

**THÈSE de DOCTORAT de l'UNIVERSITÉ PARIS VI**

Spécialité :

**SCIENCES PHYSIQUES**

présentée

par **René DOURSAT**

pour obtenir le titre de **DOCTEUR DE L'UNIVERSITÉ PARIS VI**

Sujet de la thèse :

**CONTRIBUTION A L'ÉTUDE DES  
REPRÉSENTATIONS DANS LE SYSTÈME  
NERVEUX ET DANS LES RÉSEAUX DE  
NEURONES FORMELS**

soutenue le 29 mai 1991

devant le jury composé de :

M. M. IMBERT  
M. Ch. von der MALSBURG  
M. Y. FREGNAC  
M. R. AZENCOTT  
M. E. BIENENSTOCK  
M. G. DREYFUS  
M. G. TOULOUSE

Président  
Rapporteur  
Rapporteur  
Examineur  
Examineur  
Examineur  
Examineur



Je dois l'ensemble des idées et des travaux développés dans cette thèse à Monsieur Elie BIENENSTOCK, qui en fut l'initiateur et le directeur tout au long de ces quatre années. Dès notre première rencontre, il a su me faire partager l'enthousiasme, les convictions et les espoirs qu'il retirait de ses recherches, dont j'ai pu par ailleurs apprécier le caractère original. Rapidement, je fus conquis par les idées qu'il défendait, et également impressionné par sa vaste culture. Il est ainsi devenu mon "maître" et ami, impulsant ma recherche, stimulant mon travail, redressant parfois mon courage. C'est à la confiance qu'il a bien voulu m'accorder dès le début et au soutien solide qu'il ne m'a pas ménagé par la suite, c'est aussi à l'intelligence de son jugement et à son humour salubre, que je veux rendre hommage. Qu'il reçoive ici l'expression de ma profonde reconnaissance.

Je remercie très sincèrement Monsieur le Professeur Gérard DREYFUS de m'avoir accueilli au Laboratoire d'Electronique de l'Ecole de Physique et Chimie Industrielles de la Ville de Paris, et d'avoir suivi mes travaux avec un intérêt constant et une grande ouverture d'esprit. Son dynamisme, son attention bienveillante et ses suggestions opportunes ont déterminé concrètement mon ouvrage. J'ai eu également le plaisir de pouvoir travailler dans le climat exceptionnel de confiance et de libéralité qu'il a su instaurer dans son laboratoire.

Je suis extrêmement heureux que Monsieur Christoph von der MALSBURG, Professeur à la Ruhr-Universität de Bochum, m'ait fait l'honneur de marquer son intérêt pour ce travail, en acceptant de l'examiner. Sa réflexion novatrice dans le domaine des neurosciences et des sciences cognitives a constitué la source passionnante de ma recherche. Comme j'aurai également la joie de pouvoir travailler avec lui dès l'année prochaine, je veux ici lui exprimer toute ma gratitude pour le soutien qu'il a été, et qu'il sera, dans ma carrière de chercheur débutant.

Monsieur le Professeur Michel IMBERT me fait le grand honneur de présider ma thèse, et je l'en remercie vivement. J'adresse également mes chaleureux remerciements à Monsieur Yves FREGNAC, Directeur de Recherches au CNRS, qui m'a fait le plaisir et l'amitié d'examiner cette thèse. Je suis particulièrement sensible à l'attention que ces deux neurobiologistes ont portée à ce travail de modélisation.

Je suis très reconnaissant à Monsieur Gérard TOULOUSE, Directeur de Recherches au CNRS, d'avoir accepté d'être membre du jury. Je sais l'intérêt particulier qu'il porte aux théories qui ont motivé cette thèse, à travers les judicieuses remarques dont il a pu nous faire part. Je tiens également à remercier Monsieur le Professeur Robert AZENCOTT d'avoir bien voulu participer à ce jury. Leur présence à tous deux m'honore beaucoup.

Un stage que j'ai eu l'occasion d'effectuer au Département de Mathématiques Appliquées de l'Université Brown, à Providence, m'a permis de connaître Monsieur le Professeur Stuart GEMAN, dont j'ai pu apprécier les qualités scientifiques et humaines. Il fut le promoteur d'une collaboration qui a conduit à l'étude décrite dans la première partie de cette thèse. Je lui exprime toute ma gratitude pour cet apport enrichissant.

Je garderai toujours un excellent souvenir du Laboratoire d'Electronique de l'ESPCI, où j'ai passé ces années de thèse. Les échanges que j'ai pu avoir avec les chercheurs et les étudiants du laboratoire ont beaucoup contribué à compléter ma perspective dans ce domaine très riche. Je n'oublierai pas non plus l'atmosphère quotidienne extrêmement amicale du laboratoire, où les discussions scientifiques n'ont jamais exclu un certain art de vivre.

J'adresse en particulier une pensée nostalgique et reconnaissante à Isabelle GUYON, qui fut la première personne à me parler de réseaux de neurones. Je remercie également François ANSELME, dont la gestion dynamique et intelligente du réseau informatique a été pour tous un très grand soutien. L'aide de Jérôme LECAT m'a été par ailleurs très précieuse pour mener à bien certaines simulations dans les dernières étapes de ce travail.

Mon oncle d'Amérique Jan GECSEI fut le généreux donateur de l'ordinateur que j'ai eu le loisir d'utiliser pour l'élaboration du manuscrit. Mesdames BRETAGNOLLE et FUCHS ont photocopié et relié ce texte avec tout leur savoir-faire.

Les deux premières années de cette thèse ont coïncidé avec mes troisième et quatrième années de scolarité à l'Ecole Normale Supérieure, en tant qu'"élève-fonctionnaire stagiaire", d'octobre 1987 à septembre 1989. Par la suite, j'ai reçu d'une Bourse de Formation à la Recherche, accordée par Ministère de la Recherche et de la Technologie, qui m'a permis de mener ce travail à son terme. Je tiens à souligner ici les conditions très favorables dont j'ai pu bénéficier tout au long de ces quatre années.

A ma mère, affectueusement.

# INTRODUCTION

1

## I LE DILEMME BIAIS/VARIANCE DANS LES RÉSEAUX DE NEURONES FORMELS

<b>1</b>	<b>Introduction</b>	9
<b>2</b>	<b>L'inférence non-paramétrique</b>	12
2.1	Estimation d'une distribution de probabilité	12
2.1.a	Fonction de prédiction	12
2.1.b	Minimisation de l'erreur quadratique et fonction de régression	14
2.1.c	Estimation non-paramétrique et consistance	17
2.2	Biais et Variance	19
2.2.a	Décomposition de l'erreur quadratique moyenne	19
2.2.b	Le dilemme biais/variance et la consistance	22
<b>3</b>	<b>Expériences numériques</b>	24
3.1	Méthodes	24
3.1.a	Méthode des plus proches voisins	24
3.1.b	Méthode des fenêtres de Parzen	26
3.1.c	Réseaux à couches avec algorithme de rétro-propagation	28
3.2	Expériences sur des données artificielles	30
3.2.a	Présentation du problème	30
3.2.b	Plus proches voisins	33
3.2.c	Rétro-propagation	36
3.3	Expériences sur des caractères manuscrits	38
3.3.a	Présentation des données	38
3.3.b	Plus proches voisins et fenêtres de Parzen	41
3.3.c	Rétro-propagation	43
<b>4</b>	<b>Discussion</b>	46
4.1	Réseaux de neurones et inférence non-paramétrique	46
4.2	La construction d'un "biais"	51
4.3	Conclusion	54

# II APPARIEMENTS DE GRAPHES ET RECONNAISSANCE DES CARACTÈRES MANUSCRITS

<b>1</b>	<b>Introduction</b>	59
<b>2</b>	<b>Algorithme d'appariement élastique</b>	63
2.0	Résumé préliminaire	64
2.1	Représentation des caractères par des graphes	67
2.1.a	Rappels sur les graphes étiquetés	67
2.1.b	Représentation des caractères par des grilles régulières	68
2.1.c	Coordonnées de nœuds	71
2.1.d	Représentations conformes et difformes	75
2.2	Energie d'appariement	77
2.2.a	Energie élastique de déformation	78
2.2.b	Energie de couplage des images	81
2.2.c	Energie totale	82
2.3	Optimisation de l'énergie	84
2.3.a	Forces locales et pénalités locales	85
2.3.b	Equilibre local élastique	88
2.3.c	Equilibre local complet	91
2.3.d	Equilibre global	94
2.4	Pratique	95
2.4.a	Discrétisation des images	95
2.4.b	Adaptation de l'énergie	97
2.4.c	Equilibre local	99
2.4.d	Equilibre global	102
<b>3</b>	<b>Application aux caractères manuscrits</b>	105
3.1	Problème	105
3.1.a	Présentation des données	105
3.1.b	Particularités des graphes et de l'énergie	107
3.1.c	Définition de la "distance élastique"	108
3.2	Classification	117
3.2.a	Principe	117
3.2.b	Résultats	118

# III UN MODÈLE DE DÉVELOPPEMENT ÉPIGÉNÉTIQUE DU SYSTÈME NERVEUX : RÔLE POSSIBLE DANS LA PRÉDISPOSITION AU LANGAGE

<b>1</b>	<b>Connexionnisme et représentation neurale</b>	131
<b>2</b>	<b>Le développement épigénétique du système nerveux</b>	136
2.1	Embryogenèse et neurogenèse	140
2.1.a	Différenciation neuronale	141
2.1.b	Division et migration	142
2.1.c	Croissance des axones	143
2.1.d	Mort neuronale	144
2.1.e	Formation des synapses	145
2.2	Focalisation de l'innervation	147
2.2.a	Jonction neuromusculaire	148
2.2.b	Système autonome	148
2.2.c	Cervelet	149
2.2.d	Cortex visuel primaire	150
2.2.e	Projection rétinotopique	153
2.3	Mécanismes du développement de la sélectivité	155
2.3.a	Nécessité d'un mécanisme compétitif	155
2.3.b	Nécessité d'un mécanisme coopératif	156
<b>3</b>	<b>Présentation du modèle</b>	159
3.0	Résumé préliminaire	159
3.1	Caractéristiques du réseau	161
3.1.a	Neurones et connexions	161
3.1.b	Excitation et inhibition	162
3.1.c	Contraintes morphologiques	163
3.1.d	Variables temporelles et délais	165
3.2	Equations d'évolution	167
3.2.a	Activités neuronales	168
3.2.b	Modèle simple de potentiation synaptique	171
3.2.c	Modèle des groupements synaptiques	173
3.2.d	Compétition intersynaptique	175
3.2.e	Bilan	180

<b>4 Analyse du modèle : principes de fonctionnement</b>	182
4.1 Activités seules dans un réseau homogène	185
4.1.a Première itération	186
4.1.b Itérations suivantes	187
4.1.c Stabilité de l'activité moyenne	188
4.2 Une itération élémentaire	190
4.2.a Effet des corrélations	191
4.2.b Effet de la compétition	193
4.2.c Variation totale des connexions	194
4.3 Connexions seules sous un régime forcé	198
4.3.a Propagation aléatoire : recoupements et destruction	199
4.3.b Propagation régulière : création d'une chaîne	203
4.4 Conditions de stabilité des chaînes	208
4.4.a Entretien d'une chaîne par ondes d'activation	209
4.4.b Transmission des activations sur un maillon	211
4.4.c Conditions de propagation d'une onde	213
4.4.d Conditions de déclenchement d'une onde	216
4.5 Auto-organisation des chaînes par pulsations locales	219
4.5.a Neurones pulsatiles	220
4.5.b Effet de la première pulsation	222
4.5.c Effet de la deuxième pulsation	224
4.5.d Pulsations suivantes et genèse du premier maillon	225
4.5.e Genèse d'une chaîne par empilement de maillons	231
<b>5 Simulations</b>	235
5.1 Neurones excitateurs et délais uniformes	235
5.1.a Formation d'un premier maillon	235
5.1.b Croissance d'une chaîne	242
5.1.c Autres exemples de chaînes	252
5.2 Neurones inhibiteurs et délais non-uniformes	253
<b>6 Composition et configurations complexes</b>	260
6.1 Composition dynamique de deux chaînes	260
6.2 Propriétés générales des CST	266
6.3 Croissance de configurations complexes	269
<b>BIBLIOGRAPHIE</b>	275

# INTRODUCTION

Le thème central de cette thèse est la notion de *représentation neurale*. Partant du postulat selon lequel toute "entité" (sensation, perception, concept, mot, objet extérieur, action, etc.) est représentée dans le système nerveux sous forme d'un état d'activités neuronales, nous posons la question fondamentale de savoir quelles pourraient être la structure et les caractéristiques d'un tel état de représentation. Cette question a été souvent discutée depuis le début des neurosciences modernes, mais il est d'usage de penser que le *niveau moyen* d'activité des neurones est une variable essentielle dans le codage neuronal. Ainsi, la théorie classique veut que les entités soient généralement codées par des "assemblées de cellules" (Hebb, 1949), c'est à dire des configurations spatiales de niveaux d'activité.

Nous proposons quant à nous une autre conception du format de représentation, qui sera développée tout au long de cette thèse sous différentes perspectives, à travers des modèles pratiques et théoriques, ainsi que dans des simulations numériques. Pour résumer : l'alternative que nous défendons (von der Malsburg, 1981) envisage des "*codes d'ordre supérieur*", c'est à dire implique de façon générale des *corrélations temporelles* entre les activités des neurones. Ainsi, la représentation n'est plus seulement fondée sur des moyennes temporelles individuelles  $\langle x_i \rangle$ , c'est à dire sur une distribution "d'ordre 1", mais essentiellement sur des "événements d'ordre  $n$ "  $\langle x_{i1} x_{i2} \dots x_{in} \rangle$ , et en particulier sur des corrélations à deux neurones  $\langle x_i x_j \rangle$ . Bien sûr, le schéma d'ordre 1 est motivé par l'étude des aires sensorielles primaires, où l'on a découvert que les cellules possèdent des propriétés d'activation sélective, et semblent par conséquent coder individuellement des stimuli spécifiques. Cependant, il s'agit de propriétés très locales, qui doivent être ensuite *intégrées* pour fournir une représentation globale de la scène. Il n'est alors pas plausible que cette opération d'intégration soit effectuée par un petit groupe de neurones hautement spécialisés (appelés souvent "cellules grand-mères"), ni même qu'un état distribué d'activités moyennes puisse être conservé tel quel, en raison de la confusion qui résulterait d'une superposition avec d'autres états d'ordre 1. C'est pourquoi nous pensons que l'intégration requiert un code d'ordre plus élevé que 1, pour représenter les *relations* entre des composants élémentaires initialement décorrélés. En un mot, les ordres supérieurs apportent une *structure* au format de représentation, structure qui fait essentiellement défaut aux "listes de caractéristiques" que sont les assemblées de cellules.

Cette thèse comporte trois parties qui s'attacheront, chacune à sa manière, à explorer les possibilités théoriques des codes d'ordre supérieur. Dans un premier temps, on envisagera l'étude du problème de la représentation dans le cadre de la *théorie statistique de l'estimation*. En effet, il s'avère depuis le renouveau d'intérêt pour les modèles connexionnistes que la grande majorité des modèles de réseaux de neurones (généralement inspirés du Perceptron, dans une version "Perceptron multi-couches"), sont utilisés pour résoudre des problèmes d'estimation, et plus particulièrement des problèmes de *régression*. Ainsi, ces modèles neuronaux tendent à produire une fonction d'interpolation des données, c'est à dire une approximation statistique qui ne repose sur aucune "connaissance a priori" (on parlera aussi

d'*inférence non-paramétrique*). Cette méthode fait ses preuves dans de nombreux domaines d'application, où elle peut parfois fournir de meilleures performances que les systèmes experts. Cependant, on est en droit de s'interroger sur la pertinence d'une telle démarche dans la perspective neurobiologique : apporte-t-elle une véritable réponse au problème de la représentation neurale ? Est-elle en mesure de résoudre des problèmes cognitifs complexes tels que ceux de la perception invariante ? En un mot, quand on s'intéresse au système nerveux réel, et quand on cherche à comprendre les mécanismes de la cognition, il convient de se demander si la notion d'*apprentissage*, qui est fondamentale dans toute cette approche, joue effectivement un rôle aussi important qu'on le prétend.

C'est ainsi que, dans cette première partie, nous montrerons formellement que les performances d'un estimateur statistique, quel qu'il soit, sont nécessairement limitées en pratique : on verra en particulier que *l'erreur quadratique moyenne* associée à cet estimateur (c'est à dire l'espérance de l'erreur de prédiction sur différents "ensembles d'apprentissage") se décompose en deux termes, un *biais* et une *variance*, qui montrent clairement les causes de cette limitation. Le premier représente l'écart moyen de la prédiction par rapport à la réponse attendue, et le deuxième représente les fluctuations de cette prédiction selon les exemples d'apprentissage. Les conclusions auxquelles nous aboutirons, et qui s'appuieront sur deux études numériques approfondies, montreront que l'estimation statistique peut produire de bonnes performances dans des problèmes *simples*, c'est à dire des problèmes où il n'est pas demandé à la machine de *généraliser* de façon non-triviale. Dans le cas contraire, c'est à dire si l'on attend une extrapolation pertinente à partir de la première phase d'interpolation, nous pensons qu'il est alors nécessaire de se préoccuper avant tout de la *bonne représentation* des données du problème, et que cette question est de loin la plus importante, avant même celle de l'apprentissage.

En résumé, cette étude statistique relativement formelle aura pour but de démontrer la nécessité incontournable de la recherche d'une représentation adéquate dans les réseaux de neurones, et ce pour tous les problèmes qu'ils ont à traiter. C'est aujourd'hui une conclusion à laquelle semblent adhérer un nombre croissant d'auteurs. Cependant, l'espoir principal misé dans les réseaux de neurones il y a encore quelques années était que ces réseaux découvrent "par eux-mêmes" ces bonnes représentations, en affectant aux unités cachées les propriétés adéquates (extraction des caractéristiques pertinentes) grâce à l'ajustement des poids synaptiques.

La deuxième partie décrit une réalisation concrète d'un modèle de reconnaissance de formes s'inspirant de l'approche de Christoph von der Malsburg (1981). Cet auteur propose un format de représentation original dans le système nerveux, consistant en un codage "d'ordre 2", c'est à dire un codage dans lequel les relations entre objets d'une scène visuelle sont exprimés par des *corrélations* temporelles entre activités neuronales. Von der Malsburg a élaboré cette représentation après avoir conclu que le schéma conceptuel utilisant uniquement une représentation en niveaux d'activité aboutissait à des difficultés fondamentales insurmontables, et à l'impossibilité de traiter de problèmes complexes, y

compris de reconnaissance de formes. En résumé : le format classique des assemblées de cellules *est dépourvu de structure*, donc, dans ce contexte, le seul moyen de coder l'information relationnelle indispensable pour éviter la "catastrophe de superposition" (i.e. le mélange de deux assemblées simultanément actives) est de dédier de nouvelles cellules pour chaque nouvelles combinaisons de caractéristiques ("cellules grand-mères"). Cette solution n'est évidemment pas réaliste, et il s'avère beaucoup plus naturel de coder des relations par des corrélations entre cellules que par de nouvelles cellules. Par ailleurs, à travers un mécanisme de type hebbien (rétro-action positive), ces corrélations d'activités renforcent les connexions qui les soutiennent : ainsi, une structure relationnelle sera codée de façon équivalente par les *connexions*.

Nous étudierons ici un modèle de reconnaissance de formes directement inspiré de ce principe, mais qui s'en trouve cependant éloigné dans la réalisation pratique. L'opération centrale de ce modèle est une opération d'*appariement* entre les représentations relationnelles de deux formes, c'est à dire entre deux *graphes étiquetés*. Cet appariement est conçu comme une opération d'*optimisation* dont le but est de trouver une correspondance entre ces deux graphes qui préserve au mieux à la fois les relations de voisinage géométriques entre les nœuds du graphe et les étiquettes portées par ces nœuds. Autrement dit, il s'agit de maintenir un équilibre entre deux contraintes, l'une portant sur la *déformation* de la structure, et l'autre sur la *ressemblance* entre les images déformées. Ce modèle sera étudié de façon détaillée, et les expériences numériques utiliseront des caractères manuscrits. Les performances obtenues par cette méthode seront également comparées à celles des réseaux à couches, et montreront que cette approche "relationnelle" permet de meilleures performances dans une tâche de reconnaissance des caractères manuscrits. Pour cela, on utilisera le critère des "plus proches voisins" associé à une métrique dérivée de l'opération d'appariement de graphes, au lieu de la traditionnelle métrique de Hamming entre tableaux de pixels.

Après cette deuxième partie présentant une application concrète, la troisième partie abordera le problème de la représentation neurale sous un angle beaucoup plus théorique et spéculatif. On s'intéressera de façon très générale au problème de la *compositionnalité* des processus cognitifs : ces propriétés de composition sont déjà évidentes au niveau des tâches de perception invariante, mais elles apparaissent de façon encore plus nette dans les facultés du *langage*. Le langage est en effet souvent décrit comme un "jeu de construction" (jeu de "composition"), dans lequel les objets manipulés sont des symboles dotés d'une structure combinatoire interne, cette structure leur permettant de s'assembler en *symboles composites* dont le sens dépend précisément de la façon dont les constituants sont assemblés entre eux. Ici plus que jamais, la question du format de représentation prend toute son importance, et nous proposons justement que le système nerveux utilise un code d'ordre élevé pour représenter les entités linguistiques.

Le modèle que nous exposerons s'attachera surtout à montrer que la propriété de compositionnalité peut être conçue comme un *développement ontogénétique* progressif au cours des phases précoces de la synaptogenèse, et en ceci nous rejoignons les thèses de

Chomsky selon lesquelles le langage "grandit" chez l'enfant, c'est à dire est le résultat d'un processus de maturation semblable à celui des autres organes. Cette hypothèse peut apparaître surprenante a priori, mais nous démontrerons, en utilisant des principes de développement très généraux et admis depuis longtemps dans le cas de la croissance de certaines structures du système nerveux central (système visuel en particulier), qu'il est possible d'observer l'*auto-organisation* d'un réseau aléatoire en une structure régulière de connexions, conjointement avec l'apparition de corrélations durables entre des activités neuronales initialement désordonnées. Les connexions et les corrélations se renforcent mutuellement à travers un phénomène de *coopération*, la stabilité de l'ensemble étant assurée par une contrainte de *compétition*. Des "*configurations spatio-temporelles*" complexes émergeront spontanément du réseau, et représenteront les constituants élémentaires (les "briques" élémentaires) des symboles ayant les propriétés requises citées plus haut, à savoir : une structure combinatoire interne et la possibilité de s'assembler entre eux de multiples manières, ouvrant ainsi la voie à une hiérarchie virtuellement infinie de combinaisons.

En résumé, cette thèse comporte trois volets, tous centrés autour de la notion de représentation. Le premier chapitre n'a pas la prétention de présenter une recherche originale : il s'agit plutôt d'une étude qui met en lumière le problème de l'apprentissage du point de vue statistique. Cette étude a pour but d'étayer l'opinion selon laquelle la notion de *représentation* est fondamentale et doit devenir prioritaire devant les préoccupations d'"apprentissage" ou d'"adaptativité" du système à une situation particulière. On est alors naturellement conduit, en s'inspirant des idées de von der Malsburg, à considérer des codages d'ordre 2 pour un problème de reconnaissance des formes : c'est l'objet de la seconde partie, dans laquelle on présente une application sur des caractères manuscrits, utilisant des appariements de structures relationnelles. La troisième partie, plus spéculative, essaie d'indiquer la voie dans laquelle on peut imaginer que les symboles linguistiques soient représentés dans le système nerveux, toujours dans la perspective d'un codage d'ordre élevé, mais en insistant sur la question du développement ontogénétique.