# Le connexionnisme

Bernard Victorri

# Le connexionnisme

**Bernard Victorri**

## Historique du connexionnisme

Les premiers modèles connexionnistes datent des débuts de l'intelligence artificielle, avec le célèbre *perceptron* de Rosenblatt (1962), qui fut sans doute le premier modèle de catégorisation perceptive à base de réseau neuromimétique doté d'une capacité d'apprentissage. Mais très vite, la voie de recherche ainsi ouverte a été abandonnée au profit du calcul symbolique prôné par les promoteurs de l'intelligence artificielle classique, en particulier à la suite de critiques sévères de Minsky et Papert (1969), qui ont mis en évidence les limites, jugées à l'époque indépassables, des performances du perceptron. Il a fallu attendre une vingtaine d'années avant que le connexionnisme ne revienne sur les devants de la scène, avec la publication du livre du groupe de recherche *PDP* (*Parallel Distributed Processing*), édité par McCleland et Rumelhart (1986), qui a donné une formidable impulsion aux recherches dans ce domaine. Deux facteurs principaux expliquent cette renaissance. D'une part, au plan technique, la découverte quasi simultanée par plusieurs équipes de chercheurs (Le Cun 1986, Rumelhart *et al.* 1986, Parker 1985) d'un nouvel algorithme d'apprentissage, la méthode de *rétropropagation du gradient de l'erreur*, a montré que l'on pouvait dépasser largement les limites qui avaient handicapé le perceptron. D'autre part, l'état d'esprit des chercheurs avait bien changé. En intelligence artificielle, les méthodes classiques, malgré leurs succès éclatants dans bien des domaines, se révélaient moins aptes qu'on ne l'avait espéré à réaliser un certain nombre de tâches cognitives « de base », comme par exemple la reconnaissance de forme en perception visuelle. De même, en psychologie cognitive, on était à la recherche de modèles plus dynamiques et interactifs, capables de rendre compte de la plasticité et des capacités d'adaptation des systèmes cognitifs, de leur grande sensibilité au contexte, du caractère distribué de certaines représentations, de gradations dans certaines réponses comportementales, toutes choses qu'il était difficile de modéliser avec les outils logico-algébriques issus de l'intelligence artificielle classique.

Le connexionnisme s'est donc développé à cette période en opposition avec le cognitivisme, et il est apparu comme l'ébauche d'un nouveau paradigme en sciences cognitives, capable de combler le fossé entre l'étude des comportements et l'étude des processus neurophysiologiques sous-jacents. Les travaux se sont alors multipliés, les architectures de réseaux se sont diversifiées, en même temps que les objectifs des modélisations. Il existe aujourd'hui une grande variété de modèles connexionnistes, qui couvre un vaste champ d'applications. Certains travaux ont une orientation nettement technologique : ils visent des réalisations informatiques performantes sur des tâches précises (reconnaissance de caractères, de visages, de la parole, etc.) sans se préoccuper de vraisemblance psychologique ou physiologique. D'autres au contraire cherchent à modéliser de la manière la plus réaliste possible le fonctionnement de petits groupes de neurones du système nerveux de tel ou tel animal. D'autres enfin, qui vont nous intéresser principalement ici, ont conservé l'objectif initial de faire le pont entre le fonctionnement du cerveau et celui de l'esprit. Comme on va le voir, ces modèles ont grandement contribué à renouveler la problématique de ce domaine, notamment en ce qui concerne la neuropsychologie. Même si l'on peut douter qu'il atteigne un jour intégralement l'objectif ambitieux qu'il s'est fixé, le connexionnisme joue

indéniablement aujourd'hui un rôle important dans la modélisation du fonctionnement, normal ou pathologique, des systèmes cognitifs.

## Généralités sur les réseaux connexionnistes

### 1. Mécanismes de fonctionnement

Un réseau connexionniste est constitué d'*unités* (appelées aussi *neurones formels*) reliées entre elles par des *connexions*. A chaque connexion est associé un nombre réel, son *poids* (ou *poids synaptique*), qui caractérise l'influence de l'unité source de la connexion sur l'unité cible de la connexion : la connexion est dite *inhibitrice* si son poids est négatif, et *excitatrice* s'il est positif. A chaque unité est associée un autre nombre, son *seuil*, qui peut être lui aussi positif ou négatif.

A tout instant, chaque unité est caractérisée par son *état*, appelé aussi *valeur d'activité*, un nombre qui reste généralement borné (compris entre –1 et 1, ou entre 0 et 1, suivant les cas). Pour faire fonctionner un réseau, il faut se donner une *loi de fonctionnement*, qui permet de calculer l'état de n'importe quelle unité à un instant donné en fonction de l'état de toutes les unités à l'instant précédent, des poids des connexions qui relient ces unités à l'unité en question, et du seuil de cette unité. On appelle cette fonction la *fonction d'activation* de l'unité. Par exemple, sur le réseau présenté figure 1, l'état de l'unité $u_3$ à l'instant t dépend de l'état des unités $u_1$ et $u_5$ à l'instant t-1. Si l'on note $a_i(t)$ la valeur d'activité de l'unité $u_i$ à l'instant t, $w_{ji}$ le poids de la connexion qui va de $u_i$ vers $u_j$, et $s_i$ le seuil de l'unité $u_i$, la fonction d'activation de l'unité $u_3$ aura, classiquement, la forme suivante :

$$a_3(t) = f ( w_{31} . a_1(t-1) + w_{35} . a_5(t-1) – s_3 )$$

où la fonction f est une fonction de type sigmoïde, telle que celle présentée à la figure 2.

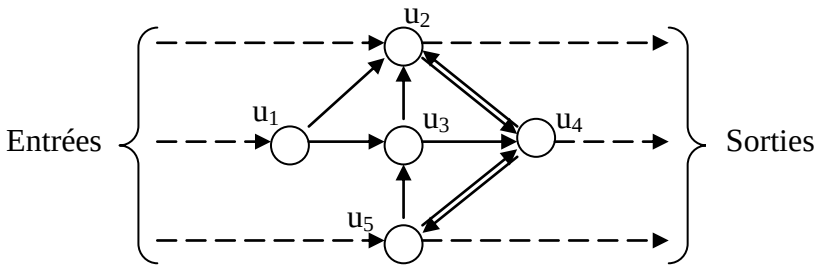


**Figure 1 :** Exemple de réseau

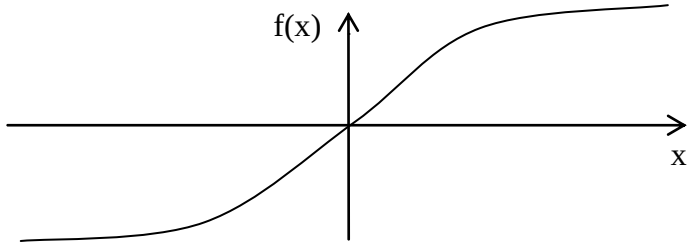


**Figure 2 :** Exemple de sigmoïde

Le principe de fonctionnement d'un réseau est le suivant. A l'instant initial, on fournit les valeurs d'activité d'un certain nombre d'unités, qu'on appelle les *unités d'entrées*[1]. On calcule alors, d'instant en instant, l'évolution de l'état de toutes les unités du réseau. Quand cet état s'est stabilisé, on relève alors les valeurs d'activité d'un certain nombre d'unités, qu'on appelle les *unités de sortie*. Ainsi, mathématiquement parlant, un réseau sert à transformer un ensemble de valeurs, qui forment un vecteur dans l'espace des entrées, en un autre ensemble de valeurs, qui forment un vecteur dans l'espace de sortie. Comme on peut le constater sur l'exemple de réseau présenté figure 1, une unité peut être à la fois unité d'entrée et unité de sortie (c'est le cas de $u_2$ et $u_5$ dans l'exemple). Une unité qui n'est ni unité d'entrée ni unité de sortie est appelée une *unité cachée* ($u_3$ dans l'exemple).
Du point de vue de la neurophysiologie, l'analogie avec un groupe de neurones interconnectés est claire : l'activité de chaque neurone (mesurée par exemple par la fréquence de ses potentiels d'actions) dépend, en première approximation, de la somme pondérée des activités des neurones dont il reçoit l'influence, excitatrice ou inhibitrice, par contact synaptique. Les entrées d'un tel groupe de neurones peuvent provenir d'autres groupes de neurones en amont, ou directement de capteurs sensoriels. De même ses sorties peuvent aller influencer d'autres groupes de neurones ou commander directement des activités motrices.

## 2. Mécanismes d'apprentissage

L'un des principaux intérêts des réseaux connexionnistes réside dans leurs capacités d'apprentissage. Comme on l'a vu, les valeurs des unités de sortie d'un réseau dépendent non seulement des valeurs des unités d'entrée, mais aussi des valeurs des poids des connexions et des seuils des unités. On peut donc modifier la correspondance entre entrées et sorties d'un réseau en changeant ces poids et ces seuils. Un processus d'apprentissage consiste à effectuer de telles modifications jusqu'à ce que la correspondance entre entrées et sorties soit considérée comme satisfaisante par l'expérimentateur.
Il existe deux types d'apprentissage pour les réseaux connexionnistes : *supervisé* et *non-supervisé.*
Dans un apprentissage supervisé, l'expérimentateur doit d'abord confectionner un *échantillon d'apprentissage* qui comporte un certain nombre d'entrées (vecteurs dans l'espace des entrées), ainsi que les sorties désirées (vecteurs dans l'espace des sorties) correspondant à ces entrées. On fait alors fonctionner le réseau, et pour chaque entrée, on compare la sortie obtenue à la sortie désirée. Si ces deux vecteurs ne coïncident pas, on modifie les poids et les seuils, de manière à ce que la prochaine fois que l'on présente cette entrée, la sortie du réseau soit plus proche de la sortie désirée. Les algorithmes d'apprentissage supervisés les plus utilisés sont dérivés de la méthode de rétropropagation qui a contribué au renouveau du connexionnisme (cf. §1).
Dans un apprentissage non supervisé, l'échantillon d'apprentissage ne comporte pas les sorties désirées. L'apprentissage consiste à modifier les poids et les seuils en fonction de l'activité même des unités. Par exemple, un des algorithmes les plus classiques s'inspire d'un mécanisme physiologique postulé par Hebb (1949). Il consiste à augmenter le poids d'une connexion si les deux unités qu'elle relie sont toutes deux très activées et à le diminuer dans le cas contraire.
Une fois l'apprentissage (supervisé ou non) terminé, on fait fonctionner le réseau sur de nouvelles entrées, qui constituent un *échantillon de test*. Si les performances du réseau sont jugées bonnes sur ces nouvelles entrées, on dit que le réseau a pu *généraliser* à partir des exemples qui ont servi à l'apprentissage. Pour qu'un apprentissage soit considéré comme

[1] Parfois, on force les unités d'entrée à conserver leur valeur initiale pendant toute la période de fonctionnement : on dit alors que ces unités sont *contraintes* (*clamped*, dans la terminologie anglo-saxonne).

réussi, il faut bien entendu que ses performances soient bonnes non seulement sur l'échantillon d'apprentissage, mais aussi en généralisation, sur l'échantillon de test.

### 3. Architectures de réseaux

Parmi les architectures les plus répandues, on distingue les *réseaux unidirectionnels multicouches* (*multilayer feedforward networks*), appelés aussi *perceptrons multicouches*, et les *réseaux récurrents* (*recurrent networks*). Dans un réseau unidirectionnel, les connexions ne forment pas de boucles : l'activité se propage directement de couche en couche, depuis la couche d'entrée jusqu'à la couche de sortie (fig. 3). En revanche, dans un réseau récurrent, il existe des boucles (c'est le cas, par exemple, pour les unités $u_3$, $u_4$ et $u_5$ du réseau de la figure 1 §2.1, qui est donc récurrent). Le réseau se comporte alors comme un système dynamique : les activités se modifient les unes les autres au cours du temps jusqu'à ce que l'on aboutisse à un état d'équilibre global, qui est un *attracteur* de la dynamique correspondante. Un réseau est dit *entièrement récurrent* si toutes les unités sont interconnectées par des liens bidirectionnels (fig. 4). Il est dit *simplement récurrent* s'il comporte des couches et que la récurrence est limitée à une ou deux de ces couches (fig.5).

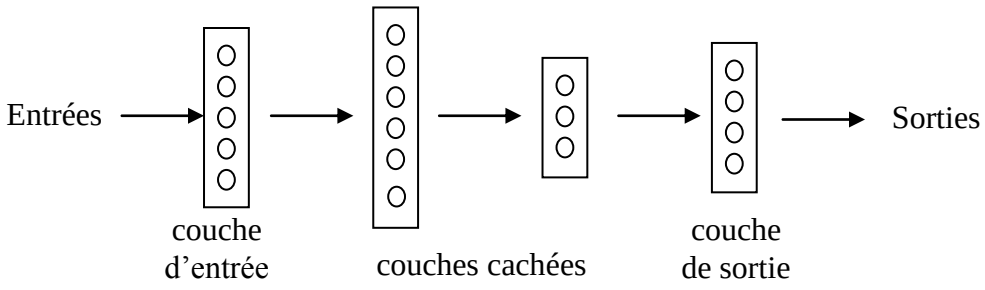


**Figure 3 :** Exemple de perceptron multicouche
*Une flèche entre deux couches signifie que toutes les unités d'une couche sont reliées à toutes les unités de l'autre couche.*

Les perceptrons multicouches, munis de l'algorithme de rétropropagation, sont d'excellents classificateurs : après apprentissage, ils peuvent fournir en sortie la classe à laquelle appartient une entrée donnée, même dans les cas difficiles (quand les frontières entre classes dans l'espace des entrées ont des formes compliquées). Plus généralement, on a pu montrer qu'ils pouvaient servir d'approximateurs universels de fonctions (Hornik *et al.*, 1989). Ils sont principalement utilisés pour simuler des tâches de reconnaissance perceptive et autres types de catégorisation cognitive.

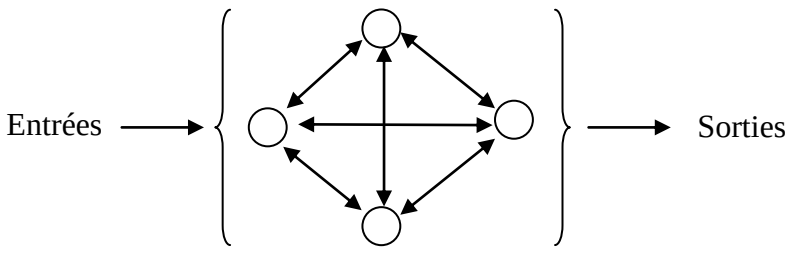

**Figure 4 :** Exemple de réseau entièrement récurrent
*Toutes les unités sont à la fois unités d'entrée et unités de sortie.*

Les réseaux entièrement récurrents ont été introduits en sciences cognitives par le physicien Hopfield (1982, 1984) à partir d'un modèle électromagnétique (les verres de spin). Ils sont utilisés notamment comme modèles de la mémoire, comme on le verra plus bas (§3.1).

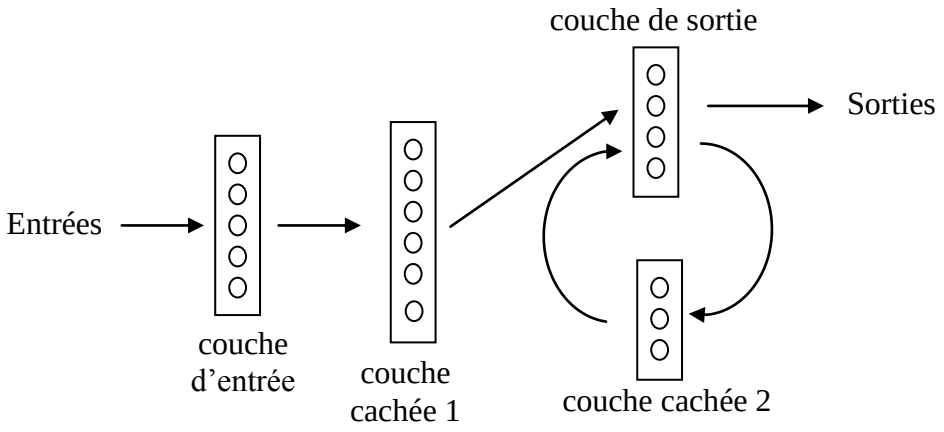


**Figure 5 :** Exemple de réseau simplement récurrent
*Il y a récurrence entre la couche de sortie et la couche cachée 2.*

La première architecture de réseau simplement récurrent est due à Elman (1990, 1991), qui a modélisé l'apprentissage de suites temporelles (en particulier des séquences grammaticales). Depuis d'autres chercheurs (cf., entre autres, Miikkulainen et Dyer 1991, St John et McClelland 1990, Victorri et Fuchs 1996) ont utilisé diverses architectures de ce type, notamment dans le domaine de la compréhension du langage.
A côté de ces architectures classiques, il existe bien d'autres types de réseaux, qui diffèrent grandement tant dans leurs architectures que dans leurs mécanismes de fonctionnement et d'apprentissage. On se contentera ici d'en citer quelques uns parmi les plus importants : les *cartes auto-organisatrices* de Kohonen (1994), le modèle de *darwinisme neuronal* de Edelman (1987), le modèle de *colonne corticale* de Burnod (1993), la théorie de la *résonance adaptative* (*ART*) de Grossberg (Carpenter et Grossberg 1991), et les *chaînes synchronisées* (*synfire chains*) d'Abeles (Abeles 1991, Abeles *et al.* 1994). Chacun d'entre eux mériterait en fait une présentation détaillée particulière pour rendre compte de ses spécificités.

## Quelques exemples de modèles

Comment les réseaux connexionnistes sont-ils utilisés en neurophysiologie, et à quoi peut servir ce type de modélisation ? Pour répondre concrètement à cette question, nous allons présenter succinctement trois modèles, assez typiques de ce qui se fait dans ce domaine.

## 1. Un modèle distribué de la mémoire

Dans le livre fondateur du groupe PDP, McCleland et Rumelhart (1986a) présentent un modèle connexionniste de la mémoire, dont la principale caractéristique est d'être *distribué* : chaque information mémorisée n'est pas localisée dans un élément précis du système, mais elle est répartie dans tout le système, tous les éléments servant à coder simultanément toutes les informations. Pour cela, ils utilisent un réseau entièrement récurrent (cf. §2.3 fig. 4) avec un mécanisme d'apprentissage non supervisé, proche de la règle de Hebb (pour une présentation et une discussion plus détaillées de ce modèle, cf. Victorri 1996).

Un vecteur d'entrée de ce réseau correspond dans leur modèle à une forme présentée au système, que celui-ci doit mémoriser. Le codage de ces formes est, comme nous l'avons dit, distribué : chaque composante d'un vecteur vaut +1 ou –1, et ce n'est qu'en comparant globalement les vecteurs que l'on peut distinguer les formes correspondantes. Par exemple, chaque forme pourrait être une image visuelle, le vecteur d'entrée codant un certain nombre de caractéristiques de cette image (+1 si la caractéristique est présente, -1 sinon). Le vecteur de sortie, obtenu après stabilisation du réseau, est interprété comme la réponse du système à la présentation d'une forme donnée. Un protocole d'expérimentation consiste à présenter une série de formes au réseau, en modifiant ses poids après chaque présentation à l'aide de la règle d'apprentissage. On peut alors tester ce que le réseau a mémorisé en lui présentant par exemple une forme incomplète et en regardant si le réseau est capable de restituer en réponse la forme complète correspondante.

Grâce à ces expérimentations, McCleland et Rumelhart ont pu montrer qu'un tel réseau était effectivement capable de mémoriser plusieurs formes différentes, apportant ainsi la preuve qu'un système entièrement distribué pouvait constituer un modèle plausible de la mémoire, exhibant d'intéressantes propriétés d'un point de vue cognitif (en particulier dans l'apprentissage de formes prototypiques). Ce type de réseau, muni de mécanismes de fonctionnement et d'apprentissage légèrement différents, a d'ailleurs fait l'objet de nombreuses études mathématiques (Amit 1989, Gardner 1988, Tsodyks et Feigel'Man 1988), qui ont permis notamment de calculer leur capacité théorique maximale en terme de stockage mémoriel.

McCleland et Rumelhart (1986b) ont aussi innové en examinant les propriétés de ces réseaux quand on les soumet à un processus de dégradation (par exemple en réduisant sensiblement un paramètre essentiel du mécanisme d'apprentissage). L'idée était d'explorer par ce biais la possibilité de modéliser des comportements qu'on peut observer dans diverses formes d'amnésie. Même si leur propres résultats ne sont pas toujours probants (Victorri 1996), ils ont ainsi ouvert la voie à toute une série de travaux sur la modélisation de pathologies à l'aide de réseaux connexionnistes : entre autres exemples, on peut citer l'intéressante modélisation de la maladie d'Alzheimer par Horn *et al.* (1993), qui ont utilisé des données neurophysiologiques sur cette maladie pour construire des réseaux dont le comportement reproduit les différents types cliniques observés (dégradation lente ou au contraire chute brutale des performances).

## 2. Un modèle de la dyslexie profonde

Un autre exemple de modélisation de pathologie est fourni par le travail de Hinton et Shallice (1991). Ces auteurs ont conçu un réseau simplement récurrent, dont l'architecture est exactement celle qui a été présentée ici fig. 5 (§2.3), pour simuler le processus de lecture. Les vecteurs d'entrée correspondent à la forme écrite de mots, et les vecteurs de sortie représentent des ensembles de traits sémantiques correspondant au sens de ces mots. Un apprentissage supervisé permet au système d'associer correctement les entrées et les sorties pour toute une série de mots. On détériore alors une région précise du réseau : soit une partie des connexions entre la couche d'entrée et la première couche cachée (ce qui correspond à

une lésion périphérique de la partie du système chargée de la reconnaissance visuelle des mots), soit une partie des connexions plus en aval (représentant les centres de traitement sémantique). On observe alors les erreurs produites par le réseau ainsi détérioré.
Les résultats obtenus sont intéressants, parce que quelque peu contre-intuitifs. En effet, il n'y a pas de corrélation entre la région détériorée et le type d'erreurs du réseau. Dans tous les cas de figure, on observe des erreurs de type sémantique (par exemple le mot présenté est *cat* et le sens donné en réponse est celui de *mice*), des erreurs de type visuel (le mot présenté est *mat* et le sens associé est celui de *cat*), et une forte proportion d'erreurs « mixtes » (mot présenté *rat* et sens associé *cat*). On aurait pu s'attendre à ce que les détériorations périphériques conduisent à plus d'erreurs de type visuel, et les détériorations plus centrales à plus d'erreurs de type sémantique, mais ce n'est absolument pas le cas. Il y a bien des réseaux qui font plus d'erreurs visuelles et d'autre plus d'erreurs sémantiques, mais cela est dû au caractère aléatoire de la sélection des connexions supprimées dans la région choisie pour être détériorée, et pas de la localisation de cette région.
Ce résultat est très important parce que, comme le note lui-même Shallice (1991), il remet en question des interprétations trop hâtives de doubles dissociations[2], qui constituent un paradigme fondamental de la neuropsychologie (Eustache et Faure, 1996). En effet, si l'on considère « de l'extérieur » le comportement d'un ensemble de réseaux détériorés, on relève des cas qui s'apparentent à de la double dissociation entre erreurs de reconnaissance visuelle et erreurs de traitement sémantique, et l'on serait tenté d'en déduire que ces cas correspondent à des détériorations de sous-systèmes différents du processus de lecture. Le modèle apporte la preuve que, dans ce cas tout au moins, une telle déduction serait erronée.

### 3. Un modèle de représentation distribuée de catégories lexicales

Les travaux de Small *et al.* (1995) portent sur un autre problème important pour la neuropsychologie. Il s'agit de l'organisation de la mémoire sémantique. L'objectif des auteurs était de montrer qu'une représentation entièrement distribuée est compatible avec une structuration du lexique en classes sémantiques, du type humains, animaux, outils, fruits et légumes, moyens de transport, etc., telle qu'elle est mise en évidence par un certain nombre de données psycholinguistiques et neuropsychologiques.
Les auteurs ont défini près de 80 « traits sémantiques » (couleur, forme, fonction, emplacement, etc.) pour caractériser une soixantaine d'objets, chaque objet étant défini sémantiquement par un vecteur indiquant la présence ou l'absence de chacun des traits sémantiques en question. Deux types de réseaux unidirectionnels (cf. §2.3, fig.3) ont été utilisés. Le premier est un réseau extrêmement simple à deux couches (perceptron sans couche cachée), prenant en entrée les vecteurs de traits sémantiques, chaque unité de sortie étant dédiée à un objet différent. L'apprentissage consiste à associer à la représentation sémantique d'un objet l'unité de sortie correspondante. En examinant les poids du réseau après apprentissage, les auteurs ont pu montré que les objets étaient regroupés en classes, toutes les unités de sortie correspondant par exemple à des outils ayant des poids très proches les uns des autres. Ainsi, bien que ces classes n'aient à aucun moment été données explicitement au système, elles ont émergé au cours du processus d'apprentissage, prouvant ainsi qu'un lexique structuré peut résulter d'un processus très simple d'apprentissage sur une représentation distribuée.
Le deuxième type de réseau mis en œuvre conforte ce premier résultat. C'est un réseau un peu plus complexe, que l'on appelle *réseau auto-associatif*. Il comporte une couche d'entrée et une couche de sortie identiques, codant toutes les deux les vecteurs de traits sémantiques, et deux couches cachées, beaucoup plus petites (la deuxième couche cachée ne comporte que 12

[2] Voir aussi l'article de Juola et Plunkett (1998), au titre évocateur : *Why double dissociations don't mean much.*

unités). L'apprentissage consiste à redonner en sortie le même vecteur que celui qui a été présenté en entrée : ainsi, à la fin de l'apprentissage, le réseau a réalisé une *compression des données*, puisque les 12 valeurs de la deuxième couche cachée suffisent à redonner en sortie les valeurs des quelques 80 traits sémantiques associés aux objets. L'étude du codage compact ainsi obtenu (analyse en composantes principales des vecteurs de la couche cachée) fait clairement apparaître les mêmes classes sémantiques que dans la première expérience, montrant ainsi la robustesse de l'émergence de cette structuration du lexique.

## Conclusion

Ainsi, l'intérêt essentiel des modèles connexionnistes en neuropsychologie réside dans leur capacité à tester, grâce à des systèmes informatiques relativement simples, le bien-fondé d'un certain nombre d'hypothèses sur le fonctionnement du système cognitif. Il est clair que ces modèles ne sont pas réalistes, au sens où ils ne peuvent prétendre simuler de manière fidèle ni la complexité des structures neuronales du cerveau, ni la complexité des activités cognitives qu'elles sous-tendent. Mais ils sont très utiles parce qu'ils mettent en évidence des mécanismes de fonctionnement et d'apprentissage d'un type nouveau, qui sont une précieuse source d'inspiration pour élaborer des modèles théoriques du fonctionnement du cerveau et de l'esprit. Comme on l'a vu sur quelques exemples, ils permettent souvent d'exhiber des systèmes dont le comportement peut paraître a priori contre-intuitif, et qui sont autant de contre exemples à des interprétations classiques en neuropsychologie. Cette utilisation du connexionnisme fait donc partie de ce que l'on peut appeler de *l'épistémologie expérimentale* : elle permet de tester, grâce à ces petits modèles informatiques, de la validité de raisonnements inductifs qui conduisent de l'observation expérimentale de comportements normaux ou pathologiques à la mise en place d'hypothèses théoriques sur l'architecture et le fonctionnement de la cognition.

## Bibliographie


Abeles M. (1991), *Corticonics : Neural circuits of the cerebral cortex*, Cambridge University Press.

Abeles M., Prut Y., Bergman H. et Vaadia E. (1994), Synchronization in neuronal transmission and its importance for information processing, in Buzsaki G., Llinas R, Singer W., Berthoz A. et Christen Y. (éds.) *Temporal coding in the brain*, Berlin, Springer-Verlag, 39-50.

Amit D.J. (1989), *Modeling Brain Function : The World of Attractor Neural Networks*, Cambridge University Press.

Burnod Y. (1993), *An adaptive neural network : the cerebral cortex*, Paris, Masson.

Carpenter G.A. et Grossberg S. (1991), *Pattern recognition by self-organization neural networks*, Cambridge, MIT Press.

Edelman G.M. (1987), *Neural Darwinism : The Theory of Neuronal Group Selection*, Nex York, Basic Books.

Elman J.L. (1990), Finding structure in time, *Cognitive Science*, 14, 179-211.

Elman J.L. (1991), Distributed representations, simple recurrent networks and grammatical structure, *Machine Learning*, 7, 195-224.

Eustache F. et Faure S. (1996), *Manuel de Neuropsychologie*, Paris, Dunod.

Gardner E. (1988), The space of interactions in neural network models, *journal of Physics*, 21A, 257.

Hebb D.O. (1949), *The Organization of Behavior*, New York, Wiley.

Hinton G.E. et Shallice T. (1991), Lesioning an attractor network : investigations of acquired dyslexia, *Psychological Review*, 98.

Hopfield J.J. (1982), Neural networks physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences*, 79, 2554-58.

Hopfield J.J. (1984), Neurons with graded responses have collective computational properties like those of two-states neurons, *Proceedings of the National Academy of Sciences*, 81, 3088-92.

Horn D., Ruppin E., Usher M., Herrmann M. (1993), Neural network modeling of memory deterioration in Alzheimer's disease, *Neural Computation*, 5, 736-749.

Hornik K., Stinchcombe M. et White H. (1989), Multilayer Feedforward Networks Are Universal Approximators, *Neural Networks*, 2, 359-366.

Jodouin J.-F. (1994), 2 vol. : *Les réseaux de neurones, principes et définitions* et *Les réseaux neuromimétiques*, Paris, Hermès.

Juola P., Plunkett K (1998), Why Double Dissociations Don't Mean Much, *Proceedings of the 20th Annual Conference of the Cognitive Science Society (CogSci-98).*

Kohonen T. (1994), *Self-Organizing Maps*, Berlin, Springer.

Le Cun Y. (1986), Learning Process in Asymetric Threshold Network, in Bienenstock E., Fogelman F. et Weisbuch G. (éds.), *Disordered Systems and biological Organization*, Berlin, Springer-Verlag.

McCleland J.L., Rumelhart D.E. (1986a), A distributed model of human learning and memory, in McCleland, Rumelhart *et al.* (1986), 170-215.

McCleland J.L., Rumelhart D.E. (1986b), Amnesia and distributed memory, in McCleland, Rumelhart *et al.* (1986), 503-528.

McCleland J.L., Rumelhart D.E. and the PDP Research Group (1986), *Parallel Distributed Processing : Explorations in the Microstructures of Cognition*, 2 vol., Cambridge, MIT Press.

Miikkulainen R. et Dyer M.G. (1991), Natural language processing with modular PDP networks and distributed lexicon, *Cognitive Science*, 15, 343-400.

Minsky M.L. et Papert S.A. (1969), *Perceptrons*, Cambridge, MIT Press.

Nadal J.-P. (1993), *Réseaux de neurones, de la physique à la psychologie*, Paris, Armand Colin.

Parker D. (1985), Learning Logic, Technical Report TR-87, Cambridge, MIT.

Rosenblatt F. (1962), *Principles of Neurodynamics : Perceptrons and the Theory of Brain Mechanisms*, New York, Spartan Books.

Rumelhart D.E., Hinton G.E. et Williams R.J. (1986), Learning Internal Representations by Error Propagation, in McCleland et Rumelhart (1986a).

Shallice T. (1991), Précis of *From neuropsychology to mental structures*, *Behavioral and Brain Sciences*, 14, 429-469.

Small S.L., Hart J., Nguyen T., Gordon B. (1995), Distributed representations of semantic knowledge in the brain, *Brain*, 118, 441-453.

St John M.F. et McClelland J.L. (1990), Learning and applying contextual constraints in sentence comprehension, *Artificial Intelligence*, 46, 217-256.

Tsodyks M.V. et Feigel'Man M.V. (1988), The enhanced storage capacity in neural networks with low activity level, *Europhysical Letters*, 46, 101-105.

Victorri B. (1996), Modèles connexionnistes de la mémoire, in Eustache F., Lechevalier B. et Viader F., *La mémoire : Neuropsychologie clinique et modèles cognitifs*, Bruxelles, De Boeck, 371-387.

Victorri B. et Fuchs C. (1996), *La polysémie : Construction dynamique du sens*, Paris, Hermès.