Chapitre 3

Modularité et apprentissage dans les réseaux de neurones récurrents

3.1. Introduction : approche dynamique et approche neuromimétique

Un agent artificiel est doté de capteurs lui fournissant des informations sur son environnement et d'actionneurs permettant d'agir sur ce même environnement. Produire le mouvement correct, « intelligent », nécessite de rassembler toutes les informations disponibles et en particulier de prendre en compte, sur le long terme, l'expérience passée et, sur le court terme, les perceptions passées. On attend d'un agent dit « dynamique » qu'il aille au-delà des actions purement réactives (actions réflexes qui ne prennent en compte que la perception immédiate) en se construisant, en interaction avec l'environnement, un répertoire de comportements. Les caractéristiques *attendues* d'un agent dynamique vouent donc une importance prépondérante aux capacités d'apprentissage, d'adaptation [1] et de généralisation, autant de caractéristiques qui sont – entre autres systèmes – des propriétés classiquement admises pour les réseaux de neurones artificiels (*Artificial Neural Networks* ou ANN). De fait, la construction d'agents dynamiques autonomes fait la part belle aux réseaux de neurones comme outils d'implémentation des comportements de l'agent même si l'origine de cet engouement est, il est vrai, autant due aux caractéristiques des réseaux neuromimétiques qu'à une volonté pas toujours assumée ou justifiée de copier leurs homologues naturels puisque « ils marchent... ».

1. On considère ici l'adaptation au sens « d'adaptation dynamique aux variations d'un environnement non stationnaire » et non au sens de l'adaptation par sélection naturelle.

Chapitre rédigé par Guillaume BESLON et Emmanuel DAUCÉ.

Il n'entre pas dans le cadre du présent ouvrage – et, *a fortiori*, du présent chapitre – de décrire, même succinctement, les modèles de réseaux de neurones artificiels et nous renvoyons pour cela le lecteur vers les nombreux ouvrages traitant de cette question (Dreyfus, Martinez, Samuelides, Gordon, Badran, Thiria et Hérault, 2002 ; Haykin, 1998). Notre but est ici de décrire les réseaux de neurones en tant que systèmes dynamiques, en nous focalisant en particulier sur leur capacité à former des comportements modulaires dans le cadre d'un apprentissage. Or, le paradigme connexionniste, loin de constituer une famille de modèles unifiée, regroupe des approches extrêmement diverses puisant leurs fondements aussi bien dans la biologie que dans les statistiques ou la thermodynamique. C'est pourquoi, avant d'entrer plus avant dans les détails de la mise en œuvre d'agents « neurodynamiques », nous nous proposons d'effectuer une rapide introduction en forme de croisement entre les propriétés *attendues* d'un agent dynamique et celles *proposées* par les différents modèles connexionnistes.

3.2. Le paradigme connexionniste à l'épreuve de l'approche dynamique

3.2.1. *Les réseaux de neurones, des systèmes dynamiques ?*

La moindre incursion dans le « monde » connexionniste montre rapidement combien sont variées les approches, nombreux les modèles et diverses les architectures relevant de ce paradigme. En outre, l'étendue des domaines d'applications des réseaux de neurones est à la hauteur de leur variété. Au sein de cette constellation de modèles et d'applications, nous allons essayer de proposer une classification succincte illustrant l'adéquation des différents modèles avec l'approche dynamique.

Classiquement, un modèle neuronal est considéré comme un graphe d'automates, généralement non linéaires (les neurones), évoluant dans le temps suivant une certaine dynamique. Si les automates utilisés s'éloignent généralement peu d'un modèle classique [2] composé d'un additionneur et d'une fonction de transfert, cela n'est pas le cas pour le graphe reliant ces automates entre eux ni pour la dynamique d'évolution du réseau. Or, pour une grande part, le graphe et la dynamique sont étroitement couplés.

La question de la dynamique des réseaux de neurones comporte deux aspects très différents, bien que fortement liés. En effet, elle concerne aussi bien l'évolution temporelle de l'activité des cellules neuronales qui composent le réseau (dynamique de

2. Nous n'aborderons pas ici le cas des modèles neuronaux dits *integrate and fire* ou des *neurones à spikes* ; voir aussi chapitre 6. Cet « oubli » est essentiellement dû à des contraintes de place, car nous pensons que ces modèles ont, et auront de plus en plus, leur place dans la galaxie des approches dynamiques puisqu'ils permettent de rendre compte de phénomènes de synchronisation plus finement que les modèles classiques qui se limitent à considérer l'activité neuronale *via* la fréquence des potentiels d'action produits par la cellule. Nous encourageons vivement le lecteur intéressé par l'approche dynamique et les modèles neuromimétiques à suivre attentivement les évolutions dans ce domaine.

propagation de l'influx nerveux) que la loi d'évolution des poids synaptiques entraînée par cette activité (dynamique d'*évolution* du réseau). Nous verrons, dans la suite de ce chapitre, que ces deux dynamiques ne sont pas suffisantes pour qu'un réseau de neurones présente un ou plusieurs *comportements* et qu'une troisième dynamique doit émerger des interactions entre le réseau et son environnement. Cependant, il nous paraît nécessaire de rappeler auparavant combien la dynamique de propagation est liée à la structure du graphe des connexions et, surtout, de relier, *via* cette dynamique, la structure du réseau aux contraintes de l'approche dynamique.

En réseau de neurones, la question de la dynamique de propagation est fondamentalement liée à l'existence de cycles dans le graphe de connexions, autrement dit à son caractère *récurrent*. En effet, l'effet premier des récurrences est d'introduire, pour une même cellule neuronale, des interactions entre son activité au temps t et son activité au temps $t + \tau$ (où τ est la longueur du chemin récurrent considéré). Contrairement aux réseaux non récurrents (typiquement les réseaux en couches généralement utilisés en classification), un tel réseau ne produira donc pas, pour une entrée donnée, une et une seule sortie mais au contraire un motif de sortie spatio-temporel dont il n'est généralement pas possible de garantir la convergence vers un point fixe ni même vers un cycle limite.

Le monde des réseaux de neurones se trouve donc globalement divisé entre les modèles non récurrents, aussi appelés réseaux *feed forward*, généralement structurés en couches de neurones, et les modèles récurrents. La relative simplicité des premiers, permettant l'analyse mathématique, a conduit à populariser leur utilisation. Ils concentrent, de fait, une très large majorité des travaux théoriques ou applicatifs en connexionnisme. A l'inverse, les réseaux récurrents restent trop souvent cantonnés à un cercle scientifique étroit, intéressé précisément par leur dynamique « complexe ». Nous souhaitons défendre ici l'idée que seuls les réseaux récurrents se prêtent à la construction d'agents dynamiques. Pour cela, il importe avant tout de rappeler la situation dans laquelle se trouve un agent situé plongé dans un environnement dynamique. Un tel agent ne reçoit pas, de la part de son environnement, une suite d'entrées indépendantes. Au contraire, les entrées de l'agent forment un motif spatio-temporel. Qui plus est, les entrées reçues (les *perceptions*) de l'agent sont généralement étroitement corrélées à ses sorties (ses *actions*)[3]. Pour que le réseau de neurones puisse, d'une façon ou d'une autre, interpréter les stimulations reçues de l'environnement, il doit donc les considérer relativement à une dynamique externe (celle de l'environnement) et, conjointement, à une dynamique interne (la sienne propre). On voit explicitement apparaître ici la notion de *couplage* entre l'agent et son environnement (voir chapitre 1), couplage que l'on doit impérativement retrouver entre le réseau de neurones et son environnement. Est-il besoin de préciser que ce couplage des dynamiques ne peut

3. Que l'on songe aux variations de perception – dont nous n'avons le plus souvent même pas conscience – induites par la simple action de tourner la tête.

en aucun cas être obtenu avec des réseaux *feed forward* puisque ceux-ci se contentent de suivre servilement la dynamique de leur environnement en associant passivement une sortie à chacune des entrées qui leur sont présentées [4]. A l'inverse, un réseau récurrent peut, sous certaines conditions, entretenir une dynamique interne propre, même en l'absence de stimulation, tandis que, lorsqu'elles sont présentes, il peut les utiliser pour modifier (si nécessaire) sa dynamique interne et réagir (cette réaction pouvant correspondre à une modification immédiate ou différée des sorties du réseau).

3.2.2. *Apprentissage supervisé* versus *apprentissage par renforcement*

Dans le cas de l'approche dynamique, il est fondamental de considérer un système *apprenant* comme un système apprenant *dans un environnement* [5]. Autrement dit, le système n'apprend – ne peut apprendre – que par rapport aux stimulations/interactions qu'il reçoit/entretient avec son environnement. Cette remarque préliminaire conduit à différencier différents paradigmes d'apprentissage en fonction du type de relation que le système (ici le réseau de neurones) entretient avec son environnement. On distingue en particulier trois approches de l'apprentissage :

– *apprentissage supervisé* : l'environnement apparaît dans ce cas, vu du système apprenant, comme une source de couples (X, D) éventuellement ordonnés, où X est un vecteur de stimulation et D la réponse (commande motrice) *attendue*, compte tenu de la stimulation. Le réseau est alors chargé de *mimer* le comportement modèle fourni par l'environnement et éventuellement de l'étendre grâce à ses capacités de généralisation ;

– *apprentissage non supervisé* : l'environnement ne fournit ici que la suite des vecteurs de stimulation X. Le réseau a donc la charge d'extraire des régularités de la suite des vecteurs X ;

– *apprentissage par renforcement* : dans ce cas, l'environnement est source de vecteurs de stimulation X et d'un signal de renforcement R décrivant qualitativement son état. Le réseau doit alors produire un *comportement* visant à maximiser le signal R reçu.

L'apprentissage non supervisé a pour modèle de référence les cartes dites « auto-organisatrices » (*Self Organizing Maps*, SOM) de Kohonen (1982). Ces cartes

4. Ce qui représente précisément leur force pour des tâches de classification ou de reconnaissance de forme.

5. Si cette remarque peut sembler de peu d'utilité lorsque les réseaux de neurones sont utilisés pour des tâches de classification ou de reconnaissance de formes, elle revêt toute sa portée dans le cas de l'apprentissage de comportements par un système dynamique puisqu'un comportement n'existe que relativement à un environnement donné.

s'adaptent à la répartition statistique de leur espace d'entrée et réalisent une classification (catégorisation) de ces entrées sur une carte monodimensionnelle ou bidimensionnelle, selon le principe du *Winner Takes All* (WTA) : suite à l'activation du réseau par une entrée, le neurone « gagnant » sur la carte désigne la catégorie du signal d'entrée, représentée par une position sur la carte.

Souvent considéré comme une forme intermédiaire, à mi-chemin entre l'apprentissage supervisé et l'apprentissage non supervisé, l'apprentissage par renforcement est en pratique un paradigme très spécifique. En effet, dans ce cas, le signal R est reçu suite à la commande produite par le réseau et aux modifications que cette réponse a induites dans l'environnement [6]. Dans le cas de l'apprentissage par renforcement, le système est donc fondamentalement impliqué dans une interaction circulaire avec son environnement. En résumé, si, en apprentissage supervisé ou non supervisé, le système apprenant *extrait* de l'information de son environnement, en apprentissage par renforcement, il crée l'information autant qu'il l'extrait (à l'extrême, il crée l'information qu'il extrait) de par la boucle de causalité circulaire dans laquelle il est impliqué et dans laquelle il implique l'environnement.

Si l'on replace ces différents modèles d'apprentissage dans le cadre de la problématique générale de cet ouvrage, la question de la relation circulaire entre un système et son environnement constitue le fondement même de l'approche dynamique. Dès lors, parmi les trois paradigmes cités, seul l'apprentissage par renforcement permet à un système d'acquérir un état d'équilibre par l'acquisition d'une trajectoire dynamique en relation avec un environnement donné. A l'inverse, l'apprentissage supervisé est destiné à copier (mimer) une dynamique prédéfinie (par l'ensemble des couples (X, D)). Il ne doit alors être considéré « que » comme une méthode d'*identification* d'un système dynamique prédéfini [7].

3.2.3. *Apprentissage hors ligne* versus *apprentissage en ligne*

Outre la classification présentée ci-dessus, les méthodes d'apprentissage sont souvent différenciées de par leur caractère hors ligne (*off-line*) ou en ligne (*on-line*) :

– *apprentissage hors ligne* : le système connaît alors deux phases distinctes : acquisition des connaissances *puis* restitution des connaissances ;

– *apprentissage en ligne* : l'acquisition des connaissances a lieu concouramment à leur utilisation.

---

6. Le signal R reçu au temps t est donc la « conséquence » des réponses du réseau aux temps $t - \delta t$.

7. L'apprentissage supervisé peut parfois être considéré comme l'acquisition d'une dynamique, mais, en l'absence de « but », il ne permet pas de guider la trajectoire puisque celle-ci est déjà contenue dans l'ensemble des couples (X, D).

Au premier abord, la différence entre apprentissage hors ligne et en ligne nous ramène à la différence entre apprentissage supervisé et apprentissage par renforcement : alors que dans le premier cas, on distingue la phase d'apprentissage – durant laquelle l'environnement se manifeste sous la forme de couples (X, D) – de la phase d'utilisation (l'environnement ne fournit plus que les vecteurs de stimulation X), dans le cas de l'apprentissage par renforcement ces deux phases sont étroitement mêlées puisque le système doit en permanence maintenir un compromis efficace entre l'*exploration* et l'*exploitation*[8].

3.3. La modularité comme principe général de construction de systèmes dynamiques versatiles

On attend d'un agent dynamique qu'il soit capable d'adopter certains *comportements* en fonction des contextes environnemental et interne. Il convient à ce stade de préciser ce que nous entendons par « comportement[9] ». Contrairement à la simple commande motrice, qui correspond à l'ordre moteur émis à un certain instant, le comportement prend place sur une durée sensible (au sens de l'observation humaine), qui peut aller de quelques secondes à quelques minutes. Un comportement peut facilement être décrit comme « ce que l'agent est en train de faire » : il recherche une proie, il interagit avec un autre agent, il fuit, il mange, il dort. Notre questionnement ici porte sur la modélisation des transitions entre différents comportements. Qu'est-ce qui fait qu'un comportement se maintient, qu'est-ce qui fait qu'un comportement disparaît au profit d'un autre ? Est-ce lié au contexte sensoriel et environnemental (l'agent, se déplaçant dans son environnement, reçoit de nouvelles stimulations qui activent de nouveaux comportements) ? Est-ce lié à une organisation interne d'un répertoire de programmes et d'actions prédéterminé (cycles éveil-repos, recherche-exploitation, patrons moteurs centraux...) ?

---

8. Cependant, dans le cas d'un agent dynamique, la nuance est plus subtile. On doit distinguer ici les situations dans lesquelles l'apprentissage a lieu « sous un regard extérieur », des situations dans lesquelles le système est entièrement livré à lui-même. En effet, dans le premier cas, l'apprentissage proprement dit est souvent à la charge d'un « superviseur » extérieur au système, comme par exemple lorsque l'acquisition de comportements a lieu *via* un mécanisme d'adaptation génétique. Or, de par sa situation, ce superviseur dispose généralement d'informations globales, hors de portée de l'agent. Comme dans le cas de l'apprentissage supervisé, nous ne considérerons pas ce cas comme de la création *endogène* (émergence) de comportements, puisque l'acquisition des couplages sensoriels et moteurs est alors contrôlée par une entité (le superviseur) située hors de la boucle dynamique.

9. La notion de comportement a été introduite par les behavioristes. Ceux-ci mettent en avant le caractère facilement identifiable et observable du comportement qui lui confère un caractère objectif que ne possèdent pas les analyses psychologiques fondées sur l'introspection.

S'intéresser aux comportements permet de mettre en avant une dynamique prenant place sur une échelle de temps plus lente que la dynamique d'activation. On ne parle plus de transition d'état (le passage d'un état à un autre prenant place sur des échelles de temps courtes) mais de transition d'un comportement à un autre (assimilable à une transition de phase). Un comportement, en tant qu'unité de perception et d'action (à caractère transitoire), peut être vu comme un « superétat[10] ». La transition d'un comportement à un autre est régie par une dynamique qui est d'une autre nature que celle qui préside aux transitions d'états.

Le fait que les comportements puissent être distingués, qu'ils se succèdent au cours du temps, illustre leur caractère *modulaire*. Tout se passe comme si différents modules, indépendants les uns des autres, s'activaient successivement (comme si l'agent changeait de programme). Cette modularité, observée d'un point de vue extérieur à l'agent, pose la question de la modularité interne : comment concevoir un agent artificiel dont les comportements aient un caractère modulaire ? Cette modularité est en effet tout à fait nécessaire si l'on souhaite construire des agents que l'on puisse qualifier de « précognitifs ». C'est bien la capacité à *choisir* un comportement en fonction du but à atteindre (problématique de la *sélection de l'action*) qui commence à caractériser des formes d'intelligence plus élaborées que le simple bouclage entre perceptions et actions.

3.3.1. *Approches ascendante et descendante*

La robotique autonome offre un cadre propice à la mise en place d'architectures dites modulaires. Une première approche, qui consiste à inscrire la modularité dans la structure de l'agent, est qualifiée d'approche descendante (*top-down*). La structure et le nombre de modules sont fixés par le concepteur, qui « fige » d'une certaine manière les possibilités d'évolution de l'agent. La modularité, dans ce cadre, consiste à doter le robot de différents programmes, plus ou moins indépendants les uns des autres ; chacun d'entre eux s'activant sélectivement selon le contexte sensoriel. Chacun de ces programmes peut être implémenté dans un module spécifique, un superviseur étant chargé de donner la main à l'un ou l'autre de ces modules et éventuellement de fusionner l'information produite (structure modulaire et hiérarchique courante en IA). La recherche d'une structure à la fois modulaire *et* évolutive amène à utiliser les propriétés de plasticité des réseaux de neurones pour construire un « réseau de réseaux » ou métaréseau. L'idée se révèle extrêmement complexe à mettre en œuvre. Des architectures incluant un grand nombre de modules ont été définies, mais le processus d'intégration a rarement mis en œuvre plus de quelques unités, en raison en particulier de la difficulté qu'il y a à multiplier les phases d'apprentissage dans les réseaux, de

10. Nous verrons dans la suite de ce chapitre qu'il s'agit d'un régime de fonctionnement.

la difficulté à synchroniser l'activité des différents modules et de l'apparition parfois fortuite de boucles d'activation récurrentes.

L'approche que nous tenterons de privilégier ici est une approche ascendante (*bottom-up*). L'idée est de partir de l'indifférencié pour aller vers la différenciation. Le nombre de modules et la nature des comportements ne sont pas fixés *a priori*, mais émergent des interactions avec l'environnement. Cette approche, de par son caractère incertain (de quel ensemble partir ? Comment faire émerger des structures ? Comment qualifier et identifier ces structures ?), apparaît encore à l'heure actuelle comme assez aventureuse pour les concepteurs de systèmes. Ce que nous cherchons à démontrer, c'est qu'il existe des situations dans lesquelles un système interne complexe, naturellement « embrouillé », s'organise et « crée de l'ordre » d'une façon qui est difficilement prévisible par le concepteur et que les comportements issus de cet apprentissage peuvent présenter une certaine efficacité, c'est-à-dire être viables dans le cadre de l'environnement donné à cet agent.

3.3.2. *Couplage des dynamiques et émergence de structures*

L'opposition entre modularités descendante et ascendante est plus profonde que la simple opposition entre intégration et émergence. En effet, modularité descendante et modularité ascendante ne se situent pas au même niveau d'analyse. Dans le cas de l'approche descendante, les modules sont agencés *a priori* (généralement par un concepteur extérieur au système). Ainsi, même si elles répondent à des contraintes liées à la structure de l'agent et/ou de l'environnement, les conditions d'activation des modules sont déterminées par la structure interne de l'agent seul. Au contraire, dans le cas de l'approche ascendante, c'est au cœur du *couplage* agent-environnement qu'apparaît la modularité. L'organisation du comportement prend ainsi sa source dans les interactions autant que dans la modularité intrinsèque au « cerveau » de l'agent. C'est pourquoi nous parlerons, dans le cas de l'approche ascendante, d'une modularité *faible*, caractérisée par ses effets plus que par ses causes.

L'approche ascendante et la modularité faible puisent leur inspiration dans le modèle de l'énaction prôné par Varela, Rosch et Thompson (1992) : c'est à travers l'action que s'affine la structure interne de l'agent, parce que cette action est structurée (inscrite dans des comportements) et structurante (pour les perceptions), autrement dit parce que cette action prend place dans un monde déjà stable. Par ailleurs, cette relation structurante entre l'agent et son environnement est fondamentalement inscrite dans le temps : c'est la relation de causalité *temporelle* entre les actions et les perceptions qui structure ces dernières ; c'est encore la stabilité *temporelle* du comportement qui permet à l'agent de vivre dans un monde structuré ; c'est enfin la succession *temporelle* des comportements qui les différencie et les rend indépendants. C'est pourquoi, même si l'on ne peut nier le caractère intrinsèquement modulaire et structuré du

cerveau (l'identification d'aires spécialisées à la surface du cortex en étant une illustration), il nous apparaît intéressant de prendre le parti pris d'étudier une architecture neuronale non structurée pour voir dans quelles conditions des structures dynamiques peuvent y apparaître et s'y déployer *dans le temps*.

3.3.3. *Attracteurs et structures dynamiques transitoires dans les grands réseaux récurrents*

Pour illustrer cette notion d'émergence de structures, nous nous proposons, dans un premier temps, de présenter les propriétés de grands ensembles de neurones à connectivité aléatoire[11]. Nous prenons comme graphe de connexions celui de Hopfield (1982) qui est totalement récurrent (voir aussi chapitre 2). Chaque neurone est relié à tous les autres. Dans notre modèle, tel qu'il a été défini dans les travaux de Doyon, Cessac, Quoy et Samuelides (1993) et de Daucé, Quoy, Cessac, Doyon et Samuelides (1998) (et tel qu'il est décrit formellement dans le chapitre 5), les forces d'interaction entre neurones sont déterminées à la création du réseau selon une loi normale. La matrice de connectivité est donc homogène[12] : le réseau est non structuré. Ce type de réseau récurrent est en tant que tel un système dynamique. Le modèle est connu pour présenter une route vers le chaos par quasi-périodicité (Doyon, Cessac, Quoy et Samuelides, 1993). Dans ce cadre, le paramètre g, qui représente la sensibilité des neurones, est le paramètre de bifurcation. Il peut être vu comme une entrée extérieure au système. Sur la figure 3.1, nous regardons l'évolution de la dynamique moyenne sous l'influence d'une évolution lente et continue de g (entre $g = 4$ et $g = 6$). Cette figure met en évidence trois régimes dynamiques : point fixe, cycle et chaos, avec des transitions brusques (bifurcations) entre les différents régimes. Chaque régime contraint les neurones à adopter un certain mode d'interaction, qui limite fortement les plages de valeurs sur lesquelles ils peuvent évoluer. Certains neurones restent par exemple totalement silencieux (avec nos paramètres, environ 50 % des neurones sont silencieux). Dans un régime cyclique, les neurones actifs sont contraints par la période du cycle global et les signaux individuels présentent tous la même période, avec des décalages de phase d'un neurone à l'autre. Lorsqu'une bifurcation intervient, elle concerne d'un seul coup tous les neurones, qui basculent ensemble dans le nouveau régime. On voit donc, sur ce premier exemple, que la dynamique des neurones peut basculer d'un mode d'organisation à un autre, sans aucun changement du schéma de connexion. Chacun de ces modes d'organisation est associé à une plage de valeurs

11. Ce modèle a été étudié de façon détaillée par l'équipe de l'ONERA Toulouse, tant pour caractériser sa dynamique à taille finie (Doyon, Cessac, Quoy et Samuelides, 1993) que pour décrire les conditions de convergence vers un comportement moyen à la limite des grandes tailles (Cessac, 1995 ; Daucé, Moynot, Pinaud et Samuelides, 2001).

12. Au sens statistique. Les valeurs particulières des forces de connexion arrivant sur un neurone sont différentes pour chaque neurone, du fait du tirage aléatoire.

du paramètre g. C'est cette forme d'organisation, stable sur un certain domaine de l'espace d'entrée (ou sur une certaine durée), que nous appelons par la suite *structure dynamique*.

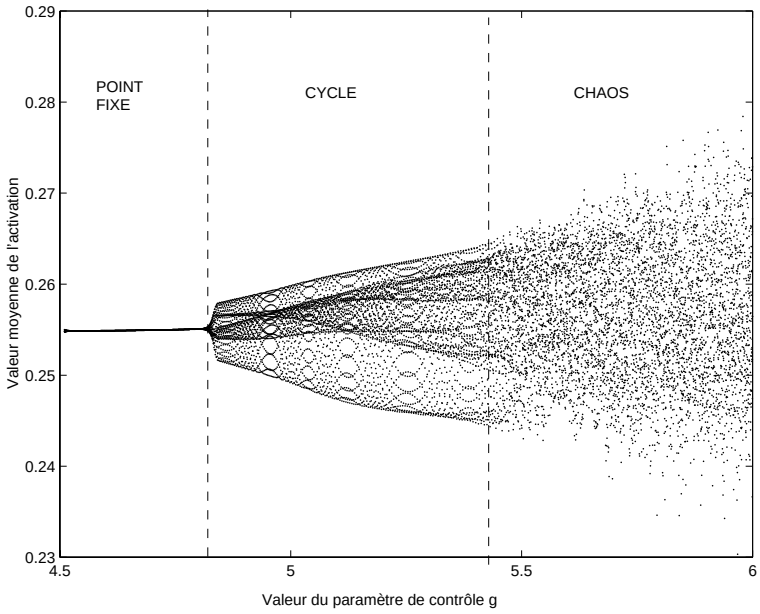

Figure 3.1. *Diagramme de bifurcation sur un réseau récurrent de 200 neurones. La figure représente l'évolution de la valeur moyenne de l'activation sur 14 000 pas de temps, tandis que l'on fait doucement évoluer le gain g entre $g = 4$ et $g = 6$*

Pour enrichir la perception, on utilise une entrée vectorielle $\mathbf{I}(t)$, telle que chaque neurone est stimulé indépendamment des autres (il y a donc N signaux d'entrée, un par neurone). Pour décrire la dynamique de ce système, nous pouvons considérer deux espaces : l'espace des états du réseau, de dimension N, et l'espace des entrées, de dimension N également. Pour une entrée figée et une valeur de g fixée, la dynamique converge vers un attracteur unique. Nous cherchons ici à décrire le comportement du système lors d'une exploration d'une partie de l'espace d'entrée. Cet espace d'entrée étant de grande dimension et ne pouvant pas être exploré exhaustivement, nous déterminons une trajectoire particulière, en veillant à ce que l'évolution de cette trajectoire soit lente par rapport à la vitesse d'évolution des états internes, afin de permettre à la dynamique interne de se stabiliser (de façon transitoire) sur des attracteurs. A chaque pas de temps, de petits ajustements sont donc apportés au vecteur d'entrée, de sorte que la corrélation entre $\mathbf{I}(t)$ et $\mathbf{I}(t+\tau)$ décroît doucement et linéairement pour τ croissant (ainsi la corrélation entre $\mathbf{I}(t)$ et $\mathbf{I}(t+1\,000)$ est de l'ordre de 0,9). La figure

3.2 (haut) nous donne l'évolution de l'activation moyenne au cours du temps, lors d'une de ces explorations de l'espace d'entrée. Nous pouvons observer une dérive à la fois de la valeur moyenne, de l'amplitude et de la nature de ce signal (on peut mettre en évidence des cycles et des bouffées chaotiques). La nature de ces bifurcations est beaucoup plus complexe que précédemment, avec des transitions du cycle au chaos, puis des transitions inverses du chaos vers des cycles. Entre ces transitions, la structure de certains attracteurs semble se maintenir de façon stable sur plusieurs centaines de pas de temps. Nous avons cherché à caractériser plus précisément ces transitions en mesurant par spectre de Fourier la période qui domine le signal (sur des fenêtres glissantes de 500 pas de temps). Cette mesure, donnée sur la figure 3.2b, met en évidence des transitions brusques et bien marquées suivies de longues plages où la période reste stable (ou dérive très légèrement). Ce qui apparaît de façon manifeste, c'est que la dynamique interne structure l'espace d'entrée, en produisant des transitions brusques qui contrastent avec l'évolution progressive et douce du signal d'entrée. On peut dire que la dynamique interne « discrétise » l'espace d'entrée, associant des attracteurs bien distincts à différentes plages de valeurs.

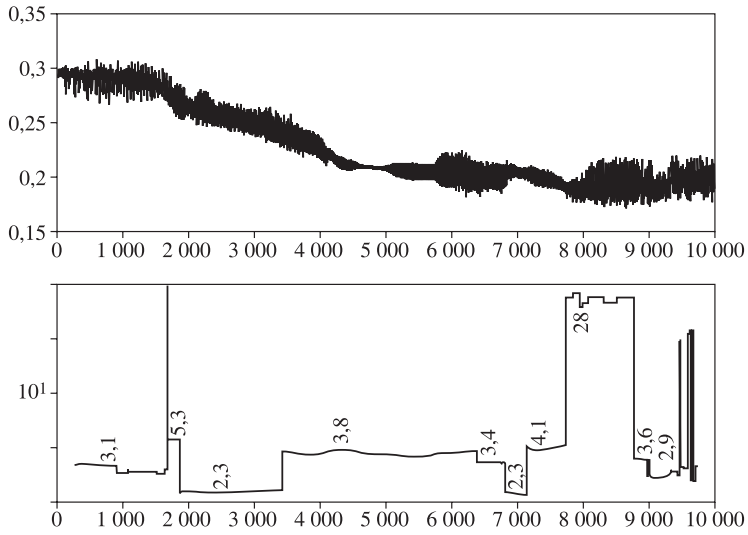

Figure 3.2. *Evolution de l'activité du réseau sous l'influence d'une évolution douce et continue de l'entrée sensorielle (voir texte). La période principale est estimée par spectre de Fourier sur une fenêtre de 500 pas de temps (cette période étant représentative de la période propre de chaque neurone actif). a) Evolution de l'activité moyenne au cours du temps. b) Evolution de la période du mode principal (en échelle logarithmique).*

A chaque transition, la répartition entre les neurones actifs et les neurones silencieux est modifiée (bien qu'en moyenne 40 % à 50 % des neurones restent actifs). Une

bonne partie des neurones actifs avant la transition le restent après, mais c'est l'organisation interne qui est profondément modifiée. Tout se passe comme si une nouvelle architecture interne se mettait en place : certains liens et certains circuits sont activés, d'autres sont désactivés. Tous les neurones se mettent à travailler « d'une autre façon ». Il apparaît clairement que deux dynamiques sont à l'œuvre au sein de ce système : une dynamique rapide (dynamique d'activation) décrit les transitions d'un état à l'autre, tandis qu'une dynamique lente décrit les transitions d'un attracteur à l'autre (transition de phase). Les réseaux récurrents tels que nous venons de les décrire présentent donc naturellement, spontanément, un caractère modulaire dans leur activité en présence d'une stimulation sensorielle non stationnaire. L'espace des entrées étant de très grande taille, le nombre et la variété de ces modes d'organisation semblent potentiellement « infinis ».

3.3.4. *L'apprentissage : émergence et stabilisation des modules comportementaux*

Si les grands ensembles de neurones récurrents présentent, nous venons de le voir, des propriétés de modularité intrinsèques, cela n'est pas suffisant pour établir un couplage comportemental entre l'animat – muni d'un tel réseau – et son environnement. En effet, il reste pour cela à *adapter* la dynamique (*les* dynamiques) du réseau en fonction de son expérience ; c'est le rôle de l'apprentissage.

3.3.4.1. *Apprentissage par renforcement et couplage des dynamiques*

Nous avons vu, dans les paragraphes introductifs, qu'il existe plusieurs schémas d'apprentissage mais que l'apprentissage de comportements au sens strict (que l'on différenciera donc du « copiage » de comportements) correspond à un apprentissage « par renforcement » et « en ligne ». Depuis le début des années 1980, et en particulier depuis les travaux de Barto, Sutton et Anderson (1983), toute une théorie statistique de l'apprentissage a été construite autour de l'apprentissage par renforcement (Sutton et Barto, 1998). Elle est basée essentiellement sur le modèle markovien qui lui donne un socle stable et pertinent [13]. Pour autant, cette approche est peu adaptée à l'acquisition d'une dynamique couplée animat/environnement. En effet, elle présuppose que l'on dispose de mécanismes de mémorisation explicite des différentes situations rencontrées et des différentes actions possibles afin de pouvoir conserver la trace des transitions explorées et les résultats obtenus. Appliquée aux réseaux de neurones (récurrents ou non), cette approche revient à les utiliser à titre d'outil de mémorisation de couples entrée/sortie (le plus souvent supervisé) et à laisser à un modèle statistique « extra-neuronal » le soin de résoudre la question, centrale, de l'apprentissage par renforcement.

13. Voir par exemple les nombreux travaux sur le TD-Learning (Sutton, 1988) ou le Q-Learning (Watkins et Dayan, 1992).

Nous avons vu au début de cette section que des transitions lentes caractérisent, pour un agent dynamique, le passage d'un comportement à un autre. Il apparaît donc naturel d'exploiter les propriétés de modularité spontanées des grands réseaux récurrents pour concevoir des agents développant des comportements modulaires dans le cadre de l'apprentissage par renforcement. Encore faut-il que cette modularité dynamique soit couplée avec la dynamique du monde dans lequel est plongé l'agent. En d'autres termes, on doit disposer de mécanismes permettant de modifier la dynamique (et les attracteurs) du réseau en fonction de la relation agent/environnement... A l'opposé des approches statistiques, nous prônons ici une correspondance directe entre l'activité du réseau de neurones et les actions produites sur l'environnement. Le réseau ne sert pas seulement à extraire de l'information de l'environnement, mais participe aussi (et en même temps) à la régulation des actions de l'agent au sein de l'environnement. La dynamique du réseau est ainsi totalement couplée à la dynamique d'interaction avec l'environnement. Dans ce cadre, l'algorithme d'apprentissage doit modifier directement la dynamique du réseau (*via* la modification des poids synaptiques) en fonction de cette dynamique elle-même et des conséquences de ce couplage (c'est-à-dire du caractère plus ou moins positif du comportement).

3.3.4.2. *L'apprentissage hebbien*

Pour mettre en œuvre l'apprentissage, nous suggérons l'utilisation de la *règle de Hebb* (1949). Cette règle, dans son énoncé initial, lie le renforcement de la liaison synaptique au fait qu'une corrélation puisse être établie de façon répétée entre l'activation du neurone présynaptique et celle du neurone postsynaptique. Malgré l'ancienneté de son énoncé, cette règle demeure, à l'heure actuelle, la plus plausible pour expliquer la plasticité synaptique observée sur les neurones réels. Outre sa plausibilité biologique, son intérêt dans notre cas est double. Premièrement, il s'agit d'une règle qui repose sur un principe de renforcement tout à fait cohérent avec les principes d'apprentissage énoncés plus haut. Deuxièmement, il s'agit d'une règle *locale,* qui agit indépendamment sur chaque lien et ne fait donc pas appel à des connaissances « supervisées » extérieures à cette interaction locale, ce qui est cohérent avec le principe d'émergence des comportements que nous souhaitons mettre en œuvre.

Depuis l'énoncé initial de Donald Hebb, cette règle a connu de nombreuses interprétations et tout autant de mises en œuvre sur les systèmes artificiels. Sur les réseaux que nous avons réalisés, nous avons choisi d'utiliser une règle différentielle (c'est-à-dire reposant sur la corrélation de la *variation* d'activité des neurones présynaptiques et postsynaptiques) et de prendre en compte le délai de propagation du signal sur l'axone, afin de capturer les phénomènes de dépendance temporelle. Une autre interprétation de la règle de Hebb, nécessaire dans le cadre de l'apprentissage par renforcement, consiste à n'« activer » la règle que lorsque le comportement est récompensé. Ainsi, un comportement neutre n'entraîne aucune modification des poids. Par ailleurs, un comportement néfaste à l'agent est puni, ce qui se traduit par une modification synaptique contraire au principe de Hebb tendant à décorréler

les neurones présynaptiques et postsynaptiques (couramment appelée règle « anti-Hebb »).

3.3.4.3. *Le problème de l'exploration*

Une des questions centrales dans le cadre de l'apprentissage par renforcement est celle de l'exploration de l'espace des comportements possibles (aussi appelé espace d'états/actions). Si l'on souhaite en effet renforcer un comportement, il faut bien le sélectionner, par essai et erreur, parmi un ensemble de comportements possibles. Dans le cadre du modèle récurrent dynamique, nous comptons sur la richesse des dynamiques chaotiques générées en interne pour produire un répertoire de comportements suffisamment vaste pour servir de source à la stabilisation de comportements bénéfiques (parmi un ensemble plus vaste de comportements qui sont soit neutres, soit néfastes). Ce répertoire peut, en outre, être étendu en utilisant un mécanisme d'exploration stochastique de l'espace d'états/actions ou, plus en conformité avec l'approche proposée, de l'espace d'état du réseau de neurones.

3.4. Un exemple : l'apprentissage du comportement de prédation

Afin de valider les principes architecturaux décrits ci-dessus, nous [14] avons utilisé un réseau récurrent pour l'apprentissage d'un comportement de prédation. Dans cette application, un agent évolue dans un environnement clos encombré par deux obstacles, dans lequel une « proie » se déplace aléatoirement et où un espace circulaire réservé constitue le « nid » de l'agent (voir figure 3.3). L'agent doit attraper sa proie (en passant dessus) puis la rapporter au nid.

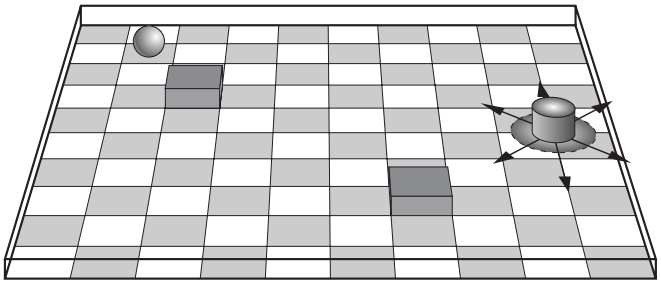

Figure 3.3. *L'environnement et l'animat : l'animat (représenté avec ses systèmes perceptifs et moteurs) est actuellement sur son nid (zone circulaire marquée au sol) tandis que la proie est à l'extrême gauche de l'environnement derrière un obstacle (boule)*

14. Les travaux décrits dans cette section ont été réalisés à l'Institut national des sciences appliquées de Lyon par Soula, Beslon et Favrel (2001).

Dès qu'une proie a été attrapée, l'environnement en crée une autre, identique, et l'agent doit alors choisir entre deux actions antagonistes : rentrer au nid ou chasser la nouvelle proie (mais il ne pourra pas l'attraper tant qu'il n'aura pas déposé l'ancienne dans son nid). La prise et la dépose de la proie sont sanctionnées, en cas de succès, par un signal de renforcement positif ($R = 1$). Inversement, lorsque l'animat heurte un obstacle (quel qu'il soit), il reçoit un signal de renforcement négatif ($R = -1$). Le but de l'animat est donc d'attraper le maximum de proies et de les rapporter au nid sans heurter d'obstacles. Le principal problème lié à cet apprentissage est qu'il met en jeu plusieurs tâches différentes (éviter les obstacles, attraper la cible et retourner au nid) mais surtout que ces différentes tâches ne sont pas présentées *simultanément* au cours de l'apprentissage mais au contraire *successivement* : l'agent doit *avant toute chose* apprendre à se déplacer en évitant les obstacles, *avant même* d'apprendre à chasser sa proie. Enfin, ce n'est qu'une fois qu'il l'a attrapée qu'il peut apprendre à rejoindre son nid. En conséquence, si l'animat n'est pas capable d'apprendre incrémentalement des comportements différents, les différentes tâches vont interférer les unes avec les autres et l'agent ne pourra pas structurer correctement son répertoire de comportements, se contentant alors de reproduire systématiquement la dernière action apprise.

La difficulté d'une telle tâche dépend évidemment des capacités de perception de l'animat. Ici, il dispose de capteurs limités mais réalistes : il perçoit son environnement par le biais de six secteurs perceptifs, régulièrement répartis et couvrant toute la périphérie de son « corps » (le « pouvoir séparateur » de son système perceptif n'est donc que de 60° ; voir figure 3.3). Chacun de ces secteurs perceptifs contient trois capteurs à réponse binaire : un capteur de proximité, un capteur de proie et un capteur de nid. De plus, l'animat dispose d'un capteur proprioceptif lui indiquant s'il porte une proie ou non. On notera que le signal de renforcement positif peut être calculé à partir des variations de ce capteur mais que la nature du renforcement négatif suppose implicitement l'existence d'un capteur de contact.

L'animat dispose de capacités motrices élémentaires basées sur une propulsion différentielle : deux roues diamétralement opposées lui permettent d'avancer ou de tourner sur lui-même.

3.4.1. *Architecture neuronale*

Le comportement de l'animat est déterminé par un réseau récurrent asynchrone, dit « réseau à mode ordonnancé » (Beslon, 1995). Ce réseau comporte vingt neurones d'entrée reliés directement aux capteurs, quarante neurones cachés et deux neurones de sortie. L'état des neurones d'entrée reflète en permanence l'état des capteurs de l'animat. De même, l'état des moteurs est déterminé en permanence par l'état des neurones de sortie. La transition perception-action est soumise à la dynamique interne du réseau récurrent, celle-ci étant elle-même influencée par les entrées du réseau.

La dynamique du réseau est adaptée par une règle hebbienne (règle Hebb/anti-Hebb, voir le paragraphe 3.3.4.2) en fonction du signal de renforcement reçu de l'environnement. Cependant, le réseau utilisé étant asynchrone, un seul neurone change d'état à chaque pas de temps. Nous avons donc utilisé un mécanisme de trace afin de renforcer les connexions en fonction du délai séparant la décharge du neurone activateur de la décharge du neurone activé. En outre, le signal de renforcement primaire est généralement insuffisant pour l'apprentissage de comportements complexes (car il n'intervient qu'à des moments très précis et ne permet donc pas de renforcer la séquence d'actions ayant permis d'y parvenir). Nous avons donc utilisé un deuxième réseau pour estimer un signal de renforcement secondaire (critique adaptative : Sutton et Barto, 1998).

Un des principaux problèmes posés par l'apprentissage de comportements est, nous l'avons vu, l'exploration de l'espace d'états/actions. Or, si l'on peut compter sur le répertoire d'attracteurs « naturels » des grands réseaux de neurones récurrents pour explorer les différents comportements, cela n'est généralement pas suffisant dans le cas d'un réseau ne comportant que quarante neurones. C'est pourquoi nous avons utilisé, dans le réseau de neurones, un facteur de « stress » permettant de bruiter l'activité du réseau et donc de faire basculer le réseau d'un attracteur à un autre. Afin de rechercher le meilleur compromis entre l'exploration (de nouveaux comportements) et l'exploitation (des comportements précédemment acquis), le facteur de stress varie en fonction de la réponse de la critique adaptative, c'est-à-dire en fonction de la qualité *estimée* de l'état dans lequel se trouve l'agent.

3.4.2. *Résultats et interprétation*

Les expérimentations ont été conduites sur une soixantaine d'agents dont les poids synaptiques initiaux étaient tirés aléatoirement. La figure 3.4 montre la courbe d'apprentissage pour l'un d'entre eux [15]. La courbe d'apprentissage montre en particulier que l'animat apprend très rapidement à réaliser les tâches de prédation et de retour au nid puisqu'il lui suffit de quelques cycles « prise/dépose » (de l'ordre de cinq cycles) pour atteindre un régime stable. On voit ainsi que le nombre de cycles réalisés entre t_0 et $t_{200\,000}$ (donc durant les premières phases de l'apprentissage) est équivalent à celui produit entre $t_{200\,000}$ et $t_{250\,000}$, ce qui illustre bien les capacités d'apprentissage.

Cependant, les performances « numériques » ne sont pas les plus pertinentes ici. Nous nous intéressons avant tout à la capacité du réseau à se structurer au cours de l'apprentissage ; les résultats doivent donc être analysés en termes de comportements.

15. La courbe moyenne présente peu d'intérêt en raison de la variabilité des temps d'exploration durant les premières phases de l'apprentissage.

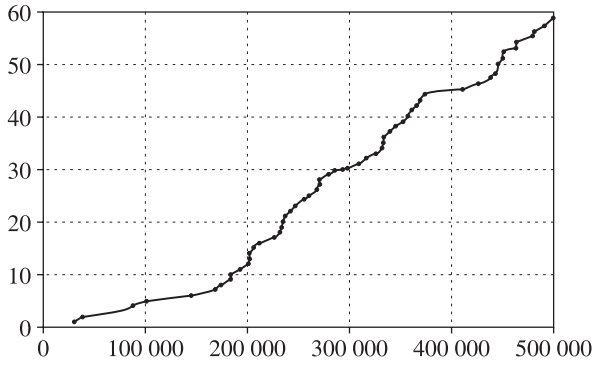

Figure 3.4. *Résultat de l'apprentissage (nombre de prise/dépose en fonction du temps)*

La figure 3.5 illustre le comportement de l'animat et l'activité neuronale correspondante : elle montre clairement les différents comportements appris, qu'il s'agisse de l'évitement d'obstacles ou de la poursuite de la cible. Dans notre modèle, l'apprentissage est incrémental (l'évitement d'obstacles doit obligatoirement être appris antérieurement au comportement de poursuite qui est lui-même obligatoirement appris avant le retour au nid) mais la figure 3.5a montre clairement que les premiers comportements appris, non seulement ne sont pas oubliés, mais surtout, peuvent être combinés efficacement pour obtenir un comportement complexe (voir par exemple les comportements C2 et C4 sur la figure 3.5 qui combinent évitement d'obstacles et *homing*).

D'autre part, du fait de la faiblesse de ses perceptions [16] et du délai lié à la transmission de l'influx neuronal à travers le réseau, l'animat doit développer des stratégies de « chasse » afin de parvenir à atteindre sa cible. Il utilise pour cela de petits mouvements autour d'une position moyenne sur la frontière de séparation entre deux capteurs de direction. Ces micromouvements – comparables à ceux observés en biologie, par exemple chez la mouche – montrent que les stratégies émergentes sont liées au couplage des dynamiques entre l'agent et son environnement. Ils montrent aussi combien un tel couplage serait difficile à mettre en place « de l'extérieur », c'est-à-dire par un concepteur humain.

L'un des intérêts majeurs mis en avant par les simulations est la rapidité de l'apprentissage. Il semble en effet que l'approche proposée, qui limite le nombre de neurones impliqués dans un comportement donné, accélère l'apprentissage en permettant à la loi de Hebb de concentrer son action sur un faible nombre de connexions. Contrairement aux algorithmes d'apprentissage par renforcement plus classiques,

16. Le pouvoir séparateur de son « œil » n'est que de 60 degrés.

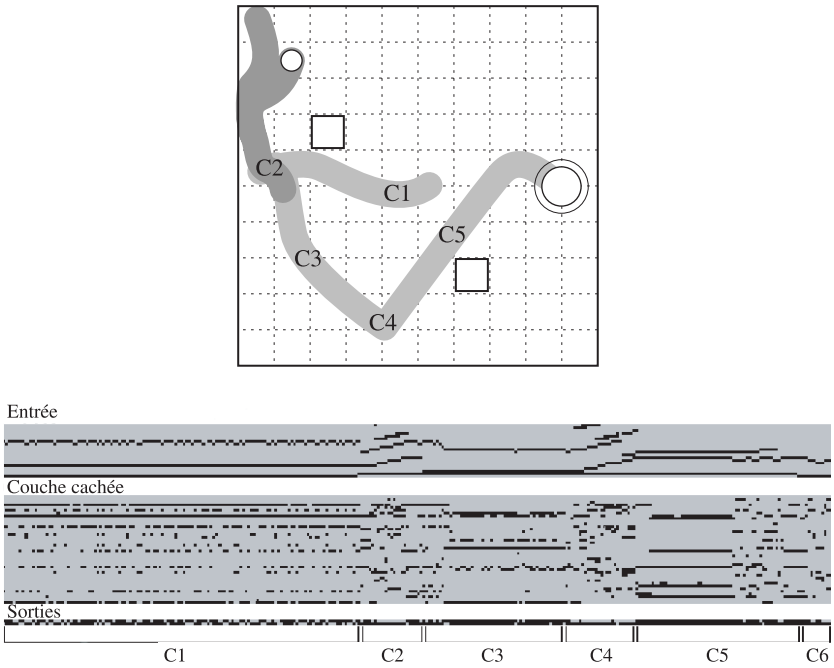

Figure 3.5. *Comportement de poursuite (C1), d'évitement d'obstacles (C2 et C4) et de retour au nid (C3 et C5). L'activité neuronale montre clairement le passage d'une dynamique à l'autre en fonction des différents comportements entrepris par l'agent. Elle montre en outre qu'un comportement ne se réduit pas à un simple attracteur mais correspond bien à un couplage des dynamiques entre le réseau et l'environnement (voir par exemple le comportement de* homing *en phase finale). a) Comportement de l'agent. b) Activité neuronale correspondante.*

l'apprentissage est très proche du *one-shot learning*. Bien que, pour l'instant, toutes les expérimentations aient été réalisées en simulation, la vitesse d'acquisition des comportements nous permet d'envisager avec sérénité des expérimentations sur un robot réel.

3.5. Conclusion

L'application des réseaux de neurones aux agents dynamiques nous oblige à aborder le paradigme connexionniste avec un nouveau regard : on ne doit plus considérer les réseaux de neurones comme des systèmes de classification purs. Ce sont, dans notre approche, des systèmes possédant une dynamique interne que l'on ne cherche pas nécessairement à stabiliser (en tout cas pas indépendamment de l'environnement dans lequel on les stabilise). On peut ainsi utiliser les dynamiques développées dans les

réseaux récurrents comme source de diversité pour la création de comportements. On peut également utiliser la propriété spontanée de réorganisation des circuits internes comme support à la modularité. Cette modularité, qualifiée de faible et ascendante, est issue à la fois de la structure de l'agent (qui contraint la modularité interne) et de la structure d'interaction qui repose sur un corps, des capteurs et des actionneurs inscrits dans un environnement plus vaste.

La démarche proposée consiste à éloigner le concepteur de l'agent qu'il conçoit (c'est-à-dire de limiter autant que possible l'influence du concepteur sur la forme que prend la « connaissance » de l'agent). La principale intervention du concepteur consiste à placer l'agent en situation d'interaction dans un environnement... et à l'y laisser, livré à lui-même.

3.6. Bibliographie

Barto, A.G., Sutton, R.S., & Anderson, C.W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*(5), 834-846.

Beslon, G. (1995). *Contrôle sensori-moteur par réseaux neuromimétiques modulaires*. Thèse de doctorat non publiée, INSA de Lyon, Lyon.

Cessac, B. (1995). Increase in complexity in random neural networks. *Journal de Physique I, 5*, 409-432.

Daucé, E., Moynot, O., Pinaud, O., & Samuelides, M. (2001). Mean-field theory and synchronization in random recurrent neural networks. *Neural Processing Letters, 14*, 115-126.

Daucé, E., Quoy, M., Cessac, B., Doyon, B., & Samuelides, M. (1998). Self-organization and dynamics reduction in recurrent networks: Stimulus presentation and learning. *Neural Networks, 11*, 521-533.

Doyon, B., Cessac, B., Quoy, M., & Samuelides, M. (1993). Control of the transition to chaos in neural networks with random connectivity. *International Journal of Bifurcation and Chaos, 3*(2), 279-291.

Dreyfus, G., Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S., & Hérault, L. (2002). *Réseaux de neurones – Méthodologie et applications*. Paris : Eyrolles.

Haykin, S.S. (1998). *Neural Networks: A Comprehensive Foundation*. New Jersey : Prentice Hall.

Hebb, D. (1949). *The Organization of Behavior*. New York : John Wiley and Sons.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, 79*, 2554-2558.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*, 59-69.

Soula, H., Beslon, G., & Favrel, J. (2001). Controlling an animat with a self-organized modular neural network. *Proceedings of EWLR'2001* (pp. 39-46). Prague, Tchéquie.

Sutton, R.S. (1988). Learning to predict by the method of temporal differences. *Machine Learning, 3*, 9-44.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, Massachusetts : MIT Press.

Varela, F.J., Rosch, E., & Thompson, E. (1992). *L'inscription corporelle de l'esprit*. Paris : Le Seuil.

Watkins, C.J., & Dayan, P. (1992). Q-Learning. *Machine Learning, 8*, 279-292.