 1–31. L. Jeffress, éditeur. J. Wiley, New York. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 288-328. Pergamon Press, Oxford, 1961-63.
- [von Neumann, 1956] VON NEUMANN J. Probabilistic logics and the synthesis of reliable organisms from unreliable components. Dans *Automata studies*, Shannon C. et McCarthy J., éditeurs, p. 43–98. Princeton University Press, Princeton (New Jersey), 1956. Reproduit dans les Collected works de John von Neumann. A. Taub, éditeur. Volume V, Design of computers, theory of automata and numerical analysis, p. 329-378. Pergamon Press, Oxford, 1961-63.
- [von Neumann, 1966] VON NEUMANN J. *Theory of self-reproducing automata*. A. Burks, éditeur. University of Illinois Press, Urbana et Londres, 1966.
- [von Neumann, 1992] VON NEUMANN J. *L'ordinateur et le cerveau*. Première édition française. La découverte, Paris, 1992. 130 p. Collection Textes à l'appui/série sciences cognitives. Traduction de The computer and the brain - the Silliman lectures. Première édition américaine, Yale University Press, New Haven, 1958. ISBN 2-7071-2164-9.
- [von Neumann et Birkhoff, 1936] VON NEUMANN J. ET BIRKHOFF G. The logic of quantum mechanics. *Annals of mathematics*. 1936, vol. 37, p. 823–843. Reproduit dans les Collected works de John von Neumann. Taub A., éditeur. Volume IV, Continuous geometry and other topics, p. 105-125. Pergamon Press, Oxford, 1961-63.
- [Wiener et Rosenblueth, 1945] WIENER N. ET ROSENBLUETH A. The role of models in science. *Philosophy of science*. 1945, vol. 12, p. 316–322.

Table des matières

Sommaire	3
Dédicace	5
Remerciements	7
Introduction	9
Références bibliographiques de l'introduction	27
Chapitre 1 L'évolution de la notion d'axiomatisation et ses relations avec l'étude du cerveau 1903-1996	29
1 Quarante ans d'évolution de la notion d'axiomatisation à travers le parcours de von Neumann (1903-1943)	31
1.1 De l'axiomatisation à la formalisation : von Neumann face à quatre problèmes de Hilbert	31
1.1.1 Le programme de Hilbert	31
1.1.2 L'axiomatisation selon Hilbert	32
1.1.3 L'axiomatisation de la physique	33
1.1.4 Axiomatiser la théorie des ensembles ?	34
1.1.5 Axiomatiser l'arithmétique : cheminement par la logique ?	37
1.1.6 Les mathématiques du fini pour elles-mêmes : mécaniser les calculs	38
1.1.7 L'axiomatisation selon von Neumann	39
1.2 Obtenir une définition opératoire de la calculabilité : von Neumann et la machine de Turing	41
1.2.1 La machine de Turing, la calculabilité et les systèmes formels	41

1.2.2 Généricité des systèmes formels et machines à réaliser des calculs	43
1.2.3 Le problème du choix des opérations élémentaires	46
1.2.4 Les difficultés liées à la matérialisation des opérations élémentaires	47
1.2.5 Pourquoi privilégier la construction de machines digitales	48
1.2.6 La trajectoire de von Neumann dans le sillage du programme de Hilbert	48
2 Les années quarante et l'unification de la conception du formalisme (1943-1956)	49
2.1 Les analogies fondatrices	49
2.1.1 La causalité circulaire et les mécanismes de rétroaction (feedback) comme concepts unificateurs	49
2.1.2 L'analogie entre machines et organismes vivants	51
2.1.3 Les hésitations de von Neumann	52
2.1.4 Les outils mathématiques et la méthodologie de la physique comme point commun	53
2.2 L'analogie comme méthodologie scientifique	55
2.2.1 L'approche méthodologique du “mathématicien”	55
2.2.2 L'approche méthodologique du “premier ingénieur”	57
2.2.3 L'approche méthodologique du “second ingénieur”	57
2.3 De l'analogie à l'identification	58
2.3.1 La réduction du modèle abstrait à sa réalisation matérielle physique	58
2.3.2 La réduction du phénomène étudié au modèle abstrait ..	59
2.3.3 La réduction du phénomène étudié à sa reconstruction matérielle sur un ordinateur	59
2.3.4 La triple réduction comme possible position ontologique au-delà des nécessités méthodologiques	60
2.3.5 La prise en main par McCulloch ou la triple réduction comme position ontologique qui s'impose	61
2.4 Ontologie et conception de l'intelligence chez Turing	62

2.4.1	La réfutation par Turing de cinq arguments opposés à la possibilité de construire des machines montrant un comportement intelligent	62
2.4.2	Comment tester, selon Turing, l'intelligence du comportement des machines ?	65
2.4.3	Interpréter dogmatiquement les résultats de Turing et des cybernéticiens	67
2.5	Les remises en cause de la méthodologie formaliste classique et les premières tentatives pour la dépasser	68
2.5.1	Le rôle des modèles selon Wiener et les premières difficultés rencontrées pour modéliser le cerveau/esprit	68
2.5.2	Les prémisses de la mise en parangon	69
2.5.3	La première voie compatible avec les résultats de von Neumann : le nécessaire retour à l'observation et à l'expérimentation neuroanatomiques	72
2.5.4	La deuxième voie compatible avec les résultats de von Neumann : le développement de logiques alternatives	73
3	L'intelligence artificielle et les sciences cognitives, quarante ans de confrontation à une ontologie formaliste (1956-1996)	75
3.1	Les héritiers de von Neumann, l'intelligence artificielle et les sciences cognitives	75
3.2	Les limites de l'intelligence artificielle et l'offensive anticomputationnelle selon Searle	77
3.3	En résumé	79
	Références bibliographiques du chapitre 1	81
	Chapitre 2 Comment les sciences cognitives conçoivent l'étude des processus	87
	Introduction du chapitre 2	89
1	Diversité et unité des sciences cognitives	91
1.1	La diversité des disciplines concernées et l'organisation autour de thèmes	91
1.1.1	La diversité des disciplines concernées	91
1.1.2	L'organisation autour de thèmes	91

1.2 L'étude de processus, élément principal de l'unité des sciences cognitives	93
1.2.1 La cognition comme ensemble de processus	93
1.2.2 Les sciences cognitives, les processus et les composants ..	96
§1 Pour les sciences cognitives, les phénomènes dynamiques dont on abstrait les processus se constituent sur des phénomènes statiques, que l'on abstrait comme des objets	96
§2 Pour les sciences cognitives, l'étude des processus prime celle des objets et en particulier celle des composants	97
§3 Une autre définition de la notion de tournant cognitif	98
1.2.3 Les processus : abstraction, généralisation, rematérialisation	98
§1 Des expériences particulières aux processus : l'abstraction des processus	99
§2 Les processus comme abstractions permettent la généralisation de leur domaine d'application	103
§3 L'ordinateur permet la rematérialisation des processus	104
1.2.4 Les sciences cognitives doivent-elles se résumer à une méthodologie ?	105
2 Le connexionnisme, un cadre général pour modéliser les processus “émergents”	107
2.1 L'article de Farmer, un cadre général pour la formalisation connexionniste	109
2.1.1 Le connexionnisme selon Farmer	109
2.1.2 Un cadre mathématique général pour les modèles connexionnistes	111
§1 Les unités : l'ensemble des sites	111
§2 Le graphe : l'ensemble des liens	111
§2 Les dynamiques	113
§4 Relation entre les systèmes dynamiques classiques et les systèmes connexionnistes	114
2.1.3 Exemples de modèles connexionnistes	117
§1 Les réseaux de neurones formels	118

§2	Les systèmes de classificateurs	120
§3	Les modèles formels du système immunitaire en termes de réseau	122
§4	Autres exemples	125
2.2	Intérêt et limites du connexionnisme pour modéliser les processus	126
2.2.1	La philosophie connexionniste de la modélisation des processus	126
2.2.2	Ce que le connexionnisme ne peut pas faire	128
2.2.3	Connexionnisme et processus cognitifs	129
3	En guise de conclusion sur la façon d'étudier les processus cognitifs	131
	Références bibliographiques du chapitre 2	137

Chapitre 3 La complexité : des théories de la quantité ? 141

1	Les modèles de la complexité/quantité, cas particuliers du cadre de Farmer	143
1.1	Les modèles de type Ising, exemple de complexité/quantité ...	145
1.1.1	Le phénomène étudié et les observations associées	146
§1	La transition ferromagnétique	146
§2	Les caractéristiques des transitions : lois puissances, exposants critiques et équations universelles	148
1.1.2	Les modèles de type Ising et les méthodes d'étude exactes	149
§1	Introduction aux modèles de type Ising	149
§2	Les modèles de type Ising, cas particuliers du cadre de Farmer	150
§3	La construction du modèle de Heisenberg	151
§4	Le modèle d'Ising et ses solutions exactes	153
§5	Rapide historique des méthodes exactes de résolution du modèle d'Ising	157

1.1.3	Les méthodes d'étude approchées	158
§1	La méthode d'approximation du champ moyen et son application au modèle de Heisenberg	158
§2	L'autocohérence de l'approximation de champ moyen	159
§3	Comparaison entre l'approximation en champ moyen et la solution exacte en dimension un	160
§4	La méthode d'approximation du "groupe de renormalisation"	160
1.1.4	Dynamiques des modèles de type Ising	161
1.1.5	Premiers commentaires sur les modèles de phénomènes critiques	165
§1	Quelques difficultés et arbitraires dans la constitution des modèles de phénomènes critiques	165
§2	Quelques difficultés dans l'étude des modèles de phénomènes critiques	166
§3	Rôle de l'homogénéité et de la quantité dans les modèles de phénomènes critiques	167
1.2	Les automates cellulaires, modèles de dynamiques et exemples de machines	169
1.2.1	Les origines des travaux sur les automates cellulaires	169
1.2.2	Les automates cellulaires dans le cadre de Farmer	170
1.2.3	Problématiques pour les automates cellulaires	171
1.2.4	Applications des automates	173
1.3	Les hypothèses justifiant l'utilisation des modèles de la complexité/quantité pour rendre compte du fonctionnement cérébral	174
1.3.1	Précisions sur les relations entre la complexité/quantité et le chaos	174
1.3.2	Cinq caractéristiques communes aux modèles de la complexité/quantité et aux conceptions des neurobiologistes sur le fonctionnement cérébral	175
§1	La première caractéristique est le nombre très important d'éléments	175
§2	La deuxième caractéristique commune est une hypothèse générale d'homogénéité	176
§3	La troisième caractéristique commune porte sur les aspects temporels et en particulier la redondance et la générnicité	177
§4	La quatrième caractéristique commune est l'hypothèse de localité	177

§5 La cinquième caractéristique commune affirme que le comportement global du système dépend du comportement local de chaque élément	177
2 Questions et outils pour aborder la complexité/quantité ..	179
2.1 L'approche de Wolfram des automates cellulaires	181
2.1.1 Introduction aux automates	181
2.1.2 Propriétés locales des automates	186
2.1.3 Propriétés globales des automates	187
2.1.4 Extensions et applications	189
2.1.5 Conclusion	190
2.2 Une nouvelle classification des problèmes sur les automates ...	191
2.2.1 Vingt problèmes sur les automates cellulaires	191
2.2.2 Les trois structures des problèmes sur les automates	192
2.2.3 Définition des éléments de l'espace S_3	193
2.2.4 Obtention de la structure de l'espace S_3	193
2.2.5 La structure S_3 et le temps infini	194
2.2.6 Relations et stabilité des relations entre les différentes structures des espaces S_1 , S_2 et S_3	194
2.2.7 Comparaison des structures S_1 , S_2 et S_3 , et des structures mathématiques de référence	195
3 Trois méthodologies de la complexité/quantité ..	197
3.1 Les modèles de gaz sur réseau	197
3.1.1 Introduction aux gaz sur réseau	197
3.1.2 Le modèle fondateur de Frisch, Hasslacher et Pomeau ...	199
3.1.3 Les propriétés des gaz sur réseau	202
3.1.4 Extension de l'applicabilité en physique	203

Problème 7	Comment les différents comportements sont-ils distribués dans l'espace des règles d'un automate cellulaire ?	224
Problème 8	Quelles sont les propriétés d'échelle des automates cellulaires ?	225
Problème 9	Quelle correspondance peut-on établir entre les automates cellulaires et les systèmes continus ?	227
Problème 10	Quelle correspondance peut-on établir entre les automates cellulaires et les systèmes stochastiques ? ...	228
Problème 11	Comment les automates cellulaires sont-ils modifiés par le bruit et d'autres types d'imperfections ?	229
Problème 12	Pour les automates cellulaires à une dimension, la complexité générique du langage rationnel est-elle non décroissante avec le temps ?	229
Problème 13	Quels ensembles limites peuvent produire les automates cellulaires ?	230
Problème 14	Quelles relations peut-on établir entre les caractéristiques computationnelles et statistiques des automates cellulaires ?	231
Problème 15	Dans quelle mesure les suites générées par les automates cellulaires sont-elles aléatoires ?	231
Problème 16	Quelle est la fréquence de la computation universelle et des problèmes d'indécidabilité rencontrés avec les automates cellulaires ?	232
Problème 17	Quelle est la nature des ensembles limites des automates lorsque la taille de ces automates devient infinie ?	233
Problème 18	Quelle est la fréquence de l'irréductibilité computationnelle des automates cellulaires ?	233
Problème 19	Quelle est la fréquence des problèmes sur les automates qui ne peuvent être traités informatiquement ?	234
Problème 20	Quelles descriptions de plus haut niveau peut-on donner à propos de l'information traitée par les automates cellulaires ?	234
Annexe C	La classification de Gutowitz	237
C.1	Pour rendre compte du comportement temporel de la quantité : introduction à l'approximation des automates par des chaînes de Markov	237
C.2	Quelques résultats	238

Références bibliographiques du chapitre 3	241
Chapitre 4 La pratique de la modélisation : la mobilisation de masse en RDA en 1989	249
1 Introduction à la pratique de la modélisation	251
1.1 Une vision classique ou “popperienne” de la justification des théories	253
1.2 Les quatre domaines de la modélisation	254
1.2.1 Les nouveaux domaines de la modélisation	254
1.2.2 Quelques caractéristiques de la modélisation	255
1.2.3 Un cadre d’interprétation	257
1.3 Différentes sophistications actuelles	258
1.3.1 L’analyse du modèle et de la simulation	258
1.3.2 La multiréalisabilité	259
1.3.3 Le glissement de théorie	262
2 La modélisation mise en pratique : la mobilisation de masse en RDA en 1989	265
2.1 Le phénomène étudié et la théorie qui lui est associée	265
2.1.1 Narration des événements	265
§1 Introduction	265
§2 La situation en RDA au début de l’année 1989 et les antécédents historiques	266
§3 La phase révolutionnaire : l’année 1989	269
§§1 L’exil	272
§§2 La loyauté	276
§§3 Le neuf novembre 1989 et la chute du régime communiste	277

2.1.2	Éléments de théorie politique permettant d'expliquer le phénomène	279
2.1.3	Les données collectées.....	280
2.2	Le modèle proposé pour ce phénomène	283
2.2.1	Présentation du modèle en langue naturelle	284
§1	Les habitants	284
§§1	La classification des habitants	285
§§2	Les caractéristiques des habitants	285
§§3	La dynamique d'activation des habitants	286
§2	Les districts	286
§§1	La classification des districts et le cas de Berlin	286
§§2	Les caractéristiques des districts	286
§§3	La dynamique des districts et les facteurs géographiques	287
§3	Le facteur global	287
§§1	Ses composants : l'exil, la répression et la libéralisation du régime	287
§§2	Sa détermination par agrégation	288
§§3	Son effet sur les districts	288
2.2.2	Une version plus formelle du modèle	288
§1	Notation	288
§2	Description des E_i	288
§3	Description de E	289
§4	Éléments dépendants de t	289
§5	Dynamiques	290
§§1	Dynamique de l'activité	290
§§2	Dynamique des réactivités	290

§6 Initialisation du modèle	290
§7 Mesure globale sur le système	290
2.3 La simulation informatique et son interprétation	291
2.3.1 Description de la simulation pour un district	291
2.3.2 Existence des différentes phases	295
2.3.3 Variations des transitions en fonction des paramètres du modèle	297
§1 Estimation des seuils possibles pour les phases	297
§2 Le rôle du nombre des voisins	297
§3 Le rôle du nombre d'itérations	299
2.3.4 Quelques remarques supplémentaires sur les cartes d'activité	300
3 Pour aller plus loin dans la pratique de la modélisation ...	301
3.1 Travaux à réaliser	302
3.1.1 Analyse du modèle	302
3.1.2 Analyse de la simulation	302
3.1.3 “Validation” du modèle	303
3.2 Apports aux sciences cognitives	305
3.2.1 Proximité des modèles de réseaux neuronaux	306
3.2.2 Difficultés pour valider les modèles en sciences cognitives ..	307
Références bibliographiques du chapitre 4	309
Chapitre 5 La complexité associée à la cognition nécessite la matérialité : une thèse pour une complexité au-delà de la complexité/quantité	313
Introduction	315

1 La complexité au-delà de la quantité	319
1.1 La diversité des éléments cérébraux	319
1.1.1 La diversité des cellules nerveuses	321
§1 Le ou les neurones et leur classification	321
§2 Ce que la comparaison entre les animaux nous apprend sur le fonctionnement des cellules nerveuses	323
§3 Ce que le développement individuel nous apprend sur le fonctionnement des cellules nerveuses	326
1.1.2 La diversité des cellules du cerveau et du néocortex	328
1.1.3 La diversité moléculaire	332
1.2 Rejet de l'étude d'une unique structure logique	333
1.2.1 Éléments pour rejeter une modélisation "trop fine" du neurone	333
1.2.2 De la thèse de l'identité logique à l'isolation normative...	335
1.2.3 Un exemple formel des difficultés que présente l'attitude normative	337
1.2.4 Les difficultés de l'attitude normative associées à un échange de quantificateurs.....	339
1.3 Mixité et hétérogénéité des structures logiques	341
1.3.1 Diversité des structures logiques	341
1.3.2 Mixité des structures logiques	341
1.3.3 Structures logiques organisationnelles d'un système et structures logiques des unités qui le constituent	342
1.3.4 Organisations et hiérarchies cérébrales	345
1.3.5 Hiérarchie de structures logiques mixtes : l'hétérogénéité...	345
2 Que nous impose cette nouvelle complexité ?	347
2.1 Ouvrir l'horizon et concevoir de nouvelles machines symboliques	347
2.2 Une autre définition de la cognition	349
2.3 Le gouffre de la réduction quantique	349

2.4 Éviter une conception réifiante de nos abstractions sous forme de processus	350
2.5 L'infini en acte	351
2.6 Le rôle de l'observateur	353
2.7 Temporalité	354
2.8 Comparaison entre la thèse de Simon et celle proposée	355
2.9 Machines symboliques et machines processuelles	356
2.10 Étudier la cognition sous une forme simplifiée qui préserve sa complexité	357
2.11 La méthodologie de la mise en parangon	358
2.12 Le langage du cerveau	359
Références bibliographiques du chapitre 5	361
Table des matières	365
Table des figures	379

Liste des figures

Chapitre 1

FIGURE 1	Les trois analogies et les trois identifications	54
----------	--	----

Chapitre 2

FIGURE 1	Relations entre une expérience, les processus mis en évidence et le thème abordé en sciences cognitives	100
FIGURE 2	Exemple dans le cas de l'apprentissage des relations entre une expérience, les processus mis en évidence et le thème abordé en sciences cognitives	100
FIGURE 3	Un système connexionniste complètement interconnecté et présenté sous ses différentes facettes : un graphe, une matrice et une liste.....	110
FIGURE 4	Un réseau à couches avec une couche cachée présenté sous ses différentes facettes : un graphe, une matrice et une liste	112
FIGURE 5	Différents types de systèmes dynamiques, caractérisés par la nature de l'espace de départ lorsqu'il existe, du temps et de l'espace d'arrivée	115
FIGURE 6	La pierre de Rosette proposée par Farmer, table des correspondances de vocabulaire entre les différents domaines (d'après Farmer)	123

Chapitre 3

FIGURE 1	Différents réseaux utilisés pour les modèles de type Ising	152
FIGURE 2	Le modèle d'Ising en dimension un, représenté sous la forme d'un cercle	154
FIGURE 3	Comparaison des valeurs des exposants et paramètres critiques obtenues avec des modèles différents (figure reprise de Diu)	162
FIGURE 4	Quelques explications sur la numérotation des valeurs des voisinages utilisée par Wolfram	180
FIGURE 5	Construction algébrique d'un automate cellulaire suivant la règle 90 selon l'appellation de Wolfram	182

FIGURE 6	Suite d’itérations de la construction géométrique du comportement d’un automate cellulaire évoluant suivant la règle 90 modulo 2 (d’après Wolfram)	183
FIGURE 7	L’évolution des automates correspondant aux règles 90 et 126 à partir d’une configuration initiale comportant un unique élément	184
FIGURE 8	Les règles de collision pour le modèle HLG sur un réseau triangulaire aux symétries hexagonales.....	200
FIGURE 9	La dynamique du tas de sable en dimension un pour une pente critique égale à un.....	204

Chapitre 4

FIGURE 1	La conception classique de la confrontation des théories aux expériences.....	252
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.....	256
FIGURE 3	La multiréalisabilité des théories et des modèles, ainsi que la multiplicité des mesures globales possibles sur une simulation ..	260
FIGURE 4	La figure de Rosen sur la modélisation et ses différentes interprétations	264
FIGURE 5	La carte de la RDA avec la localisation des districts où se trouvaient les principaux groupes contestataires	268
FIGURE 6	Trois tableaux résumant les données obtenues pour les manifestations, la répression et la levée des sanctions	270
FIGURE 7	Deux courbes et deux tableaux résumant les données obtenues pour l’exil.....	274
FIGURE 8	Présentation schématique des deux niveaux du modèle	282
FIGURE 9	Deux représentations différentes d’une même “carte d’activité”	292
FIGURE 10	Mise en évidence de la “transition entre deux phases”	294
FIGURE 11	Comparaison des “cartes d’activité” obtenues lorsque l’on modifie le nombre d’habitants et le nombre de voisins	296
FIGURE 12	Comparaison des “cartes d’activité” obtenues lorsque l’on modifie le nombre d’itérations	298

Chapitre 5

FIGURE 1	Bestiaire montrant quelques aspects de la diversité des cellules nerveuses et en particulier des formes des arborisations dendritiques	320
FIGURE 2	Une illustration des modifications des cellules nerveuses et en particulier des arborisations dendritiques au cours de l'évolution	324
FIGURE 3	Principaux niveaux d'organisation cérébrale, du cerveau entier aux contacts synaptiques	344