II.4. Plausibilité cognitive d'un appareil conceptuel autonome

1. Introduction

Nous proposons ici un modèle neuronal simplifié de la détection d'incohérences. Ce modèle n'a pas encore donné lieu à des implantations, mais notre objectif est de montrer que les traitements symboliques ne sont pas nécessairement étrangers à toute modélisation biologiquement plausible.

Le modèle fait intervenir un ensemble particulier de neurones, la *couche "*$\mathcal{C}$*"*, spécialisé dans le traitement des concepts logiques. Cette couche implante les concepts logiques suivant un schéma ensembliste, et permet la détection des situations incohérentes, caractérisées par l'activation simultanée de représentations conceptuelles incompatibles.

Ce schéma permet de donner un début de réponse satisfaisant à plusieurs problèmes soulevés indépendamment : problème de la systématicité et de l'unicité matérielle des processus conceptuels, problème de la compositionnalité et de la constituence, problème de la dualité extension / intension.

2. Traitement symbolique vs. traitement connexionniste

Les mécanismes logiques sont peut-être, parmi les mécanismes symboliques, ceux qui sont considérés comme les plus éloignés de ce que les modèles connexionnistes habituels sont capables de reproduire. L'une des rares manières d'approcher la performance des traitements logiques avec une architecture connexionniste consiste à remplacer les implications logiques [$p \Rightarrow q$] par des inférences statistiques [Smolensky 1988]. Nous avons montré ailleurs (voir chapitre II.2) à quel point une telle approximation de la performance logique est non satisfaisante. Rappelons simplement que l'inférence statistique est symétrique (corrélation), alors que l'implication ne l'est pas ; qu'elle exige un échantillon important, alors que l'implication peut être formée sans expérience du domaine (cf. chap. II.2) ; de plus la reconnaissance des incompatibilités logiques (par ex. [*p & non q*]), performance humaine quotidienne, est distincte d'un traitement statistique, puisqu'elle peut s'exercer sur des situations indépendamment de leur fréquence.

Or la systématicité de certains processus logiques nous a amenés à postuler l'existence d'un appareil conceptuel autonome (voir chapitre II.3). Cet appareil est supposé héberger les représentations matérielles de tous les concepts. Comment imaginer un fonctionnement neuronal pour un tel appareil, qui serait capable d'effectuer des opérations logiques comme la reconnaissance d'incohérences ?

L'exigence de systématicité a parfois été perçue, à tort, comme incompatible avec une implantation neuronale. Par exemple l'argument donné par J. Fodor et Z. Pylyshyn [1988], qui repose en grande partie sur la systématicité des traitements linguistiques et logiques, est souvent cité pour opposer traitement symbolique et implantation connexionniste, alors qu'il consiste simplement à imposer des contraintes à une telle implantation, contraintes auxquelles certaines architectures connexionnistes semblent pouvoir satisfaire [Chalmers 1990].

Les modèles qui cherchent à "réconcilier" les contraintes liées à l'approche connexionniste et celles liées à l'approche symbolique sont de deux types : d'une part les modèles ascendants ou homogènes, qui cherchent à faire émerger les symboles de fonctionnements sub-symboliques [Grumbach 1990 ; Hofstadter & McGraw 1993], et d'autre part les approches descendantes ou hétérogènes, qui garantissent une autonomie au niveau symbolique et accordent simplement un ancrage aux symboles [Harnad 1990].

Les approches hétérogènes (descendantes) sont souvent critiquées pour deux raisons. Premièrement parce que leurs implantations sont elles-mêmes hétérogènes : comme nous venons de le voir, le fonctionnement symbolique n'a pas de correspondant plausible dans ce que les architectures connexionnistes proposent couramment. Or celles-ci sont généralement reconnues comme plus proches du fonctionnement neuronal naturel.

La deuxième raison vient du problème de l'ancrage. Comment obtenir que les symboles du niveau supérieur aient une signification intrinsèque fournie par le niveau inférieur ?

Nous avons traité cette question de l'ancrage dans le cas des concepts (chap. II.3). Nous avons décrit les concepts comme des détecteurs, nous avons défini l'ancrage d'un concept par les objets que ce concept peut détecter dans une situation donnée, et nous avons décrit plusieurs mécanismes par lesquels de nouveaux concepts pouvaient être formés pour détecter de nouveaux objets.

Pour justifier la plausibilité de notre modèle, qui est résolument hétérogène, il nous faut donc maintenant suggérer une architecture vraisemblable pour le traitement symbolique des concepts. C'est ce que nous nous proposons de faire pour un cas particulier, qui jouit pour nous d'une grande importance, celui de la détection des incohérences.

3. Implémentation neuronale de la détection

Dans le chapitre II.3, nous avons présenté les concepts comme des détecteurs, au sens fonctionnel, mais en laissant entendre qu'ils étaient réalisés matériellement par des détecteurs neuronaux. Notre premier objectif ici va donc être de proposer une implantation neuronale pour les détecteurs (qu'ils interviennent dans le traitement conceptuel ou perceptif).

Dans [Dessalles 1993b], nous avons proposé une modélisation neuronale du phénomène de détection. Ce phénomène, que nous invoquons ici pour caractériser les concepts, intervient aussi dès les premiers niveaux de la perception et opère selon nous à chaque changement de niveau d'intégration. Nous avons donné des arguments pour suggérer le fait que le phénomène de détection ne peut pas être le fait de neurones uniques. Nous avons donc proposé un modèle qui décrit les détecteurs comme des ensembles localisés de neurones susceptibles d'entrer en activité synchrone, et nous avons qualifié ces détecteurs de *collectifs*, dans la mesure où aucune information n'est détenue par un seul neurone(1). Nous reproduisons ici une partie de la description de ces détecteurs collectifs.

3.1. Détection collective et cohérence temporelle

Nous partons de l'axiome suivant, dicté par un souci de plausibilité biologique : aucune fonction cognitive ne peut être supportée par un seul neurone. Ce sera en particulier le cas de la fonction de détection.

Tout modèle plausible du fonctionnement d'un détecteur se doit par conséquent de respecter cette contrainte de non-centralisation de l'information détectée. En particulier, l'activité du détecteur doit être portée

---

[1] Le mot *collectif* est plus fort que le mot *distribué*. Dans un système collectif, les neurones individuels ne sont pas porteurs d'information, contrairement aux systèmes distribués, comme les réseaux connexionnistes habituels, dans lesquels chaque neurone détient une partie de l'information (fournie ou apprise).

par un phénomène physique défini au niveau *collectif* et dépourvu de signification au niveau d'un composant du détecteur. Il en est de même pour le système physique (connexion du détecteur) qui provoque l'activation : il doit lui aussi rester collectif.

Si un détecteur collectif est réalisé par un ensemble de neurones, par quoi se traduit son activation ? Pour quelques raisons que nous allons expliciter, nous suggérons le fait que cette activation peut être caractérisée par une activité cohérente des neurones qui composent le détecteur.

Dans le cas des neurones naturels, la cohérence se traduit par une émission synchronisée de potentiels d'action. Les neurones émettent spontanément des potentiels d'action, mais généralement sans coordination avec d'autres neurones. Nous définissons donc un détecteur comme un ensemble de neurones susceptibles de se synchroniser entre eux : le détecteur est actif si les neurones sont synchrones, et inactif si l'activité des neurones est désordonnée.

La cohérence temporelle est un bon candidat pour caractériser l'activation d'un détecteur collectif neuronal pour au moins trois raisons.

Premièrement, il s'agit d'une relation simple entre les neurones, et nous sommes "en droit" d'attendre une relation simple. En effet, le problème étant symétrique par nature, nous pouvons exiger du motif d'activation qu'il soit lui-même symétrique, ou stochastiquement symétrique (*i.e.* rôle statistiquement équivalent pour chaque neurone). La synchronisation est l'une des relations symétriques les plus simples que l'on puisse imaginer. L'avantage d'une relation simple est que son déclenchement pourra être simple lui aussi. On comprend aisément comment, dans une hiérarchie d'intégration perceptive, la synchronisation des neurones d'un détecteur de niveau donné *n* peut être obtenue par synchronisation avec les neurones afférents (de niveau *n-1*), autrement dit par l'action des détecteurs afférents actifs au niveau *n-1*. Nous détaillerons cet aspect plus loin.

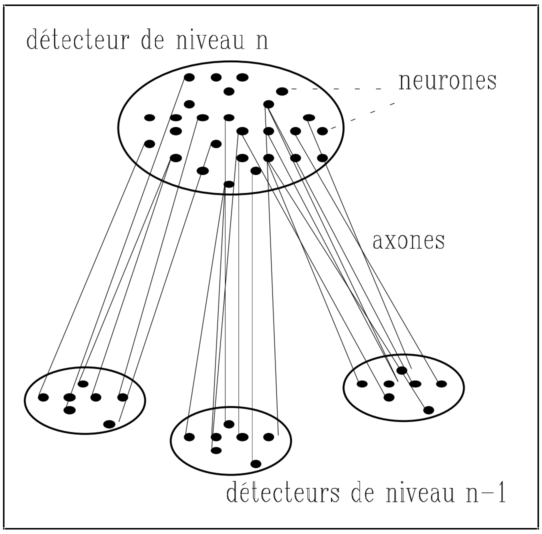


figure 1 : implémentation neuronale de deux niveaux de détection collective. Les neurones d'un détecteur reçoivent des synapses excitatrices de leurs voisins (connexions latérales), ainsi que des synapses excitatrices en provenance de neurones appartenant à des détecteurs de niveau inférieur (ou à des capteurs). Si peu de détecteurs de niveau *n-1* sont actifs de manière synchrone, peu de neurones du détecteur de niveau *n* peuvent sommer une activité suffisante, et cette activité "se perd" dans les connexions latérales. En revanche, l'activité d'un nombre suffisant de détecteurs de niveau *n-1* permet au détecteur de niveau *n* de "s'embraser" : ses neurones deviennent synchrones.

Le deuxième argument vient de ce que la synchronisation neuronale est biologiquement vraisemblable. Elle est observée par exemple dans le cortex visuel, où l'on vérifie que les neurones répondant à une même zone du champ visuel, mais situés dans des aires différentes (comme l'aire V4 qui traite la couleur et l'aire V5 (aire MT) où les mouvements sont détectés), deviennent synchrones lorsqu'ils répondent au même objet du champ visuel [Zeki 1992](2). Le même phénomène a été observé dans des zones sous-corticales comme le bulbe olfactif [Freeman 1991].

La troisième raison qui nous pousse à considérer la synchronisation comme un phénomène plausible pour marquer le résultat d'une détection collective vient de notre connaissance de mécanismes inter-neuronaux qui la rendent possible. Par exemple, une stabilisation des synapses qui suit la loi de Hebb peut rendre un réseau de neurones artificiels *cohérent* : deux neurones d'un réseau cohérent seront soit toujours simultanément actifs, soit jamais actifs en même temps.

Il reste à expliquer comment l'activation est transmise d'un niveau à l'autre, et grâce à quel câblage. Il s'agit ensuite de se demander comment un tel câblage peut s'établir.

3.2. Déclenchement d'une détection collective neuronale

Nous sommes confrontés à la question de savoir comment l'activité d'un détecteur collectif peut être déclenchée par l'activité de détecteurs collectifs du niveau inférieur, autrement dit nous devons préciser le type de câblage, au niveau neuronal, qui permet "d'alimenter" les détecteurs collectifs. Plus précisément, dans notre optique, il s'agit de comprendre comment l'activité synchrone de neurones de niveau *n-1* peut déclencher l'activité d'un groupe de neurones donné au niveau *n*. Nous devons en quelque sorte définir ce qui caractérise les niveaux d'un point de vue structurel.

3.2.1. Des détecteurs collectifs cohérents

L'hypothèse la plus simple que nous pouvons faire semble être qu'un détecteur collectif est formé d'un ensemble "architecturalement cohérent" de neurones : les connexions internes au détecteur sont excitatrices, et les connexions avec des neurones externes (*i.e.* les neurones de la même couche, mais externes au détecteur) sont inexistantes(3). Il peut s'ensuivre un phénomène de seuil, caractéristique de tout détecteur : si très peu de neurones du détecteur sont actifs de manière synchrone, alors les excitations qu'ils propagent ont peu de chance de se rencontrer en entrée d'autres neurones du détecteur pour y être sommées. Les neurones restent donc insensibles à l'activité de leurs voisins, et le détecteur garde une activité désordonnée. En revanche, sous l'effet d'une excitation synchrone extérieure, un nombre suffisant de neurones du détecteur peuvent devenir actifs et ainsi "embraser" le détecteur (cf. figure 1).

Un tel schéma suppose l'existence d'un câblage excitateur interne au détecteur, et d'un câblage excitateur en provenance des détecteurs afférents. La question se pose alors de savoir comment un tel câblage peut s'établir. Cette question est essentielle dans les systèmes naturels, pour lesquels il faut résoudre le contraste entre l'apparente complexité du câblage (10^{15} synapses) et la relative concision de son codage génétique (il y a moins de 10^5 gènes au total, et seule une petite partie de ces gènes a une influence sur la spécificité du câblage neuronal). Nous abordons maintenant cette question du câblage.

3.2.2. Des détecteurs collectifs pseudo-locaux

Sans prétendre fournir une réponse précise, nous pouvons rappeler ici quelques éléments de réponse souvent mentionnés dans des contextes similaires. Le câblage excitateur interne peut être associé à la topologie des détecteurs. Si nous supposons que les détecteurs sont *pseudo-locaux*, c'est-à-dire qu'ils sont constitués non pas d'un seul neurone (hypothèse localiste extrême), mais de neurones proches, alors un système de connexions latérales rappelant le "chapeau mexicain" de T. Kohonen [1984] peut définir un ensemble de détecteurs collectifs avec un coût de codage architectural très faible. De plus, Lumer et Huberman [1992] ont montré qu'un tel ensemble de neurones (chez ces auteurs, il s'agit d'oscillateurs connectés entre eux avec des valeurs de couplage qui décroissent exponentiellement avec la distance de connexion) pouvait se synchroniser.

---

[2] Imaginer un petit objet rouge en mouvement horizontal. Les neurones détectant, dans cette zone du champ visuel, la couleur rouge (dans l'aire V4) et le mouvement horizontal (dans l'aire V5) seront synchrones entre eux.

[3] Des connexions externes inhibitrices permettraient d'augmenter le contraste, comme le note T. Kohonen [1984].

De plus, le câblage excitateur afférent peut être lui aussi réalisé avec un faible coût grâce aux propriétés topologiques de tels détecteurs collectifs pseudo-locaux. Il suffit que des connexions en provenance d'un détecteur afférent se projettent à l'intérieur du détecteur cible. La mise au point fine peut être assurée par une sélection synaptique. Seules sont conservées les synapses qui sont fonctionnelles au sens de Hebb : lorsque tous les détecteurs afférents, ou une grande partie d'entre eux, sont actifs, le détecteur cible devient actif aussi, même si le câblage est grossier. Les synapses qui favorisent cette synchronisation sont alors stabilisées. En revanche les synapses qui se perdent dans des détecteurs voisins, non actifs, dégénèrent. Une telle stabilisation peut s'opérer sans apport d'information extérieure, par exemple grâce à des générateurs aléatoires qui respectent la topologie [Shatz 1992](4). Le coût de codage d'une telle architecture reste faible.

4. Modélisation neuronale de la détection d'incohérence

Nous proposons de considérer que la détection des incohérences, qui est une compétence humaine facile à mettre en évidence par l'observation des conversations, peut, malgré son caractère éminemment symbolique, être réalisée par un mécanisme neuronal plausible. Selon notre définition, une incohérence relie des concepts logiques. Nous suggérons le fait que tous les concepts logiques connus ou potentiellement connus d'un individu sont hébergés dans une structure unique que nous dénommons "appareil conceptuel". C'est cet appareil qui, dans notre modélisation, réalise la détection d'incohérences.

4.1. Architecture de l'appareil conceptuel

Nous allons tenter ici de suggérer un fonctionnement plausible pour un "appareil conceptuel" autonome. Nous avons décrit les concepts comme des détecteurs. Nous considérons maintenant qu'ils sont *réalisés* par des détecteurs neuronaux collectifs pseudo-locaux. Ces détecteurs sont reliés physiquement aux étages intégrés des appareils perceptifs par des faisceaux d'axones comme schématisé sur la figure 1. Leur entrée en activité dépend donc des perceptions.

Nous supposons que l'appareil conceptuel consiste en un ensemble de neurones impliqués dans des détecteurs collectifs dans lesquels les concepts sont implantés. Pour fixer les idées, imaginons cet ensemble bidimensionnel et appelons-le *couche* $\mathcal{C}$ (figure 2). Les connexions internes à la couche $\mathcal{C}$ sont excitatrices au sein des détecteurs (que nous supposons pseudo-locaux, au sens défini plus haut).

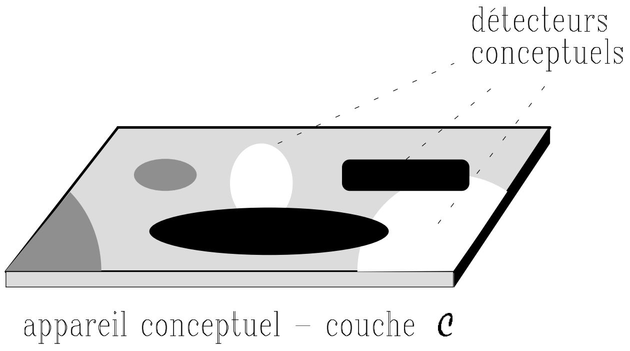


figure 2 : représentation imagée de la couche "$\mathcal{C}$ ". Les représentations conceptuelles sont constituées d'ensembles de neurones connectés par des synapses excitatrices, qui constituent des détecteurs. Tous les neurones de la couche $\mathcal{C}$ sont impliqués dans de tels détecteurs. Cette couche est reliée aux représentations intégrées de la perception et du langage. La représentation d'un concept actif *dans un contexte donné* est traduite par l'activité synchrone d'un ou plusieurs de ces détecteurs.

---

[4] Shatz a observé de tels générateurs aléatoires dans des rétines d'embryons.

L'appareil conceptuel tel que nous le décrivons héberge donc les représentations conceptuelles (détecteurs pseudo-locaux). Il reçoit des faisceaux de connexion en provenance des appareils perceptifs. Mais nous devons prendre en compte aussi le fait que les concepts sont intimement liés aux mécanismes linguistiques, et qu'ils sont capables d'évoquer des perceptions. On est donc en droit d'imaginer les connexions correspondantes. Toutefois, nous n'en parlerons pas ici, car ces aspects n'entrent pas dans le cadre de la modélisation partielle que nous décrivons.

4.2. Principe de fonctionnement de la couche "$\mathcal{C}$"

Chaque situation perçue va activer un certain nombre de détecteurs de la couche $\mathcal{C}$. Les neurones de ces détecteurs deviennent actifs de manière synchrone. Cette synchronisation est liée à la perception de la situation : il n'y a qu'une seule synchronisation possible à un moment donné, autrement dit la couche $\mathcal{C}$ ne traite qu'une seule situation à la fois. Les neurones n'appartenant pas aux détecteurs activés par la situation ont une activité désordonnée.

> *L'agencement des détecteurs est tel qu'une incohérence est détectée lorsque tous les neurones de la couche $\mathcal{C}$ sont synchronisés. La couche $\mathcal{C}$ elle-meme se comporte alors comme un détecteur activé.*

La figure 2 montre de manière très schématique ce que peut être la couche $\mathcal{C}$. Les détecteurs conceptuels peuvent se recouvrir, et tout neurone de la couche appartient au moins à un détecteur. Noter qu'un neurone donné appartenant à un détecteur non activé peut néanmoins être en phase avec la synchronisation de référence : il suffit qu'il appartienne aussi à un détecteur activé. Ce raisonnement est valable pour tous les neurones du détecteur. Ce dernier peut donc se retrouver actif en quelque sorte "malgré lui", c.à.d. sans que son activité ait été provoquée par ses propres connexions afférentes (en provenance des systèmes perceptifs).

4.3. Principe du câblage de la couche "$\mathcal{C}$"

L'agencement des détecteurs permet une représentation directe des relations logiques. Si plusieurs concepts sont simultanément détectés dans une situation, les neurones de tous ces détecteurs dans la couche $\mathcal{C}$ seront synchrones. Nous constatons donc que ce qui correspond à la conjonction logique (présence simultanée) est représenté par la *réunion* des détecteurs correspondant aux concepts conjoints.

Nous devons en fait attribuer à la couche $\mathcal{C}$ des propriétés duales de celles accordées généralement à l'extension des concepts : le "et" logique correspond à *l'union* des détecteurs correspondants ; l'implication $p \Rightarrow q$ est représentée par l'inclusion inverse des détecteurs $D_q \subset D_p$, et l'incohérence, par la couverture de toute la couche. De cette manière, et cela constitue l'élément central de notre modèle :

> *L'ensemble des neurones de la couche $\mathcal{C}$ encode les relations logiques grâce à un schéma ensembliste.*

relation logique entre concepts		*relation ensembliste entre détecteurs*	
conjonction	$r = p \,\&\, q$	union ensembliste	$D_r = D_p \cup D_q$
implication	$p \Rightarrow q$	inclusion ensembliste	$D_q \subset D_p$
incompatibilité	$[p \,\&\, q] \Rightarrow \mathbf{F}$	totalité	$D_p \cup D_q =$ couche $\mathcal{C}$

4.4. Formation des concepts dans la couche "$\mathcal{C}$"

L'agencement des détecteurs conceptuels dans la couche $\mathcal{C}$ doit être issu de l'encodage des relations logiques. Comment cela se peut-il ? Nous ne prétendons pas être en mesure de résoudre ce problème, mais simplement suggérer le fait qu'il n'est pas insoluble.

Nous nous plaçons dans le cadre d'un mécanisme d'appariement pour expliquer la formation des concepts (chap. II.3). Nous postulons donc l'existence, préalable à toute expérience, d'un certain nombre de détecteurs conceptuels, que nous appelons détecteurs primitifs(5). Leur agencement permet la perception de certaines incohérences, c'est-à-dire que la synchronisation de plusieurs d'entre eux entraîne la synchronisation de toute la couche $\mathcal{C}$.

Dans notre modèle, le câblage interne qui caractérise les détecteurs primitifs de la couche $\mathcal{C}$ ne se modifie pas après sa mise en place. Il est caractérisé par des synapses fortement excitatrices, il définit la topologie les détecteurs primitifs, et donc leurs relations d'incompatibilité. Nous devons donc nous demander comment sont codés les concepts non primitifs et leurs relations d'incompatibilité.

Les concepts formés par les humains semblent exprimables par des conjonctions de concepts connus [Johnson-Laird 1983:424; Medin et al. 1987; Richard 1990:68]. Si l'on remonte jusqu'aux concepts primitifs postulés dans notre modèle (ce qu'un individu est peut-être incapable de faire de manière explicite), tout concept c peut être défini en théorie comme une conjonction de concepts primitifs : $c = c_1 \;\&\; c_2 \;\&\; . . . \;\&\; c_n$. Les neurones de la couche $\mathcal{C}$ vont traduire cette relation : le détecteur conceptuel D_c qui représente le concept c regroupe, au sens de la réunion ensembliste, les neurones impliqués dans les détecteurs primitifs D_{c_i}. Il nous faut donc caractériser matériellement les capacités de détection de D_c, ce qui revient à justifier l'existence d'un détecteur non primitif comme D_c en tant que groupe de neurones de la couche $\mathcal{C}$ capables de se synchroniser.

Nous avons vu qu'un détecteur pseudo-local devait son existence à un câblage excitateur interne et à un câblage afférent en provenance d'autres détecteurs. L'ensemble de neurones D_c possède en partie ce câblage excitateur interne. Toutefois D_c est hétérogène, formé de régions (les détecteurs primitifs D_{c_i} qu'il réunit) fortement connectées en interne, mais éventuellement non connectées entre elles. Une solution possible pour caractériser D_c consiste à postuler l'existence d'un *second câblage* de la couche $\mathcal{C}$. A côté du câblage de base aux synapses fixées sur des valeurs fortement excitatrices, nous supposons qu'il existe un deuxième câblage dont les synapses sont, elles, susceptibles de changer de valeur. D_c sera ainsi caractérisé par des synapses excitatrices de ce deuxième câblage, dont les valeurs sont issues de l'expérience selon un mécanisme à préciser, et qui relie des neurones de D_c (sans tenir compte des limites des détecteurs primitifs qui le composent)(6).

Nous appellerons ces détecteurs non primitifs, capables eux aussi de se synchroniser pour certaines configurations des perceptions intégrées, *détecteurs mémorisés*.

---

[5] Cette hypothèse concernant les primitives conceptuelles est parfois dénoncée comme non parcimonieuse. Il faut cependant bien comprendre qu'elle est indissociable de celle du mécanisme d'appariement. Rappelons que l'alternative, c'est-à-dire le refus de tout appariement, obligerait à considérer que les connaissances conceptuelles résultent d'un apprentissage de type "indifférent" (i.e. qui tolère tout changement de repère dans le codage de ses entrées). Nous avons montré qu'un mécanisme indifférent n'est pas assez efficace pour expliquer l'apprentissage des connaissances manipulées, par exemple, dans les conversations (chap. II.2). De plus, un mécanisme indifférent ne permet qu'un encodage *implicite* des connaissances, ce qui exclut toute manipulation logique (chap. II.3). Or les humains sont capables de manipulations logiques (*e.g.* le mécanisme de recevabilité d'une explication). L'abandon des systèmes indifférents nous conduit forcément à postuler l'existence de "connaissances" *a priori*. Dans le mécanisme d'appariement que nous proposons, ces connaissances prennent la forme de primitives conceptuelles.

[6] Noter que c'est ce deuxième câblage qui rend nécessaire, au niveau matériel, la centralisation physique de l'appareil conceptuel que nous avions postulée dans le chapitre II.3.

> **Détecteur primitif** : la couche $\mathcal{C}$ comporte, du fait de son câblage interne de base, un certain nombre de détecteurs, appelés détecteurs primitifs, capables de se synchroniser pour certaines configurations des perceptions intégrées. Tout neurone de la couche appartient à un tel détecteur. La structure de ces détecteurs primitifs, ainsi que leurs capacités de détection, sont supposées indépendantes de l'expérience de l'individu.
>
> **Détecteur mémorisé** : les perceptions intégrées relatives à certaines situations peuvent activer simultanément plusieurs détecteurs primitifs. Une telle configuration (réunion de détecteurs primitifs) peut être mémorisée, dans certaines conditions. Nous suggérons pour cela l'existence d'un câblage secondaire dans la couche $\mathcal{C}$ dont certaines synapses reliant des neurones de D_c seraient positionnées, dans les conditions de mémorisation et selon un mécanisme à définir, sur des valeurs excitatrices.

Les détecteurs mémorisés sont davantage que des réunions ensemblistes de détecteurs primitifs. Ce sont des représentations "structurées" (voir définition p.175), qui réunissent d'autres détecteurs mémorisés moins étendus.

Ce qui distingue l'établissement de D_c, en tant qu'ensemble de neurones susceptibles de se synchroniser, du type de mémorisation que l'on observe dans un réseau associatif habituel, c'est qu'il peut être réalisé en une fois. Nous ne sommes pas en mesure de suggérer un mécanisme détaillé de ce processus de mémorisation. Ce que nous suggérons, et que nous avons suggéré dans d'autres contextes (chap. II.3 et III.2), c'est que l'établissement d'un nouveau détecteur comme D_c , autrement dit la formation d'un nouveau concept *c*, se produit préférentiellement lorsqu'une incohérence a été détectée. Voyons comment.

Le fait que la couche $\mathcal{C}$ entre complètement en synchronisation (c.à.d. qu'elle devienne active exactement comme les détecteurs qui la composent) est utilisé comme signal pour une révision des concepts activés par la situation. Cette révision peut entraîner une création de concept ou une modification des concepts existants, à travers le mécanisme de l'*explication* que nous avons décrit ailleurs (voir chapitre III.4). Ce genre de phénomène, dans le contexte de cette modélisation neuronale, peut être décrit en 4 phases :

1- mise en synchronisation de nombreux détecteurs primitifs de la couche $\mathcal{C}$, due souvent à une perception sommaire (moins la perception est détaillée, plus les suppositions "par défaut" sont actives ; en termes connexionnistes, la généralisation au niveau des mécanismes perceptifs est plus importante dans le cas d'observations partielles) ;

2- détection d'une incohérence par entrée en synchronisation de tous les neurones de la couche $\mathcal{C}$;

3- révision de la perception (comme dans l'étape 1), l'attention plus soutenue conduisant à des informations plus complètes, et donc éventuellement à moins de détecteurs activés ;

4- mémorisation éventuelle de l'exception : un aspect de la perception avait été négligé dans la phase 1 ; du fait de sa prise en compte, la couche $\mathcal{C}$ n'est plus entièrement synchrone. Par différence avec la situation en -2-, un concept (groupe de détecteurs) peut être mémorisé ou redélimité, et ses liens (relation d'incompatibilité) avec les autres concepts actifs établis ou corrigés.

5. Points forts du modèle

5.1. Compatibilité avec les contraintes de la systématicité

L'une des principales caractéristiques de ce modèle est d'héberger en une même structure les représentations de tous les concepts. Ceci constitue, on s'en souvient (cf. chap. II.3), une exigence liée à la systématicité de certains processus conceptuels : systématicité des résultats, systématicité de prise en compte. Nous avions pris l'exemple de la recevabilité d'une explication, dont la systématicité exigeait qu'elle fût un processus matériellement unique (systématicité des résultats) et que sa réalisation matérielle opérât à partir d'un appareil conceptuel lui-même unique hébergeant toutes les représentations conceptuelles (systématicité de prise en compte).

Pour une incohérence que nous représentons de la façon suivante :

$$[p_1 \;\&\; p_2 \;\&\; .\,.\,.\; p_n] \Rightarrow \mathbf{F}$$

q est une tentative d'explication recevable dès que q = *non* p_i. Dans le modèle neuronal que nous venons d'esquisser, la couche $\mathcal{C}$ constitue ce lieu où toutes les représentations conceptuelles sont hébergées, et le mécanisme de recevabilité peut directement opérer sur la couche $\mathcal{C}$. Lorsque p_i cesse d'être perçu, la situation est révisée (au sens de l'étape -3- dans le scénario esquissé ci-dessus), et le détecteur correspondant au terme p_i a toutes les chances de n'être plus synchrone(7). Cela a pour effet de supprimer la perception de l'incohérence, car par définition p_i était nécessaire à l'incompatibilité. L'inactivation de p_i, dans le contexte de l'incohérence ci-dessus, *définit* le concept (*non* p_i) qui est ainsi reçu comme une tentative d'explication, puisque sa présence supprime l'incohérence initiale.

En revanche, il faut noter que l'évocation d'un concept r qui serait sans rapport avec l'incompatibilité n'a aucun effet sur l'état de la couche $\mathcal{C}$ qui reste entièrement synchrone : r ne sera pas reçu comme une tentative d'explication(8).

5.2. Constituence des représentations

Dans leur célèbre article, J. Fodor et Z. Pylyshyn [1988] insistent sur le fait que les représentations mentales, en particulier les concepts, doivent avoir un certain nombre de propriétés, dont nous allons mentionner certaines. Tout modèle de la cognition doit, selon ces auteurs, respecter ces propriétés. Nous voulons indiquer ici que notre modèle respecte les propriétés indiquées par Fodor et Pylyshyn. La principale qui nous intéresse ici est la propriété de *constituence*, qui concerne les représentations mentales structurées :

> **Représentations structurées** : Les éléments d'un ensemble de représentations mentales sont structurés si certaines de ces représentations sont atomiques, et que les autres sont moléculaires, *i.e.* résultent de la combinaison de composants qui sont eux-mêmes des représentations atomiques ou moléculaires.

Les concepts logiques, dans notre modèle, ont des représentations structurées. Les représentations atomiques sont les détecteurs primitifs, et les représentations moléculaires sont les détecteurs mémorisés, qui résultent de la réunion ensembliste d'autres détecteurs, primitifs ou mémorisés.

> **Constituence** : un ensemble de représentations structurées a la propriété de constituence si les composants des représentations sémantiquement interprétables de cet ensemble sont eux-mêmes sémantiquement interprétables.

Dans notre modèle, l'interprétation sémantique correspond aux capacités de détection. Un détecteur est mémorisé par la fixation de la co-activité de plusieurs détecteurs, dans des conditions particulières (consécutives à la détection d'une situation problématique, par ex. incohérente). Tout nouveau détecteur ainsi formé possède des capacités de détection, il est donc sémantiquement interprétable. Les détecteurs qui le composent, c.à.d. les détecteurs qui étaient activés au moment de sa mémorisation, sont aussi sémantiquement interprétables pour la même raison. La propriété de constituence des représentations est donc vérifiée dans notre modèle.

---

[7] Si ce n'est pas le cas, autrement dit si la perception continue à évoquer p_i alors que *non* p_i est évoqué par ailleurs (par ex. par le langage), il y a détection d'une autre incohérence par la couche C qui masque l'incohérence initiale.

[8] Ceci est vrai même si r est relatif à un élément de la situation : r=*non* s où s a été déclenché dans un premier temps. La désactivation de s n'affectera pas la détection de l'incohérence, si s n'était pas un élément de l'incompatibilité. Certes le détecteur D_s associé à s ne sera peut-être plus excité par les perceptions qui avaient provoqué son activation dans la première phase. Cependant, les neurones de D_s resteront synchrones avec tous les autres neurones de la couche $\mathcal{C}$, car ils appartiennent à d'autres détecteurs que D_s. Ainsi, s ne sera pas détecté comme une tentative d'explication, car la couche C restera entièrement synchrone.

Noter que la constituence est obtenue, dans notre modèle, par les propriétés d'inclusion matérielle des constituants dans les représentations, comme l'avaient plus ou moins prévu Fodor et Pylyshyn :

> [Properties like constituency] "constrain the physical realizations of symbol structures. In particular, the symbol structures in a Classical model are assumed to correspond to real physical structures in the brain and the combinatorial structure of a representation is supposed to have a counterpart in structural relations among physical properties of the brain. For example, the relation 'part of', which holds between a relatively simple symbol and a more complex one, is assumed to correspond to some physical relation among brain states." [Fodor & Pylyshyn 1988]

Nous pouvons extrapoler en disant que des représentations structurées sont nécessairement matériellement structurées :

représentations matériellement structurées : des représentations structurées sont matériellement structurées si elles contiennent (au sens de l'inclusion ensembliste) leurs composants.

Nous pouvons écrire symboliquement : constituence ⇒ structuration ⇒ structuration matérielle.

Le fait que notre modèle vérifie les contraintes identifiées par Fodor et Pylyshyn est tout sauf anodin. L'argument de Fodor et Pylyshyn vise principalement les architectures connexionnistes classiques qui, selon eux, ne respectent aucune des trois propriétés ci-dessus (ainsi que d'autres qui en découlent). Ils décrivent les architectures connexionnistes "orthodoxes" comme utilisant la transmission d'activité entre les noeuds d'un graphe. Ces noeuds, qui peuvent être des neurones ou des groupes de neurones, sont les unités sémantiques du système. Ces noeuds sont tous considérés comme sémantiquement atomiques, et l'architecture connexionniste ainsi décrite ne contient pas de représentation matériellement structurée. Elle ne peut donc pas héberger des représentations structurées, et vérifie encore moins la propriété de constituence (9).

Le modèle neuronal que nous proposons dans ce chapitre pour la détection de l'incohérence semble respecter la propriété de constituence, principalement parce que ses représentations sont matériellement structurées. Par exemple, le détecteur correspondant à [*p&q*] correspond à l'union des neurones impliqués dans les détecteurs associés à *p* et à *q*. Il *contient* donc les représentations de *p* et de *q*.

Qu'en est-il de la représentation de [*p ou q*] ? La contrainte de constituence exige que si [*p ou q*] a une représentation, alors elle doit inclure la représentation de *p* et celle de *q*. Cela semble poser un problème majeur à notre modèle. En effet, dans notre représentation neuronale des relations logiques, on devrait s'attendre à ce que [*p ou q*] ait comme représentation *l'intersection* des représentations de *p* et de *q*, ce qui serait tout à fait contraire à l'exigence de structuration matérielle qui voudrait que la représentation de [*p ou q*] contînt celle de *p* et celle de *q*. De plus, les représentations des concepts dans la couche $\mathcal{C}$ sont pour nous des détecteurs primitifs, ou la réunion de plusieurs détecteurs primitifs. Rien ne garantit que l'intersection de deux détecteurs sera une telle représentation : elle risque d'inclure des "morceaux" de détecteurs primitifs (10), et donc de ne pas avoir de capacité de détection propre.

En fait, c'est dans ce dernier point que réside la solution. Pour nous, [*p ou q*] N'A PAS de représentation directe dans la couche $\mathcal{C}$, dans la mesure où généralement $D_p \cap D_q$ n'est pas un détecteur. En termes cognitifs, *p ou q* n'est pas un objet élémentaire de la pensée (11). Ce qui peut être représenté, c'est l'incohérence [*non p & non q*] ⇒ ***F***. La sortie de cette incohérence exige "que *p* ou que *q*" soient vrais. D'après notre modèle, un individu ne pense [*p ou q*] qu'en ayant à l'esprit l'impossibilité que *non p* et *non q* soient tous deux vrais(12). Le problème de la structuration matérielle ne se pose donc pas pour des expressions comme [*p ou q*] qui, dans notre modèle, n'ont pas de représentation directe.

---

[9] Certaines architectures connexionnistes semblent échapper à cette description [Chalmers 1990].

[10] Dans sa version actuelle, le modèle ne suppose aucunement que les détecteurs primitifs soient disjoints.

[11] Cela rejoint la préférence manifeste que montrent les individus pour les définitions conjonctives, cf. p. 173.

[12] On peut dire que [*p ou q*] a une représentation mentale dynamique (*p* et *q* sont "pensés" en séquence), contrairement à la représentation statique de [*p & q*].

Il en est exactement de même pour la négation logique *non p* d'un concept *p*(13). (*non p*) n'existe pas en tant que repésentation mentale directe (sauf dans les cas où *non p* coïncide avec un détecteur). Comme nous l'avons annoncé dans le chapitre II.3, l'individu pense *non p* seulement après avoir pensé que *p* participe à un contexte problématique. (*non p*) n'intervient donc que comme concept temporaire, en tant qu'invalidation d'une première relation problématique. Noter que dans le contexte de cette relation problématique, (*non p*) correspond souvent à un détecteur. Par exemple, nous avons vu dans le chapitre I.1 un extrait où, dans le contexte de la conversation, *remorque* et *voiture break* apparaissaient comme l'exacte négation l'une de l'autre.

Citons encore une propriété exigée par Fodor & Pylyshyn pour un système cognitif, la propriété de compositionnalité, qu'ils dénient aux architectures connexionnistes "orthodoxes", et qui semble vérifiée par notre modèle. Nous exprimons cette propriété ainsi :

> **Compositionnalité sémantique** : un ensemble de représentations structurées vérifie la propriété de compositionnalité si l'interprétation sémantique d'une représentation *R* peut être calculée à partir de l'interprétation sémantique des constituants de *R* et de leurs relations dans *R*.

Les représentations de notre modèle sont clairement compositionnelles : la sémantique d'un détecteur *D* réside dans ses capacités de détection, et celles-ci peuvent être calculées par conjonction des capacités de détection des composants de *D*.

5.3. Compatibilité avec les modèles associatifs

Nous avons défini les concepts par leurs propriétés logiques (capacité à entrer dans une relation d'incohérence, cf. chap. II.3. Mais nous n'avons jamais nié l'importance des propriétés associatives des concepts (propriétés qu'ils partagent avec d'autres représentations mentales). Il semble que ces propriétés soient d'ailleurs considérées comme fondamentales par certains auteurs.

Parmi les aspects de notre modèle que nous pouvons considérer comme des points positifs, nous devons précisément noter le fait qu'il permet la cohabitation des processus logiques et des processus associatifs qui opèrent sur les concepts. En effet les concepts, même s'ils entrent dans des relations logiques comme la relation d'incompatibilité, ont néanmoins de nombreuses propriétés de type associatif, qu'ils partagent d'ailleurs avec d'autres représentations mentales (voir chap. II.3). Les associations et analogies comme celles qui sont étudiées dans le programme COPYCAT [Hofstadter & Mitchell 1988] seraient réalisées, dans notre modèle, par des transferts d'activation au sein du deuxième câblage de la couche $\mathcal{C}$ que nous avons dû postuler pour justifier de l'existence de concepts mémorisés. Toutefois, nous sommes loin d'avoir élucidé le fonctionnement de ce câblage, en particulier les aspects *dynamiques* de ce fonctionnement qui semblent essentiels pour le déclenchement des analogies.

5.4. Simplicité et plausibilité neuronale

Le modèle du traitement des incohérences que nous venons de proposer repose sur peu de principes. Tout d'abord le mécanisme de la détection, réalisé par des détecteurs pseudo-locaux qui sont caractérisés par des synapses internes excitatrices, et dont l'activation se traduit par une synchronisation. Ce type de détection ne suppose aucune propriété des neurones autres que celles qui sont bien établies sur un plan biologique.

Nous avons postulé ensuite l'existence de détecteurs "précâblés" dans la couche $\mathcal{C}$. Les organismes vivants disposent de nombreux dispositif neuronaux précâblés, comme les détecteurs de segments orientés des aires visuelles primaires, mis en évidence dans les expériences de D. Hubel et T. Wiesel [1979], ou comme les mécanismes innés de déclenchement (les *Innate Release Mechanisms* postulés par les éthologistes [Lorenz 1978]).

Nous avons essayé de montrer que le coût architectural, et donc génétique, de détecteurs comme ceux de la couche $\mathcal{C}$ est relativement faible. Le principe de fonctionnement de la couche $\mathcal{C}$ elle-même présente l'intérêt d'être le même que celui des détecteurs qu'elle contient : c'est sa synchronisation qui traduit la détection d'une

---

[13] Je remercie Bernard Victorri d'avoir attiré mon attention sur ce point concernant la négation.

incohérence. Nous pouvons imaginer, en termes évolutifs, qu'elle soit issue d'un détecteur pseudo-local standard. Son câblage interne comporte deux types de câblage, l'un fixe, l'autre de type associatif. Elle a donc une structure interne, ce que n'a pas le détecteur de base. Nous pouvons supposer que la couche $\mathcal{C}$ n'est pas une création unique, et émettre l'hypothèse que des structures complexes comme la couche $\mathcal{C}$ sont impliquées dans d'autres opérations cognitives, y compris la perception.

5.5. Parallélisme et temps de réponse

Le sentiment qu'une situation est incohérente est souvent un sentiment quasi instantané. Contrairement aux systèmes symboliques couramment implantés dans les ordinateurs, les individus ne prennent pas le temps d'enchaîner des inférences pour juger de l'incompatibilité de faits perçus. Le modèle que nous proposons permet de rendre compte de ce temps de réponse étonnant.

Le problème de la détection d'une incohérence ne doit pas être confondu avec le problème, de complexité exponentielle, de savoir si n termes donnés sont toujours incompatibles (problème de la satisfiabilité). Le problème résolu par notre modèle neuronal est celui de la *satisfaction* d'une relation d'incompatibilité dans une certaine interprétation. L'interprétation (au sens logique, *i.e.* la valuation qui assigne une valeur de vérité aux propositions) est fournie par la situation perçue. Le fonctionnement de la couche $\mathcal{C}$ permet de savoir si, compte tenu de ces valeurs de vérité, une contradiction sémantique est obtenue. Ce problème de satisfaction d'une relation d'incompatibilité est ainsi résolu de manière parallèle par notre modèle.

5.6. Lien logique - ensembles

Ce modèle de la détection des incohérences permet, comme souvent en pareil cas, de résoudre un problème qui était passé inaperçu. En effet, la logique correspond souvent à des raisonnements abstraits totalement déconnectés de toute interprétation sur des collections d'objets (penser à un raisonnement du genre "S'ils se sont disputés, alors il refusera de lui écrire !"). Par ailleurs, tout humain semble capable de raisonner sur des ensembles (possession et partage d'ensembles d'objets, inclusion de territoires, etc.). Par exemple nous pouvons colorier des intersections ou des réunions dans un diagramme de Venn sans jamais réfléchir aux liens logiques entre les appartenances des éléments. Appelons cette capacité "compétence spatio-ensembliste", dans la mesure où elle semble mettre en jeu nos capacités de représentation spatiale.

Or la logique cadre parfaitement avec une interprétation spatio-ensembliste : l'intersection et la conjonction, l'implication et l'inclusion, la disjonction et la réunion, etc. sont autant de paires de notions dont les termes peuvent être définis indépendamment, et que l'esprit humain ne peut pourtant pas dissocier dès qu'il prend en compte la relation d'appartenance. Par ex. $\forall x \,;\, x \in A \Rightarrow x \in B$ si et seulement si $A \subset B$. Au point que dans la tradition philosophique, les concepts ont toujours eu deux définitions simultanées ! L'une logique (intensionnelle) et l'autre ensembliste (extensionnelle). Par exemple J-B. Grize [1967] présente la "logique des classes" comme faisant presque double emploi avec la logique des propositions !

> Mathématiquement, il y a homomorphisme entre les deux. Si $\mathcal{P}(E)$ est l'ensemble des parties d'un ensemble E, et Φ la bijection de $\mathcal{P}(E)$ dans l'ensemble des prédicats logiques sur E (*i.e.* les fonctions de E dans $\{V;F\}$) : $\Phi(A)(x) = V \text{ ssi } x \in A$, alors on a par exemple $\Phi(A \cap B) = \Phi(A) \,\&\, \Phi(B)$.

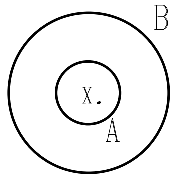
Certes nous avons clairement opté pour une définition intensionnelle des concepts logiques (chap. II.3), et nous avons indiqué que l'extension des concepts ne pouvait pas être définie parmi les objets du monde réel. Mais il reste que dans les situations axiomatisées (ou si l'on préfère, dans les contextes très clairement circonscrits), les concepts logiques peuvent jouer leur rôle de détecteur de manière parfaite : on peut décider sans ambiguité, dans ces cas-là, quel est l'ensemble des objets qui correspondent au concept. On retombe alors sur cette adéquation obligée logique / ensembles.

Or cette adéquation est, *a posteriori*, pour le moins troublante. Elle nous est tellement naturelle que nous omettons de nous en étonner, mais elle *est* étonnante. Pourquoi notre compétence logique et notre compétence spatio-ensembliste montrent-elles une telle convergence parfaite ?

Jean Piaget est l'un des rares auteurs à avoir soulevé cette question, et à avoir suggéré une réponse. Pour Piaget, notre compétence logique est construite, elle *dérive* de l'évolution des opérations que nous maîtrisons au cours de notre enfance, entre autres de notre expérience spatio-ensembliste [Piaget & Inhelder 1947:530].

G. Lakoff aussi a repéré la convergence entre notre compétence spatio-ensembliste et notre compétence logique. Lui aussi fait dériver la deuxième de la première :

> "One of the most interesting properties of image-schemas is that they have built-in logics. For example, BOUNDED REGION schemas, also called CONTAINER-schemas, have essentially a Boolean logic. Consider two CONTAINER-schemas, A and B, such that A is in B, and an object X is in A. We 'see' instantly, without doing any logical deduction, that X is in B. All we need to do is to shift our focus to the relationship between X and B." [Lakoff 1988]



Nous considérons pour notre part que la compétence spatio-ensembliste et la compétence logique *ne dérivent pas* l'une de l'autre. Ce genre d'hypothèse nous semble invraisemblable, compte tenu des observations que nous avons pu faire. Piaget observe la compétence logique en recherchant la trace d'un raisonnement hypothético-déductif. Or l'étude des conversations nous conduit à d'autres conclusions : les capacités conversationnelles des jeunes enfants, dès 3ans, mettent en jeu des compétences logiques qui dépassent largement certaines de leurs capacités spatio-ensemblistes telles qu'elles sont apparentes dans les tests psychologiques [Garitte & Dessalles 1994].

Nous pensons donc que ces deux compétences apparaissent indépendamment chez l'enfant. Mais dans ce cas, comment expliquer qu'il existe entre elles une telle adéquation, que l'on peut caractériser par un homomorphisme mathématique ?

Le modèle neuronal que nous avons décrit ici apporte un élément important qui rend beaucoup moins surprenante cette adéquation mathématique entre compétence logique et compétence spatio-ensembliste, sans poser l'exigence, selon nous peu vraisemblable, d'une origine spatiale ontogénétique de la logique ni, inversement, d'une origine logique de la perception spatiale. Selon notre modèle, les relations logiques (relations d'incompatibilité, d'implication, etc.) sont représentées de manière *ensembliste* dans le cerveau. Dans ces conditions, la coïncidence semble beaucoup moins "miraculeuse", car il est beaucoup plus facile d'accepter qu'il existe une adéquation mathématique entre notre compétence spatio-ensembliste et l'implantation neuronale, ensembliste, des relations logiques. Noter toutefois que la relation d'appartenance impose, dans la perception spatio-ensembliste, des relations duales de celles qui existent entre les neurones de la couche $\mathcal{C}$.

6. Limites du modèle

Le modèle neuronal de la détection d'incohérences esquissé ici est loin d'être complet. Un certain nombre de points restent à préciser, avant que l'on puisse songer à une implémentation :

- la nature des détecteurs primitifs. Ceux-ci sont souvent postulés, mais rares sont les auteurs qui se risquent à donner des exemples de concepts innés. A titre de simple illustration et sans aucune justification, on peut suggérer par exemple le concept d'entité spatio-temporelle (cf. la "whole object assumption" [Markman 1990]).
- l'agencement des détecteurs primitifs dans la couche $\mathcal{C}$. Cet agencement prévoit des incompatibilités indépendamment de toute expérience. Peut-on suggérer, à titre d'exemple, l'impossibilité d'avoir deux localisations distinctes et simultanées pour une entité spatio-temporelle ?
- la mémorisation des concepts. Dans notre modèle, nous n'avons pas décrit ce mécanisme de mémorisation. Nous avons simplement indiqué qu'il n'était pas du type associatif habituel, mais qu'il utilisait l'information issue de la détection d'une incohérence donnée par la couche $\mathcal{C}$.
- le fonctionnement dynamique du câblage associatif de la couche $\mathcal{C}$, qui permet de déclencher des associations et des analogies.

D'un autre côté, le modèle est aussi incomplet parce qu'il est très loin de rendre compte de l'ensemble des processus conceptuels. Nous n'avons fait que mentionner le fait que les concepts pouvaient évoquer des perceptions, provoquer l'activation de schémas d'action intégrés, et qu'ils avaient des liens très étroits avec le système linguistique. De même, les traitements conceptuels que nous avons observés dans les conversations sur les modes improbable et (in)désirable (cf. chap. I.1) ne sont pas couverts par ce modèle neuronal.

Ces points ne sont pas rédhibitoires compte tenu de notre objectif qui est de suggérer un mécanisme plausible pour la détection d'incohérences. Toutefois, nous pensons qu'il serait particulièrement intéressant de proposer des implémentations de ce modèle, qui permettrait de prouver sa fécondité, son intérêt pour la modélisation cognitive et enfin, peut-être, son intérêt technique pour l'apprentissage automatique.

7. Conclusion : "ramener la logique au neurone"

Dans le contexte scientifique actuel, la séparation symbolique/sub-symbolique est tellement affirmée que la réalité cognitive des processus symboliques est parfois niée [Smolensky 1988], principalement parce que l'on n'est pas en mesure de lui trouver une implantation neuronale plausible. L'objectif de ce chapitre était justement de suggérer une implantation plausible d'un mécanisme symbolique, le mécanisme logique de détection d'incohérence.

Notre modèle fait intervenir un appareil neuronal spécifique, la couche "$\mathcal{C}$", pour le traitement de l'incohérence et des relations logiques. Il est en accord avec les exigences de la systématicité, de la constituence et de la compositionnalité. Il est de plus plausible sur le plan neuronal, car les structures qu'il fait intervenir (détecteur collectif) résultent des propriétés de base des neurones et présentent un coût architectural faible.

Notre modèle est bien sûr incomplet. Mais, rappelons-le, notre objectif était avant tout de démystifier le caractère symbolique de certains traitements réalisés par les individus, en montrant qu'une architecture neuronale relativement simple pouvait expliquer des performances comme la détection d'incohérences pour lesquelles on est obligé de postuler un traitement de type logique. D'une certaine façon, nous avons voulu montrer qu'il n'était pas déraisonnable de vouloir "ramener la logique au neurone", de relier les propriétés d'un traitement symbolique particulier, la détection d'incohérences, aux propriétés que l'on reconnaît généralement aux cellules du cerveau.

remerciements : je remercie Olivier Hudry et Jean-Bernard Auriol pour les discussions qui m'ont permis de progresser dans la définition de ce modèle neuronal de la détection d'incohérences. Je remercie Cédric Chappelier et Bernard Victorri pour leurs critiques détaillées de ce texte.