



HAL
open science

Modélisation de la propagation de l'information cérébrale par graphes causaux qualitatifs

Marc Lafon

► **To cite this version:**

Marc Lafon. Modélisation de la propagation de l'information cérébrale par graphes causaux qualitatifs. Modélisation et simulation. Université Paul Sabatier - Toulouse III, 2000. Français. NNT: . tel-00010176

HAL Id: tel-00010176

<https://theses.hal.science/tel-00010176v1>

Submitted on 16 Sep 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ TOULOUSE III - PAUL SABATIER

U.F.R. Mathématiques, Informatique et Gestion.

T H E S E

Pour obtenir le grade de

Docteur de l'université Toulouse III

Discipline : Informatique.

Présentée et soutenue par

Marc LAFON

le 3 mai 2000.

**Modélisation de la propagation de
l'information cérébrale par graphes causaux
qualitatifs**

JURY

M. Michel Cayrol, Professeur, <i>IRIT, Université Toulouse III,</i>	Président,
Mme Marie-Odile Cordier, Professeur, <i>IRISA, Université Rennes I,</i>	Rapporteur,
Mme Liliana Ironi, Cherheur <i>LAN-CNR, Pavie,</i>	Rapporteur,
M. Jean-Luc Nespoulous, Professeur, <i>Laboratoire J. Lordat, Université Toulouse II,</i>	Rapporteur,
Mme Josette Pastor, Ingénieur de recherche, HDR, <i>INSERM Unité 455, Toulouse,</i>	Directrice,
Mme Louise Travé-Massuyès, Chargée de recherche, HDR, <i>LAAS/CNRS, Toulouse,</i>	Co-Directrice.

Marc LAFON

MODELISATION DE LA PROPAGATION DE
L'INFORMATION CEREBRALE PAR GRAPHES
CAUSAUX QUALITATIFS

Directrice de thèse :

Josette PASTOR, Habilitée à Diriger des Recherches en Informatique,
INSERM Unité 455 – Toulouse
Neuro-Imagerie fonctionnelle, plasticité cérébrale et pathologie neurologique.

Co-Directrice de thèse :

Louise TRAVEMASSUYES, Habilitée à Diriger des Recherches en Automatique,
LAAS/CNRS – Toulouse
Groupe Automatique et Symbolique.

- Résumé -

Dans les études d'activation, l'activité du cerveau humain est mesurée grâce à diverses techniques d'imagerie neuro-fonctionnelle. Ces études mettent en évidence les réseaux d'aires cérébrales mis en jeu lors de la réalisation de tâches cognitives. Mais la complexité du traitement cérébral et l'incomplétude de l'observation compliquent l'interprétation de ce fonctionnement et nécessitent la construction d'un outil de modélisation et de simulation. Pour parvenir à une compréhension de l'organisation fonctionnelle du cerveau humain, un recours à une représentation explicite est nécessaire.

Dans cette optique, nous proposons un formalisme original, fondé sur des graphes d'influences causales et une représentation qualitative de l'information. Les graphes nous permettent de représenter la double organisation du cerveau, en réseaux anatomiques (structurels) et fonctionnels. Nous proposons une architecture hiérarchique qui réduit la complexité des modèles tout en augmentant leur réutilisabilité et nous utilisons un formalisme homogène pour décrire les deux niveaux. Une modélisation explicite de l'information cérébrale nous conduit à une représentation qualitative mixte, sous la forme d'un couple (intervalle, symbole) pour prendre en compte l'amplitude et la catégorie de l'information. Fondés sur ce formalisme, nous proposons le simulateur BIOCAEN limité aux processus automatiques qui constituent la majeure partie du fonctionnement cérébral.

La contribution de notre travail réside dans la mise à disposition d'un environnement permettant d'exprimer, à un haut niveau d'abstraction, des hypothèses sur la propagation de l'information cérébrale. Elles sont décrites par des modèles où l'architecture causale sert de support à l'expression des réseaux anatomiques d'aires cérébrales, chacune étant représentée par un sous-réseau fonctionnel. La confrontation des résultats de la simulation aux données de l'imagerie neuro-fonctionnelle permet la validation des hypothèses.

Mots Clés : Modélisation cognitive, graphe causal hiérarchique, simulation qualitative.

INSERM Unité 455,
Services de Neurologie, C H U .Purpan,
F-31059 TOULOUSE CEDEX .

TABLE DES MATIERES

Introduction	9
Chapitre 1. Études d'activation cérébrale.....	11
1. Le cerveau humain	13
1.1. Subdivisions anatomiques du système nerveux central.....	13
1.2. Fonctionnement cellulaire	17
2. Techniques d'imagerie fonctionnelle	20
2.1. Techniques de surface	20
2.2. Tomographies	22
2.3. Conclusion	25
3. Interprétation des résultats d'études d'activation	25
Chapitre 2. Modèles computationnels de l'activité cérébrale	28
1. Introduction	28
2. Modélisation fonctionnelle	31
2.1. Approche symbolique.....	31
2.1.1. Les premiers modèles	32
2.1.2. La théorie ACT*	33
2.2. Approche connexionniste	36
2.2.1. Description de base des réseaux de neurones formels	37
2.2.2. Apprentissage	41
2.2.3. Transparence sémantique et RNF	45
2.3. Conclusion	46
3. Modélisation structurelle	47
3.1. Différents niveaux d'organisation du cortex	48
3.1.1. Le niveau cellulaire	48
3.1.2. Le niveau modulaire	49
3.1.3. Le niveau régional.....	50
3.1.4. Le niveau global	50
3.2. RNF et modélisation structurelle	51
3.3. Formalismes dédiés à la modélisation structurelle	52
4. Conclusion.....	55
Chapitre 3. Formalismes représentant des connaissances imprécises et approche causale.....	57
1. Introduction	57
2. Le calcul des probabilités	58
2.1. Événement probabiliste	58
2.2. Probabilité	58
2.3. Modèle probabiliste	59
3. Logique Floue	60
3.1. Logiques multivaluées.....	61
3.1.1. Fondement des logiques multivaluées	61
3.1.2. Logique L_3	62
3.1.3. Logique L_{N1}	63
3.2. Théorie des ensembles flous.....	64
3.3. Logique possibiliste	66
4. Raisonnement qualitatif.....	68
4.1. L'algèbre des signes	69
4.2. Les ordres de grandeur	70
4.2.1. Ordres de grandeur absolus	70
4.2.2. Ordres de grandeur relatifs	72
4.3. Les intervalles	73
4.4. La simulation qualitative	75
5. L'approche causale	77
6. Conclusion.....	80

Chapitre 4. Formalisation de l'activation fonctionnelle	81
1. Notions fondamentales	81
1.1. Processus cognitifs modélisés	82
1.2. Architecture des modèles	82
1.2.1. Modélisation à 2 niveaux	84
1.2.2. Graphes d'influences	86
1.2.3. Modèles génériques	87
1.3. Représentation de l'information	88
1.3.1. Amplitude de l'information	89
1.3.2. Type de l'information	90
1.3.3. Conclusion	91
2. Architecture des modèles	92
2.1. Les graphes causaux de processeurs	92
2.1.1. Les nœuds du graphe	93
2.1.2. Les connexions du graphe	93
2.2. Modèle de composant	97
2.3. Modèles de composants génériques	99
2.4. Graphe global	101
2.5. Conclusion	103
3. Représentation de l'information	103
3.1. Amplitude	103
3.1.1. Les intervalles étendus	104
3.1.2. Relations entre intervalles étendus	105
3.1.3. Opérations sur les intervalles étendus	107
3.1.4. Relations multivaluées	109
3.2. Type	111
3.2.1. Relations sur les types	112
3.2.2. Opérations sur les types	113
3.3. Conclusion	114
4. Simulation dynamique du système	114
4.1. Horloge	115
4.2. Fonctions de mise à jour	117
4.2.1. Condition d'activation	118
4.2.2. Gains dynamiques	120
4.2.3. Le délai	120
4.2.4. Temps de réponse	121
4.2.5. Conclusion	122
4.3. Composition et propagation	123
4.4. Calculabilité d'un graphe global	124
5. Conclusion	126
Chapitre 5. Le simulateur BIDCAEN	127
1. Algorithmes	127
1.1. Le simulateur Ca-En	127
1.2. Organisation générale du simulateur	128
1.3. Description des modèles	129
1.3.1. Description des valeurs constantes	130
1.3.2. Description d'un atome	131
1.3.3. Description d'une influence	133
1.3.4. Description d'un composant	135
1.4. Algorithme de compilation	135
1.4.1. Génération du graphe global	137
1.4.2. Simplifications du graphe global	137
1.5. Algorithme de simulation	139
1.6. Implémentation	141
2. Application et validation	142
2.1. Modèle générique de zones associatives	142
2.1.1. Interface d'entrée du modèle	144
2.1.2. Nœud d'intégration	144
2.1.3. Nœud d'inhibition	146
2.1.4. Nœud d'activation	146

2.1.5. Nœud de propagation	146
2.1.6. Nœud de mémoire de travail	146
2.1.7. Surcharge de la zone	147
2.1.8. Conclusion	147
2.2. Modèles de la réponse de l'aire visuelle primaire à un stimulus simple	148
2.2.1. Présentation de l'étude d'activation	148
2.2.2. Premier modèle : La capacité d'apprentissage de l'aire visuelle primaire	150
2.2.3. Second modèle : La boucle cortico-thalamique	151
2.2.4. Simulation	153
3. Conclusion	156
Conclusion	157
1. Formalisme général pour la modélisation cognitive	157
1.1. Représentation de l'information	157
1.2. Traitement automatique de l'information	158
2. Formalisme explicite	159
2.1. Haut degré d'abstraction de l'information	160
2.2. Représentation hiérarchique homogène	160
2.3. Modularité et réutilisabilité des modèles	161
3. Position de notre approche	162
3.1. Approche structurelle hybride	162
3.2. Simulation qualitative	163
4. Perspectives	164
4.1. Contrôle de l'accroissement de l'imprécision	165
4.2. Inférence distribuée sur les types	167
4.3. Architecture hiérarchique généralisée	168
5. Conclusion	169
Bibliographie	170
Index	175
Annexe A : Exemples commentés de descriptions BIO CAEN	177
1. Description du modèle générique de zones associatives	177
2. Modélisation de la réponse de l'aire visuelle primaire à un stimulus simple	179
2.1. La capacité d'apprentissage de l'aire visuelle primaire	179
2.2. Modèle de la boucle cortico-thalamique	180

TABLE DES ILLUSTRATIONS

Num éro	Page
1. Décom position anatom ique du cerveau	14
2. G randes subdivisions du cortex.	15
3. Localisation d'aires fonctionnelles.....	16
4. Structure du neurone	18
5. Proj ection d'un potentiel évoqué m esuré par EEG , (Thierry et al. 1998).....	21
6. Im ages tom ographiques	23
7. A rchitecture de la théorie A C T* d'après Anderson (1989).	34
8. Neurone form eld d'après Rum elhart (1989).	38
9. Schém a sim plifié du résea ux en couche de lecture de m ots d'après McC lelland et Rum elhart (1981)	40
10. Résea u de Kohonen, d'après A lliot et Schiex (1994).....	43
11. Connexion entre deux plans définie par une proj ection topographique, d'après Sakai et al. (1997).	53
12. Fonction d'appartenance de l'ensem ble flou associé au prédicat « est une personne âgée »	64
13. Fonction d'appartenance trapézoïdale.....	65
14. Interprétation contrainte de A rB en fonction du rapport A /B dans $O(M)$, d'après M avrovouniotis et Stephanopoulos (1988).....	72
15. Illustation neuronale du couple A m plitude /Type.....	91
16. D ifférentes sortes de connexions.....	94
17. Points de connexion d'un com posant.....	95
18. Schém as de connexions.....	96
19. Exem ple de la définition d'un graphe.	98
20. Exem ple d'une définition récursive.	99
21. Instanciation des param ètres d'un m odèle générique.....	100
22. Construction du graphe global.	102
23. Condition d'activation d'une influence.....	119
24. D e m ultiples influences convergentes.....	123
25. U n graphe global incalculable.....	125
26. O rganisation d'une sim ulation sous B I D C A E N	129
27. A lgorithme de com binaison des valeurs m arginales pour l'opérateur ForA LL.....	133
28. A lgorithme de com pilation de B I D C A E N , prem ière partie, obtention du graphe global.	136
29. A lgorithme de com pilation de B I D C A E N , seconde partie, test de validité du graphe global.	138
30. A lgorithme de sim ulation de B I D C A E N	140
31. M odèle fonctionnel générique d'une aire associative.	143
32. Résultats de l'expérience de Fox et Raichle, d'après Fox et Raichle (1985).	149
33. M odèle fonctionnel du cortex strié, d'après Pastor et al. (2000).	151
34. M odélisation de la boucle cortico-thalam ique, d'après Pastor et al. (2000).	152
35. Extraits des fichiers de m esures (M esures de l'am plitude de l'atom e V L A N).....	153
36. Résultats globaux de la sim ulation, d'après Pastor et al. (2000).....	154

REMERCIEMENTS

Je tiens à remercier le professeur François Chollet pour m'avoir accueilli au sein de l'unité 455 de l'INSERM dont il est le directeur.

Je remercie Josette Pastor, Ingénieur de recherche à l'unité 455, tout d'abord, pour avoir accepté de diriger mes travaux, puis, pour l'avoir réellement fait, pour toujours avoir été présente et avoir accompli la lourde tâche qu'a représenté ma formation. Je la remercie donc pour son grand courage et pour sa gentillesse.

Je remercie Louise Travé-Massuyes, directrice du groupe AS au LAAS/CNRS, pour avoir accepté de co-diriger ma thèse et pour son contact sympathique, d'avoir apporté une expertise vitale à mon travail et avoir grandement complété mon apprentissage.

Je remercie le professeur Marie-Odile Cordier d'avoir bien voulu examiner mes travaux, pour les commentaires très complets qu'elle a émis et qui ont contribué à l'achèvement de ce document. Je la remercie aussi pour son contact chaleureux et très convivial.

Je remercie le professeur Liliana Ironi pour avoir franchi la barrière de la langue en acceptant de faire partie de mes rapporteurs. Je la remercie pour la clarté de ses commentaires et pour l'évident travail de compréhension qu'elle a produit à l'égard de mon travail. *Grazie mille.*

Je remercie le professeur Jean-Luc Nespoulous pour avoir bien voulu relire et commenter mes travaux malgré leur technicité. Je le remercie pour son ouverture d'esprit, ses innombrables qualités et la gentillesse qu'il a manifesté à mon égard.

Je remercie le professeur Michel Cayrol d'avoir accepté de présider mon jury, mais aussi et surtout pour l'enseignement de qualité qu'il prodigue à l'université Paul Sabatier et dont j'ai pu bénéficier.

Je remercie le docteur Pierre Celsis, chercheur à l'unité 455, pour m'avoir adopté dans son équipe et pour avoir su toujours trouver une réponse compréhensible à mes oreilles et ce, quelle que soit ma question. Je remercie aussi le docteur Bernard Doyon pour l'étonnante richesse de ses remarques et pour ses commentaires enthousiasmants. Et enfin, je remercie le docteur Jean-François Démonet, pour la source d'inspiration qu'il incarne, pour sa disponibilité à mon égard et son ouverture d'esprit. Je remercie enfin l'ensemble de cette triade pour leur gentillesse à mon égard.

Je remercie Bertrand Lacotte pour le travail qu'il a effectué avant moi sur ce projet, pour son esprit créateur et très objectif, ainsi que pour la base de travail solide qu'il m'a confiée et qui a énormément contribué à mon avancement.

Je témoigne de mon infinie gratitude à tous ceux qui ont contribué de près ou de loin à rendre mon travail plus agréable et à m'enrichir personnellement durant ces dernières années, les chercheurs, le secrétariat et le personnel soignant de l'unité 455, ainsi que tous les étudiants de tous horizons et laboratoires. La liste est bien trop longue pour contenir dans ces pages.

Je rends grâce, plus largement encore, à tous mes amis, mes *ex*-camarades, les étoiles filantes et les immortels, à mon frère, ceux qui sont toujours jeunes et les plus sages, à tous ces êtres exceptionnels qui m'entourent, pour leurs merveilleuses qualités et leurs aimables défauts. Je m'excuse ne n'avoir pas pu être aussi présent, aimant et dévoué qu'ils l'ont été à mon égard. Encore une fois, je suis bien incapable d'être exhaustif, alors je ne donnerai pas de noms, en espérant qu'ils se reconnaîtront. Je les garde tous dans mon cœur, là où ils resteront toujours. Je dédie l'aboutissement de ce travail à mes parents, que j'aime tant et sans qui je n'aurais jamais pu être à la hauteur de cette tâche, en témoignage de ma gratitude et de mon affection.

A Danielle et Michel LAFON.

*Un gros meuble à tiroirs encombré de bilans,
De vers, de billets doux, de procès, de romances,
Avec de lourds cheveux roulés dans les quittances,
Cache moins de secrets que mon triste cerveau.
Charles Baudelaire (Spleen et Idéal, LXXIX).*

Si l'étude de la pensée remonte à l'antiquité, l'établissement de son origine dans l'activité du cerveau est beaucoup plus récente. Depuis deux cents ans, l'étude du fonctionnement du cerveau humain a motivé de nombreuses recherches scientifiques dont les résultats ont bouleversé les connaissances et les croyances que l'on pouvait avoir sur l'organisation et le fonctionnement du cerveau. Plus particulièrement, les progrès technologiques de ces vingt dernières années ont permis la naissance d'un nouveau paradigme expérimental, les études d'activation. Il est désormais possible d'observer, par l'intermédiaire de la *neuro-imagerie fonctionnelle*, l'activité du cerveau pendant qu'il réalise des tâches cognitives et ce, sans aucun traumatisme.

Les études d'activation ont permis de mieux connaître l'organisation fonctionnelle du cerveau. Pour résoudre une tâche donnée, c'est tout un réseau de régions du cerveau qui semble être mise en œuvre. Chaque aire localisée est une portion du cortex de taille variable dont les frontières sont variables en fonction des individus. Dans l'accomplissement d'une tâche cognitive particulière, les aires impliquées sont dépendantes les unes des autres, mais elles procèdent à un traitement spécifique et indispensable à la résolution de la tâche cognitive. Les chercheurs en neuropsychologie tentent de définir le lien entre les fonctions cognitives réalisées et les structures mises en œuvre pour y parvenir. Cependant, beaucoup de tentatives pour établir un lien définitif entre structure cérébrale et fonction cognitive se sont soldées par des échecs. Ceci peut s'expliquer par une absence d'interprétation en termes des mécanismes fondamentaux de traitement de l'information cérébrale, déterminés par la neurophysiologie et la neuroanatomie. Les moyens de neuro-imagerie ne permettent que d'observer la conséquence de la mise en œuvre de ces mécanismes (i.e. l'activation de certaines aires cérébrales). La compréhension des mécanismes ne peut venir que de la modélisation. Il est donc indispensable d'établir des modèles physiologiquement plausibles des liens entre structures et fonctions cérébrales.

La plupart des recherches faites dans ce sens sont fondées sur les réseaux de neurones formels (RNF) et se proposent de déterminer quelle fonction émerge d'un réseau de micro-structures cérébrales (les neurones). Les modèles fondés sur les RNF perdent toute plausibilité physiologique lorsqu'ils sont appliqués aux réseaux de macro-structures cérébrales (les aires). C'est pourquoi,

l'objet de notre travail est double. Dans un premier temps, nous proposons un formalisme qui permet de décrire des modèles physiologiquement plausibles du traitement de l'information dans les réseaux d'aires cérébrales. Dans un second temps, nous construisons un simulateur qui permet de mettre ces modèles en œuvre.

D'une part, nous voulons proposer un cadre formel dans lequel pourront s'exprimer le plus grand nombre de modèles possibles. D'autre part, nous ne cherchons pas, comme c'est le cas dans d'autres domaines (physique par exemple) à simuler le comportement d'un système connu. Actuellement, le niveau de connaissances des mécanismes régissant l'activité cérébrale n'est pas suffisant pour permettre une telle aventure. Notre ambition est d'offrir une méthodologie et un outil qui permettent de décrire des hypothèses, compatibles avec les connaissances acquises, qui pourraient expliquer les observations issues des études d'activation. La nuance est cruciale. Le domaine des neurosciences étant très dynamique et les connaissances progressant sans cesse, les hypothèses sont nombreuses et évoluent constamment.

Nous proposons donc un formalisme dans lequel nous pouvons exprimer facilement et explicitement des hypothèses portant à la fois sur les relations entre les aires cérébrales et sur le fonctionnement propre à chacune d'elles. Pour ce faire, nous utilisons une représentation par graphes orientés, cohérente avec l'organisation distribuée du cerveau. Ces graphes sont organisés hiérarchiquement pour diminuer la complexité des modèles et pour représenter explicitement leurs différents niveaux d'abstraction. La propagation est modélisée par des opérateurs qualitatifs propres à maîtriser l'imprécision caractéristique de « l'information cérébrale », que nous définissons comme une abstraction du signal cérébral.

Compte tenu du fait que cette thèse s'adresse à un public pluridisciplinaire, nous avons fait un effort pédagogique dans la description de l'état de l'art. Nous sommes conscients que, dans chaque discipline, certaines parties des trois premiers chapitres pourront paraître quelque peu naïves aux yeux du spécialiste. Cependant, les références bibliographiques ont été particulièrement sélectionnées pour permettre le développement des sujets abordés. Après une rapide description, anatomique et fonctionnelle, du cerveau et un aperçu des techniques d'observation de l'activité cérébrale, nous poserons les objectifs de notre problématique. Puis nous verrons les différentes approches formelles de la modélisation de l'activité cérébrale ainsi que leurs caractéristiques fondamentales. Nous pourrons alors introduire notre approche et définir notre formalisme, ainsi que le simulateur BIOCAEN. Finalement, nous critiquerons notre travail et nous discuterons de ses futures évolutions possibles.

ÉTUDES D'ACTIVATION CÉRÉBRALE

Les premières tentatives de liens entre la structure cérébrale et les fonctions cognitives datent de bien avant notre ère (e.g. Herophile, médecin grec, 300 A.C., croit y situer le siège de l'âme). Les premières études réellement abouties datent, cependant, du siècle dernier, avec les travaux du médecin allemand, Franz Jozef Gall, créateur de la phrénologie, en 1810, qui localisait les fonctions mentales suivant la conformation du crâne humain. Cette théorie prête aujourd'hui à sourire mais ses hypothèses de base étaient pertinentes. Gall était en effet le premier à considérer le cortex cérébral comme un lieu privilégié pour la réalisation des fonctions cognitives et stipulait que l'on pouvait les y localiser (Mazoyer & Belliveau 1996, Laplane 1985). C'est la phrénologie qui marque le véritable début de la *théorie de la localisation*, selon laquelle l'esprit n'est pas une entité indivisible désincarnée, mais en un ensemble de fonctions localisables dans le cerveau. Elle reste aussi un très bon exemple de la difficulté à localiser des fonctions cognitives sans techniques d'observation directe de l'activité du cerveau.

Durant tout le XIX^e siècle, les observations anatomiques se succèdent. Citons, par exemple, la découverte de la division en couches corticales par Baillarger en 1840, ou l'atlas du cerveau établi par Cruveilhier en 1829 ou bien encore celui de Korbianan Brodmann qui date de 1909 et est encore utilisé aujourd'hui. Les connaissances sur l'organisation fonctionnelle du cerveau se développent, elles aussi, sous la forme d'études cliniques. Les relations entre les fonctions cognitives et la structure cérébrale sont, alors, déduites des corrélations entre la perte d'une fonction cognitive et la localisation d'une lésion cérébrale. Cette localisation était observée *post-mortem*. C'est, par exemple, le cas des travaux sur l'aphasie que Paul Broca réalisa en 1861 et qui lui permirent de localiser une aire cérébrale impliquée dans le langage et qui depuis porte son nom (Mazoyer & Belliveau 1996, Mesulam 1990). Cependant les corrélations sont souvent ardues à réaliser, surtout lorsque le déficit est important. De plus, la grande variabilité des lésions et les facultés de réorganisation du cerveau rendent difficile la transposition des observations du cerveau lésé au cerveau sain (Evans 1996). On peut également souligner une observation générale et d'une importance fondamentale. Dans certains cas, des patients présentant des lésions situées en des endroits différents du cerveau, présentent un déficit touchant une même fonction. Symétriquement, une même lésion peut

entraîner des déficits impliquant plusieurs fonctions cérébrales. Cette observation peut être entendue comme une limite à la localisation des fonctions cérébrales dans le cerveau humain par la méthode anatomo-clinique (Laplaine 1985).

Durant les deux premiers tiers de ce siècle, la seule alternative pour l'étude de l'organisation était, d'une part, l'étude clinique post-mortem, chirurgicale (Philippon 1988) ou en utilisant une technique d'imagerie anatomique (e.g. exploration par radio-isotope, angiographie) et, d'autre part, les travaux sur les animaux. Mais, dans ce dernier cas, la transposition à l'homme des résultats reste en général assez problématique, puisque les organes sont différents. De plus, la transposition est limitée à des fonctions réalisables par l'animal étudié, ce qui exclut, entre autres, les études sur le langage (Mesulam 1990).

Les études d'activation sont nées avec toutes les nouvelles techniques d'imagerie fonctionnelle, datant pour les plus anciennes des années soixante. Grâce à ces techniques, on a pu étudier, et surtout mesurer, l'activation cérébrale provoquée par la réalisation d'une tâche cognitive, chez des patients cérébro-lésés, mais aussi chez des sujets sains. Ces nouvelles études ont permis d'obtenir des informations plus précises sur les localisations des fonctions cérébrales et, notamment, elles ont mis en évidence une activité plus distribuée sur l'ensemble du cerveau. Une fonction cognitive donnée entraîne une activité en plusieurs endroits du cerveau anatomiquement repérés, les *aires cérébrales*. Ces aires ne sont pas entièrement dédiées à la résolution d'une seule fonction cognitive, contrairement à ce que les études cliniques des lésions pouvaient laisser croire. Une même aire cérébrale peut, en effet, être activée pour la résolution de différentes fonctions. Pourtant, le rôle de l'aire dans la mise en œuvre de la fonction est indispensable, puisqu'une lésion de l'aire entraîne une perte plus ou moins importante des aptitudes à résoudre la fonction. Ainsi, l'étude de l'activité cérébrale est-elle passée, ces vingt dernières années, de la localisation d'une aire dédiée à une fonction, à la localisation d'un groupe d'aires impliquées dans la fonction (Pulvermüller 1996, Mesulam 1990, Posner *et al.* 1988). Pour comprendre le rôle de chaque aire impliquée dans l'accomplissement d'une fonction cognitive, il faut diviser cette fonction en primitives fonctionnelles et effectuer des recoupements avec d'autres fonctions possédant des primitives semblables.

La technique de base d'une étude d'activation est théoriquement très simple. On demande au sujet d'effectuer une opération cognitive (e.g. compter jusqu'à dix, lire un texte, mémoriser un chiffre, etc) et pendant qu'il la réalise, on mesure l'activité de son cerveau. Cette description paraît effectivement fort simple, mais elle ne l'est évidemment pas. La grande difficulté réside dans

l'élaboration d'un protocole expérimental qui ne mette en jeu que la fonction cognitive que l'on souhaite observer. Si cela n'est pas possible, on doit utiliser une tâche de référence, qui regroupe les fonctions à supprimer, et que le sujet accomplit aussi. On compare alors les activités générées par les deux tâches de manière à détacher l'activité propre à la fonction que l'on souhaite étudier (Sergent 1994, Posner *et al.* 1988). On peut aussi de cette manière déterminer les différences entre deux fonctions en soustrayant l'activité de la première à la seconde, et inversement. Ce mécanisme est aussi employé en comparant l'activation dans une tâche cognitive à l'activité du cerveau au repos, pour supprimer l'effet de l'activité cérébrale spontanée sur les mesures. En effet, le cerveau étant un système très réactif au milieu qui l'entoure, l'activité liée à la tâche du sujet est toujours noyée dans cette activité spontanée, qui est un mélange de processus incontrôlables, traitements automatiques ou inconscients.

Après une rapide description de l'anatomie du cerveau humain, aussi simple que possible, nous verrons les principales techniques d'imagerie fonctionnelle actuellement utilisées. Sans trop entrer dans les détails, nous verrons quels types de mesures les chercheurs en neuropsychologie peuvent attendre de ces techniques et comment s'y prendre pour obtenir des images de l'activité cérébrale. Alors nous pourrions mieux faire comprendre les objectifs de notre recherche.

1. LE CERVEAU HUMAIN

Il ne s'agit pas, ici, de décrire l'ensemble des mécanismes biologiques impliqués dans le fonctionnement du cerveau humain, loin de là. Nous allons simplement, dans un premier temps, décrire sommairement comment on se repère dans le cerveau humain, sa géographie en quelque sorte. Dans un second temps, nous donnerons quelques détails sur la structure cellulaire du cerveau, ce qui nous permettra de mieux comprendre ce que l'on observe réellement avec les techniques d'imagerie cérébrale. Pour des informations plus complètes de nombreux ouvrages de référence peuvent être consultés (Rosenzweig & Leiman 1991, Hedaen 1972).

1.1. Subdivisions anatomiques du système nerveux central

Le *système nerveux central* (SNC) est composé de l'*encéphale* et de la moelle épinière. L'encéphale est la partie du système nerveux contenue dans la boîte crânienne et est communément reconnue comme

étant le siège de la conscience. L'encéphale peut être subdivisé anatomiquement en trois grandes parties, le *cerveau* proprement dit¹, le *cervelet* et le *tronc cérébral* reliant le cerveau à la moelle épinière.

En tant que partie de l'encéphale, le cerveau est protégé par la boîte crânienne. Pour l'atteindre, outre le scalp, il faut pénétrer l'épaisse carapace des os du crâne. On arrive alors aux *méninges*, qui sont des membranes enveloppant totalement le tissu nerveux et sont plus ou moins épaisses. On distingue trois membranes : la dure-mère, l'arachnoïde et la pie-mère. L'arachnoïde se dédouble pour contenir un liquide, le liquide céphalo-rachidien, qui n'est pas uniformément réparti sur la totalité de la membrane, formant ainsi des sortes de poches. Tout ceci constitue autant d'enveloppes d'épaisseurs non uniformes et de densités différentes qui sont des filtres pour une observation externe de l'activité du cerveau.

On convient de diviser le cerveau en trois parties principales (Figure 1) qui sont le télencéphale, le diencéphale et le mésencéphale. Ce dernier est le plus profond. Le diencéphale est composé entre autre, d'une paire de structures ovoïdes, appelé le thalamus (droit et gauche) dont on sait qu'il intègre les information sensorielles en direction du télencéphale. C'est le télencéphale qui constitue la majeure partie du cerveau. Il coiffe le diencéphale. C'est au niveau du télencéphale que se concentrent les observations de l'imagerie cérébrale car c'est ici que se situe l'activité cérébrale la plus élevée, en termes de fonctions cognitives.

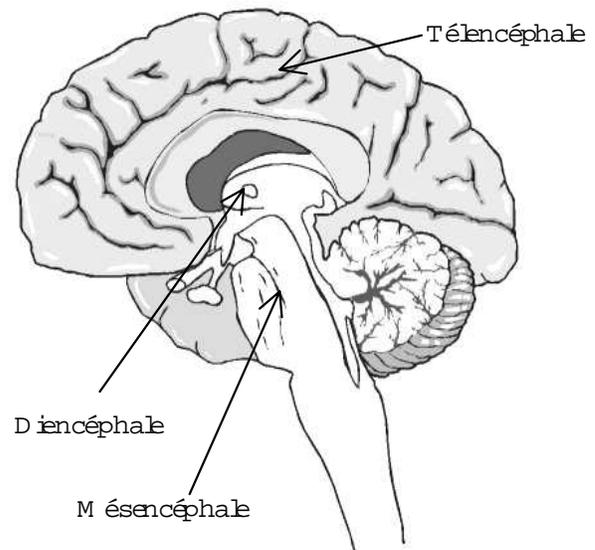


Figure 1 : Décomposition anatomique du cerveau

Le télencéphale est séparé par l'axe médian en deux *hémisphères* (droit et gauche) dont la surface, totalement plissée, constitue le *cortex cérébral*. Cette surface est plus vaste qu'il n'y paraît puisqu'elle atteint presque les 1,5 m². Les figures formées par les sillons et les circonvolutions du cortex varient d'un individu à l'autre. Les deux hémisphères ne sont d'ailleurs pas totalement symétriques chez une même personne. Les berges des sillons sont appelés *gyri* et leurs creux sont appelés *saisures*, lorsqu'ils sont profonds, ou *sulci*, lorsqu'il le sont moins.

¹ Au sens anatomique restreint du terme (Le Robert).

Ces figures morphologiques permettent la division du cortex en *lobes* et la localisation d'aires du cerveau sans recourir à aucun système métrique. On peut par exemple aisément distinguer 3 scissures à la surface du cerveau, la plus visible est la scissure hémisphérique, qui sépare les deux hémisphères. Vient ensuite la scissure de Rolando, ou scissure centrale, qui divise verticalement les deux hémisphères à peu près en leur milieu. On distingue enfin la scissure de Sylvius, ou scissure latérale qui, elle, est horizontale de chaque côté du cerveau, en plein centre de chaque hémisphère (Figure 2). Ces scissures permettent de diviser chaque hémisphère en 4 lobes : le lobe frontal, le lobe pariétal, le lobe temporal et le lobe occipital. Un autre lobe, le lobe limbique, n'est pas visible à la surface du cortex, car situé sur la surface interne du télencéphale. Le corps calleux constitue le seul axe de connexion entre les deux hémisphères. Il détermine de ce fait un axe primordial dans l'interaction entre les hémisphères.

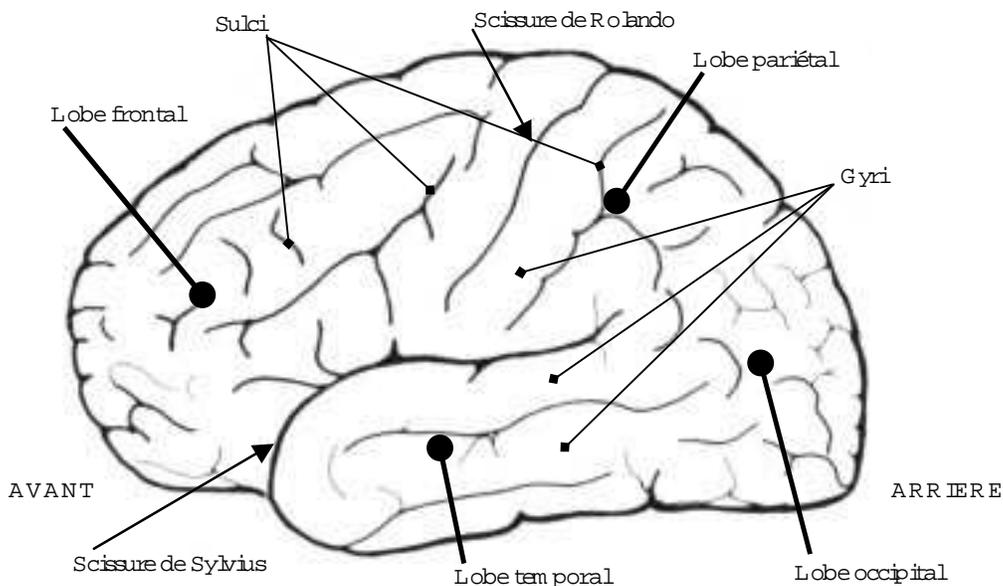


Figure 2 : Grandes subdivisions du cortex.

Il existe aussi un découpage fonctionnel du cortex (Figure 3). On y différencie des portions assez vastes dont l'activité semble plus ou moins dédiée à des tâches spécifiques. C'est le cas, par exemple, du cortex moteur où l'on retrouve des aires impliquées dans le contrôle musculaire. Les études cliniques, animales et, maintenant, les études d'activation ont permis de dresser un découpage fonctionnel très précis des aires motrices primaires (gyrus précentral de chaque hémisphère). Cette cartographie fonctionnelle est appelée *somatotopie* (Churchland & Sejnowski 1992), elle permet d'associer à chaque partie du cortex moteur primaire un segment corporel correspondant (e.g. mains, visage, mollets, pieds...). Un tel découpage reste encore une exception et il est soumis à la

variabilité interindividuelle de la morphologie du cortex. Mais on sait que les aires auditives primaires (gyri temporaux supérieurs) ont une organisation tonotopique, chaque portion de l'aire correspondant à un segment de fréquence sonore (Ardila 1993). De même, les aires visuelles primaires (gyri occipitaux) possèdent une organisation rétinotopique, c'est à dire qu'elles reproduisent en quelque sorte les images imprimées sur la rétine.

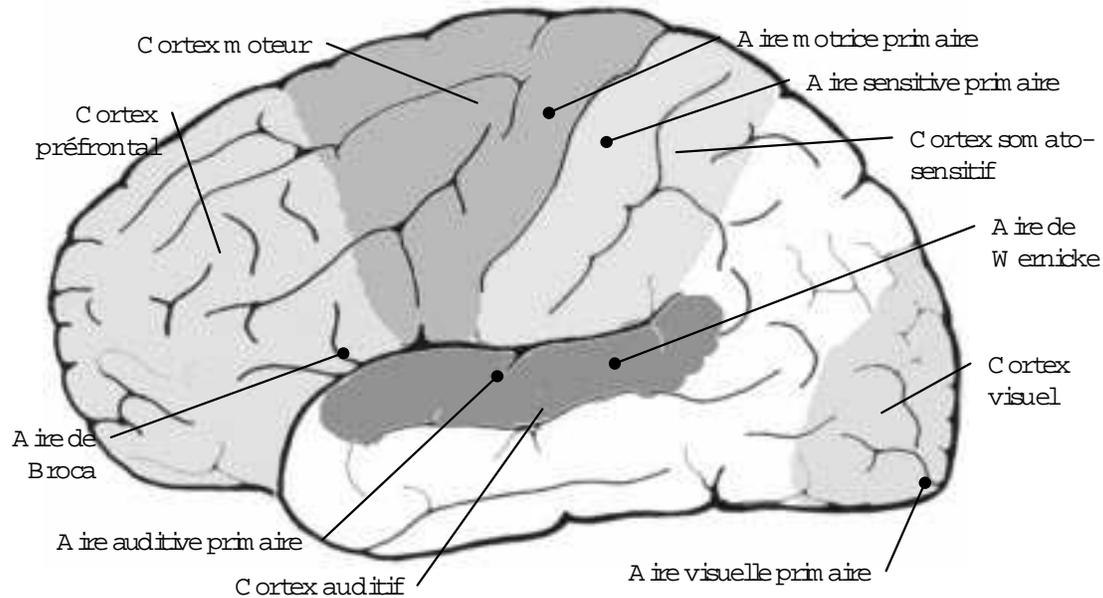


Figure 3 : Localisation d'aires fonctionnelles.

Les aires primaires sont, d'une manière générale, des zones fonctionnelles relativement bien localisées car elles sont les premières aires du cortex en relation avec les stimuli remontant des récepteurs sensoriels ou, pour la motricité, en relation avec les muscles. Ainsi, toute stimulation auditive activera l'aire auditive primaire ou, par exemple, une lésion d'une partie de l'aire motrice primaire entraînera une paralysie du segment corporel associé. On sait aussi localiser des aires impliquées dans des tâches plus évoluées, grâce, notamment, aux études cliniques ou en activation. C'est le cas, par exemple, des aires de Broca et Wernicke toutes deux impliquées dans les fonctions du langage et portant le nom des chercheurs qui, les premiers, ont su les identifier (Mesulam 1990).

La résolution d'une tâche cognitive met en relation de nombreuses aires du cortex. Chacune de ces aires, qui ne sont pas forcément voisines, joue un rôle bien précis, indispensable au traitement de la tâche. Les aires impliquées dans la tâche peuvent être mises en évidence par les techniques d'imagerie fonctionnelle. C'est à ce niveau de représentation que les observations sont faites. Les techniques d'imagerie fonctionnelles, conjuguées aux études anatomiques, mettent en évidence

l'organisation en réseau des aires, interconnectées par des faisceaux de fibres neuronales (Pulvermüller 1996, Fischbach 1992, Posner *et al.* 1988).

Une coupe du cortex peut, elle aussi, avancer une preuve anatomique de l'organisation du cortex en aires hétérogènes. Le cortex peut en effet être décomposé en six couches distinctes, c'est la *cytoarchitecture* du cortex. Dans chaque couche, la densité et l'organisation des neurones est différente. Les variations relatives de l'importance de ces couches, sur l'étendue du cortex, ont permis à Brodmann de dresser une cartographie du cortex, délimitant subjectivement 52 aires, connues sous le nom d'*aires de Brodmann* (Zilles 1996, Hedaen 1972). Ces variations sont une preuve anatomique d'une différenciation structurelle des populations neuronales. Le cortex cérébral est un substrat hétérogène. Ainsi, on peut s'attendre à un traitement localisé ou, du moins, différent selon l'endroit où il est effectué. Les zones fonctionnelles ne correspondent cependant pas toujours à une aire de Brodmann (Mesulam 1990). C'est par exemple le cas des aires de Broca et de Wernicke (Mesulam 1990). Ceci est dû à la technique de délimitation des aires de Brodmann qui doit être réalisée par une observation au microscope qui, compte tenu de la variabilité interindividuelle, reste sujette à l'appréciation de l'observateur (Zilles 1996).

La cartographie de Brodmann ne suit pas l'anatomie externe du cortex, les scissures et gyri. Chez le vivant, on se réfère donc à cette cartographie en utilisant un atlas neuroanatomique de référence, comme par exemple l'atlas de Jean Talairach. Compte tenu des différences morphologiques du cortex d'un individu à l'autre, l'usage d'un atlas est indispensable pour localiser les aires cérébrales. L'atlas de Talairach permet une localisation grâce à des coordonnées cartésiennes. Pour localiser une zone d'un cerveau donné, on doit se ramener, par des transformations géométriques, dans l'espace de l'atlas (Fox & Lancaster 1996). On peut, de cette manière, retrouver une aire de Brodmann, sans avoir à réaliser une coupe du cerveau étudié. L'usage de l'atlas est aussi, et surtout, un bon moyen pour lutter contre la variabilité interindividuelle de la morphologie du cortex.

1.2. Fonctionnement cellulaire

Les principales cellules du système nerveux sont les cellules nerveuses ou *neurones*. Ces cellules sont très particulières, tout d'abord de par leur structure, mais aussi de par leur activité chimique. Les neurones sont aussi les seules cellules de l'organisme qui ne se divisent pas. De ce fait, toute cellule détruite n'est pas remplacée. Ceci explique les pertes de facultés cognitives après une lésion cérébrale. Il est cependant possible que le cerveau lésé récupère progressivement un certain nombre de ses facultés. Cette récupération n'est néanmoins pas due à un remplacement des cellules

détruites mais à une réorganisation, limitée, des neurones intacts. Ce mécanisme de réorganisation est une preuve de la *plasticité* du cerveau. Cette plasticité est à prendre en compte lors d'une étude clinique car elle peut fausser l'appréciation du déficit.

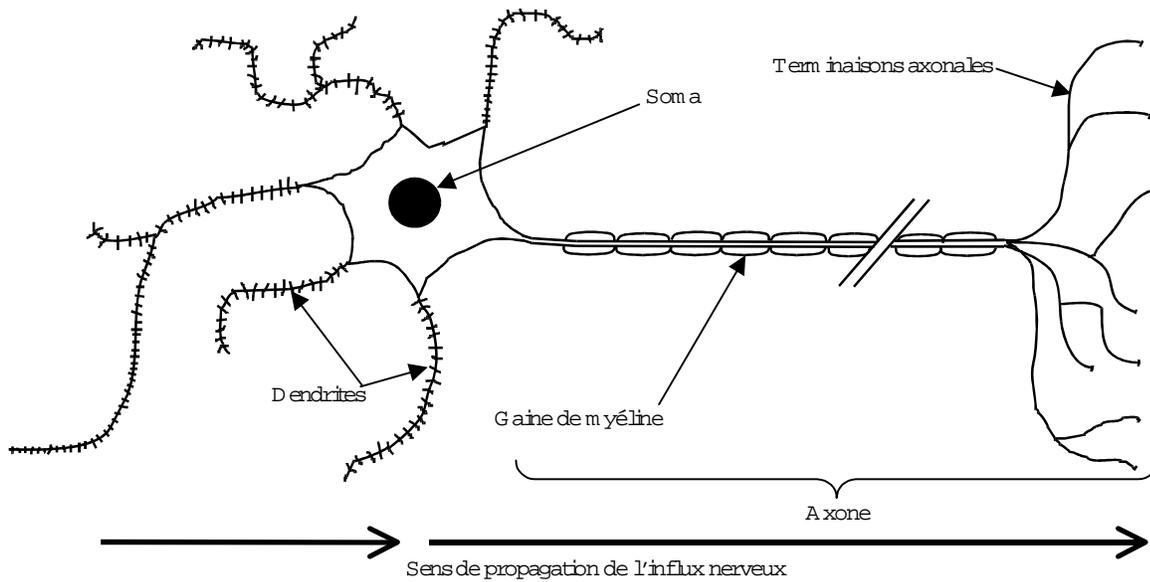


Figure 4 : Structure du neurone

La structure d'un neurone peut être divisée en trois parties, les *dendrites*, le *soma* et l'*axone* (Figure 4). Le soma est le corps du neurone, là où est situé le noyau de la cellule. Les dendrites sont des expansions qui entourent le corps cellulaire. Elles forment un panache que l'on appelle *arbre dendritique*. L'axone est aussi une expansion du corps cellulaire mais il est plus long que les dendrites et il se termine par de nombreuses divisions. Il n'y a qu'un seul axone par cellule. Chacune des extrémités des terminaisons axonales est renflée, c'est le bouton synaptique. Ces boutons se connectent aux dendrites d'autres neurones pour transmettre l'influx nerveux, le point de connexion est appelé *synapse*. Un neurone est en moyenne connecté à 10.000 autres neurones.

Cette division en dendrites/soma/axone est valable pour tous les neurones du système nerveux, tout comme le fonctionnement général, tel qu'il est décrit dans les paragraphes suivants. Comme on l'a déjà vu avec la cytoarchitecture, les neurones se différencient par leur morphologie (i.e. la taille et la forme de leur arbre dendritique et de leur axone) ainsi que par les connexions qu'ils établissent avec les autres neurones. Les caractéristiques structurelles interviennent donc dans la différenciation des rôles des divers neurones, au même titre que leurs fonctionnalités chimiques (i.e. susceptibilité à des neurotransmetteurs différents).

Le fonctionnement du neurone est de nature biochimique et électrique. Cette activité, caractérisée par un transport d'ions de sodium entre le milieu intra et extracellulaire, crée une impulsion électrique, ou *potentiel d'action*. Ce dernier se transmet des dendrites vers le soma. Puis, lorsque le potentiel interne dépasse un *seuil d'activation*, l'influx se propage du soma vers les terminaisons axonales, via l'axone. On dit alors qu'il y a décharge.

Le potentiel d'action circule à une vitesse variable, d'un neurone à l'autre, de 1 à 100 m/s le long de l'axone, ce qui est très lent comparé par exemple à la conductivité du cuivre. La différence de potentiel le caractérisant a une amplitude de 100 millivolts (mV) approximativement et est toujours égale pour un neurone donné. Au repos la différence de potentiel entre l'intérieur et l'extérieur de la cellule est de -70 mV, cette différence devient positive lorsqu'il y a activation. Les variations du champ électrique créé par l'activité des neurones, sont la mesure la plus directe de l'activité du cerveau. Pour transmettre l'influx à un autre neurone, le bouton synaptique libère des substances chimiques qui exciteront ou inhiberont le neurone adjacent. Il y a excitation lorsque le potentiel interne du neurone est rapproché du seuil d'activation (i.e. entrée d'ions dans le corps cellulaire). Il y a inhibition si le potentiel est réduit. Ces actions sont les deux seules actions possibles pour une synapse. Sachant qu'une synapse est soit inhibitrice, soit excitatrice, le champ d'action d'un neurone est directement fonction du nombre et de la qualité des synapses qu'il possède. On voit ici que la complexité du fonctionnement cérébral n'est pas due à la complexité du fonctionnement du neurone, même si des questions persistent (Fischbach 1992). C'est la grande interconnectivité des neurones qui rend l'exploration fonctionnelle quasi impossible à un niveau cellulaire.

Le transport du potentiel d'action nécessite une grande consommation d'énergie (i.e. oxygène, glucose et protéines). Or, les neurones sont des cellules qui ne possèdent pas de grandes réserves énergétiques et ont rapidement de grands besoins énergétiques lors d'une activité soutenue. Ce besoin se traduit par une augmentation du débit sanguin à proximité du neurone. Les variations du débit sanguin sont le second type de mesure que l'on peut avoir sur l'activité neuronale. Cette mesure est moins précise que la mesure du champ électrique, car on ne connaît pas la distance séparant le neurone du vaisseau sanguin. On ne connaît pas non plus le rapport entre l'activité du neurone et l'augmentation du débit, ni le délai nécessaire à l'organisme pour subvenir aux besoins des neurones (Mazoyer & Belliveau 1996, Sergent 1994).

Dans tout le système nerveux central (SNC), il est possible de différencier la *substance grise* de la *substance blanche*. La première, qui forme, entre autres, le cortex cérébral, est constitué pas les somas des neurones, de couleur sombre. Les axones, eux, forment la substance blanche. Ils sont

recouverts d'une gaine de substance isolante, la *myéline*, qui accroît la vitesse de conduction des axones. Cette organisation permet aux neurones de se connecter en des points du cortex très éloignés de l'endroit où se trouve leur soma. Les faisceaux d'axones de la substance blanche constituent autant de connexions entre les différentes aires du cortex. Ces connexions orientées sont la base des relations entre les aires fonctionnelles. C'est via ces connexions privilégiées que les aires échangent *l'information cérébrale* qui peut être matérialisée sous la forme de configurations (ou *patterns*) de décharges neuronales.

2. TECHNIQUES D'IMAGERIE FONCTIONNELLE

Il y a actuellement quatre techniques d'imagerie fonctionnelle utilisées. On peut aisément les classer en deux familles, les techniques de surface, et les tomographies. Comme on vient de le voir (cf. partie 1.2), il y a deux manières d'obtenir une évidence externe de l'activité du cerveau. On peut observer l'activité électrique des neurones ou la variation du débit sanguin engendrée par cette activité. La seconde méthode offre des localisations plus précises que la première qui, elle, possède une meilleure résolution temporelle. Les deux méthodes sont donc complémentaires.

2.1. Techniques de surface

Les techniques de surface, *l'électroencéphalographie* (EEG) et la *magnétoencéphalographie* (MEG), mesurent l'activité électrique des neurones. La première mesure directement le champ électrique produit par les potentiels d'action. La seconde mesure le champ magnétique créé par le champ électrique. Ces mesures sont réalisées à la surface du crâne par des capteurs (i.e. des électrodes pour l'EEG). Les valeurs mesurées sont très faibles, de l'ordre de la centaine de micro-volts ou du milliardième de gauss. Elles sont obtenues avec une résolution temporelle de l'ordre de la milliseconde.

Cette grande précision temporelle est le principal point fort de ces techniques. Leur point faible tient au fait que chaque capteur mesure l'ensemble de l'activité du cerveau depuis sa propre position. Chaque capteur obtient donc une mesure locale de l'activité totale du cortex et il est très difficile, voire impossible, de mesurer les signaux de structures trop profondes (e.g. thalamus). Dans le but d'améliorer ces mesures, on peut augmenter le nombre de capteurs ou, pour l'EEG seulement, on peut implanter les électrodes directement dans le cerveau. La méthode devient alors invasive et elle n'est employée que pour des raisons thérapeutiques, en préopératoire.

Sachant que le champ magnétique est directement produit par le champ électrique, on pourrait croire que ces deux techniques offrent des mesures identiques, mais ce n'est pas le cas. En effet, lorsque le champ électrique traverse les méninges et la voûte crânienne, il est distordu et partiellement absorbé. Le champ magnétique, lui, l'est bien moins, et ses mesures sont donc plus précises (Pernier *et al.* 1992). De plus les capteurs mesurent bien les champs dirigés vers eux, mais ils mesurent moins bien les champs qui leur sont perpendiculaires. Or, en schématisant, la mesure du champ magnétique est orthogonale à celle du champ électrique. Ainsi les mesures de la MEG peuvent compléter celles de l'EEG.

Les techniques de surface mesurent un signal « délocalisé », distordu et noyé dans la masse de l'activité spontanée du cerveau. Pour pouvoir observer les mesures liées à l'activité étudiée, que l'on appelle *réponse évoquée*, on applique la technique dite des grandes moyennes. Cela consiste à répéter l'expérience un grand nombre de fois (i.e. plusieurs dizaines voire plusieurs centaines de fois) pour obtenir beaucoup d'échantillons de mesures, calés sur la présentation des stimuli, et dont on fait la moyenne. L'activité concernant l'expérience se répète de manière synchronisée d'un échantillon à l'autre, alors que l'activité spontanée varie aléatoirement. Cette dernière va donc disparaître de la moyenne et la réponse évoquée va apparaître. Cette méthode contraint néanmoins le chercheur à développer un protocole expérimental dans lequel l'expérience peut être répétée un grand nombre de fois par le sujet.



Figure 5 : Projection d'un potentiel évoqué mesuré par EEG, (Thierry *et al.* 1998).

Une fois que l'on a obtenu des mesures claires des réponses évoquées, une par capteur, on peut interpoler les valeurs des capteurs puis projeter le tout sur un modèle 3D de la voûte crânienne (Figure 5). Les valeurs mesurées sont associées à une échelle de couleur. On peut alors observer la progression du champ électrique, ou magnétique, à la surface du crâne, au cours du temps. Cette observation permet de se faire une première idée des endroits du cerveau impliqués dans la résolution de tâches cognitives, mais surtout d'obtenir déjà une chronologie précise du traitement cérébral (Pernier *et al.* 1992).

Pour localiser la, ou les, zone(s) corticale(s) génératrice(s) du champ observé, l'affaire devient plus compliquée. Pour ce faire, on doit faire appel à des modèles physico-anatomiques. Le plus simple consiste à considérer les aires génératrices comme des dipôles électriques et à représenter la tête comme un ensemble de sphères concentriques (e.g. le scalp, le crâne, les méninges et le cerveau) de conductivités différentes. Il faut alors donner *a priori* le nombre et la position des dipôles impliqués. Le problème direct consiste à calculer le champ créé par les dipôles donnés. Le problème inverse consiste à faire corroborer le champ ainsi modélisé et le champ réellement observé. Pour ce faire, un logiciel va déplacer chaque dipôle pour réduire la différence entre les champs, recalculant à chaque étape le problème direct, le tout à chaque pas d'échantillonnage.

Tout le problème réside dans le fait que l'on doit initialiser le processus en plaçant *a priori* les dipôles. Or, il existe toujours plusieurs configurations produisant un même champ. On procède donc par incrémentation du nombre de dipôles. Ils sont placés à la position la plus probable, d'après l'expertise de la cartographie de surface. On peut arrêter le processus lorsque l'on obtient une configuration stable de dipôles. Cependant, lorsqu'il existe encore un doute sur la localisation des dipôles, le dernier recours est le bon sens de l'expert. Malgré ces problèmes de localisation, les techniques de surface restent indispensables pour obtenir une chronologie précise des événements cérébraux.

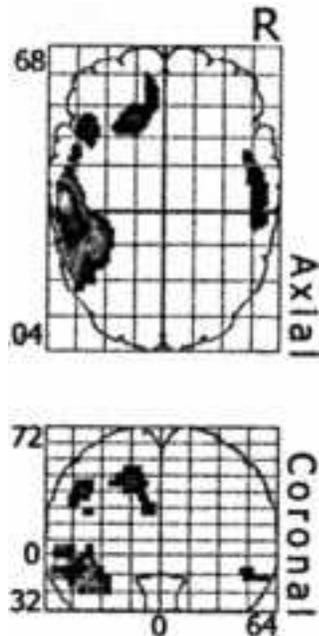
2.2. Tomographies

Les techniques de tomographie réalisent des images (*[radio]graphie*) du cerveau en coupes (*tomos*). On peut donc obtenir des images en profondeur de l'activité cérébrale. On utilise actuellement deux techniques récentes, la tomographie par émission de positron (TEP) et l'imagerie par résonance magnétique fonctionnelle (IRMf). Ces techniques ne mesurent pas directement l'activité du cerveau mais la consommation énergétique imputée à cette activité via la variation du débit sanguin cérébral. Cette approche est déjà en soi une limite pour une grande précision des mesures, car il y a de nombreuses inconnues (délai, amplitude, proximité, configuration du réseau sanguin) dans l'établissement du lien entre l'activité des neurones et les variations du débit sanguin. Ces techniques acquièrent des images avec des délais relativement grands et ont une résolution spatiale suffisamment faible, du moins par rapport à la taille des neurones. Ceci rend donc moins cruciale la connaissance du rapport entre débit sanguin et activité neuronale.

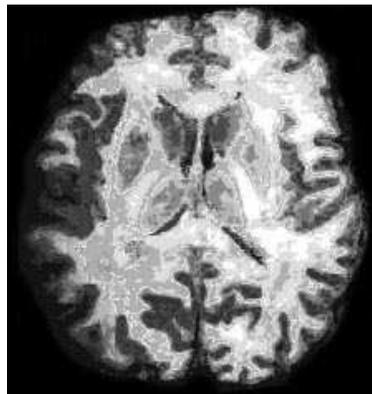
La TEP mesure la désintégration d'un marqueur radioactif injecté dans l'organisme. On injecte au sujet une substance marquée radioactivement, à des doses très faibles et donc considérées comme

inoffensives. Cette substance, par exemple de l'eau marquée à l'oxygène-15, mélangée au sang, va se répandre dans l'organisme du sujet et s'accumuler là où il aura une augmentation du débit sanguin. Les isotopes radioactifs émetteurs gamma utilisés ont une demi-vie assez brève (e.g. 123 secondes pour l'oxygène-15). Lors des réactions entraînées par la désintégration d'un noyau radioactif, deux photons antiparallèles sont émis.

On place le sujet, sous perfusion intraveineuse, dans la caméra à positron, et il y restera aussi immobile que possible durant toute la durée de l'expérience. Cette caméra est un cylindre, parsemé de capteurs, de manière à capter les photons des deux cotés. Il faut compter environ 40 secondes, selon l'isotope utilisé, pour obtenir une image en trois dimensions de l'ensemble du cerveau (Figure 6 a). Durant toute cette durée, la tâche du sujet doit être maintenue. La caméra ne relève que des couples de photons arrivant simultanément sur les capteurs. Après un traitement statistique des mesures, on peut calculer la position du site d'émission. La résolution spatiale obtenue est de l'ordre de 10 à 100 mm³ (un cube de 2.5 à 5 mm de coté).



a) Image TEP



b) Image IRM anatomique



c) Image IRM fonctionnelle

Figure 6 : Images tomographiques

(Images, extraites de Thierry *et al.* 1998 (a) et de travaux de G. Thierry (b, c), publiées avec l'autorisation de l'auteur)

L'IRM, quant à elle, utilise les propriétés de la résonance magnétique nucléaire. Il est en effet possible, en présence d'un champ magnétique, de faire entrer en résonance des atomes en leur envoyant une onde radio. La fréquence de cette onde est spécifique à chaque type d'atome.

Lorsqu'un atome entre en résonance, il absorbe l'énergie de l'onde incidente. Il est alors possible de détecter quelle est la concentration des atomes à l'intérieur d'un objet en corrélant la fréquence de l'onde envoyée et le volume d'énergie absorbé. On peut de cette manière obtenir des images tomographiques des tissus du cerveau (Figure 6 *b*). Cette technique est une technique anatomique (IRMa). L'IRM fonctionnelle (IRMf) se concentre uniquement sur les propriétés magnétiques de l'hémoglobine. Cette dernière ne possède pas les mêmes propriétés suivant qu'elle est liée à de l'oxygène ou non. On peut ainsi observer les différentes concentrations de l'oxygène transporté par le sang dans le cerveau (Figure 6 *c*, Mazoyer & Belliveau 1996). Cette méthode d'observation est appelée méthode BOLD.

Comme pour la TEP, le sujet doit rester immobile durant toute la durée de l'expérience. La résolution spatiale de l'IRMf est comparable à celle de la TEP. L'aspect totalement inoffensif de la technique (pas d'injection de produit radioactif) est souvent contrebalancé par l'inconfort du sujet et le bruit de la machinerie. Le véritable avantage de l'IRMf réside dans sa vitesse d'acquisition des images. On peut acquérir une image élémentaire (une coupe) environ toutes les 150 millisecondes, soit une image du cerveau (environ 10 coupes) en 1.5 secondes, ce qui représente un avantage certain sur la TEP. Cependant, même si ce délai est faible, il ne permet pas d'obtenir une chronologie des événements cognitifs. On peut, pour augmenter la vitesse d'acquisition, réduire le nombre de coupes, mais on réduit alors la taille de la zone observée. On se concentre généralement sur une bande de quelques centimètres à mi-hauteur du cortex. On augmente donc la fréquence d'acquisition des images au détriment du champ d'observation.

Les deux techniques tomographiques sont, comme les techniques de surface, soumises au problème de l'activité spontanée du cerveau qui entraîne une « variation spontanée du débit sanguin ». Il ne paraît pas raisonnable d'appliquer une méthode des grandes moyennes à ces techniques, car les durées d'acquisition des mesures, le confort du sujet et le coût de mise en œuvre de ces techniques limitent la durée des protocoles. On se contente donc de moyennes plus modestes.

D'une manière générale toutes les images sont d'abord normalisées suivant un atlas (e.g. atlas de Talairach), dans le but de réduire les différences interindividuelles. On transforme les images anatomiques de chaque sujet, obtenues par IRMa, par translation, rotation et homothétie pour les faire coïncider avec l'atlas. On applique ensuite les mêmes transformations aux images fonctionnelles (IRMf ou TEP). Les images d'une même série sont ensuite réalignées pour corriger les éventuels mouvements de la tête d'un bloc à l'autre. Pour pallier aux variations anatomiques, les images sont coregistrées, c'est à dire, qu'en prenant un sujet comme référence, les différences

anatomiques sont mises en correspondance. Les images sont finalement lissées pour terminer leur homogénéisation. Seulement alors, on peut procéder à l'analyse de la variance du signal dans une série de blocs d'images.

On peut alors appliquer une méthode de soustraction d'images, en utilisant une tâche de référence, qui peut être le repos. De cette manière, on fait ressortir les différences relatives d'activité pour une aire donnée suivant les tâches réalisées. En d'autres mots, on observe si telle aire est plus activée dans la première tâche que dans la seconde, ou inversement. Il faut remarquer que, même si la résolution spatiale dans une coupe est meilleure en IRMf (1 à 2 millimètres) qu'en TEP (5 millimètres), la TEP permet une acquisition volumique (jusqu'à 41 coupes) et à une sensibilité (rapport signal sur bruit) bien meilleure de l'IRMf.

2.3. Conclusion

Face aux caractéristiques complémentaires des techniques citées ci-dessus, de plus en plus d'études d'activation font appel à la combinaison de techniques de surface et de techniques tomographiques (e.g. EEG et IRMf (Alary *et al.* 1998), MEG et PET (Joliot *et al.* 1998), EEG et PET (Thierry *et al.* 1998)). On peut alors obtenir une corrélation spatiale et temporelle de l'activité cérébrale.

Une étude préalable en TEP peut ainsi aider à localiser les dipôles d'une étude en EEG. Cependant, ce genre de « double étude » reste très compliqué à mettre en œuvre car les contraintes des différentes techniques s'ajoutent. Les mesures ne pouvant être obtenues en même temps, il faut faire passer deux fois l'expérience au sujet et être certain de reproduire à chaque fois la même activité. Tous les protocoles expérimentaux ne peuvent pas se plier aux exigences combinées des deux types de techniques.

3. INTERPRETATION DES RESULTATS D'ETUDES D'ACTIVATION

Les études d'activation proposent des images de la *macro-organisation* du cerveau humain sollicitée par une tâche cognitive. Ces images mettent en évidence, soit l'ordonnement des événements cérébraux, soit l'ensemble des localisations du cerveau impliquées dans la résolution de la tâche. On ne peut pas observer à la fois l'ordonnement et la localisation des événements cérébraux. Les localisations observées en IRM et en TEP sont composées de milliers de neurones interconnectés.

On les appelle *aires fonctionnelles*. Elles interagissent via des faisceaux d'axones pour accomplir la fonction cognitive sollicitée. C'est l'activation de ces aires qui produit les événements cérébraux observés en EEG ou en MEG.

A l'intérieur de l'ensemble des aires activées, l'exécution du traitement n'est pas distribué au hasard. Chacune des aires impliquées possède un rôle qui lui est propre et complémentaire de celui des autres aires, comme le montrent les études cliniques ou les études animales (Mesulam 1990). Mais contrairement à ce que l'on pensait il y a encore quelques années, une seule aire n'est pas entièrement dédié à une unique fonction et, à l'opposé, le traitement n'est pas entièrement diffus (Pulvermüller 1996, Sergent 1994, Mesulam 1990, Posner *et al.* 1988). On peut, en fait, considérer cette activité comme l'activité d'un réseau où les processeurs sont les aires fonctionnelles et les transmetteurs sont les connexions anatomiques reliant les aires. On peut distinguer deux types de transmetteurs, les connexions de proximité, reliant deux aires contiguës, et les connexions distales, liant des aires éloignées via des faisceaux d'axones.

Chaque aire procède à un traitement qui lui est plus ou moins spécifique. Mais, pour une tâche déterminée, ce traitement n'a réellement de sens qu'au travers des connexions que l'aire entretient avec les autres aires activées. La structure du réseau des aires activées conditionne la façon dont la tâche est accomplie (Hickok & Poeppel 2000, Sergent 1994, Démonet *et al.* 1992). Mais une simple observation de ce réseau ne peut suffire à comprendre *comment* la tâche est résolue. Pour ce faire, il faut savoir quel est le rôle dévolu à chacune des aires le constituant et comment elles communiquent entre elles. Il faut donc représenter les mécanismes intra-aires du traitement de l'information cérébrale et les modes inter-aires de propagation de cette information.

Or, l'information de nature électrochimique, transitant au travers d'un faisceau d'axones, représente la décharge de certains des neurones à l'origine de ce faisceau. Il n'existe pas de moyen non invasif d'observer les neurones qui déchargent et ceux qui ne le font pas. La seule mesure possible, chez l'homme, est une mesure énergétique globale de cette information. Qu'elle soit obtenue par une technique de surface ou tomographique, cette valeur représente une valeur indirecte, distordue et moyenne. Autant de facteurs qui nous conduisent à considérer cette mesure comme une mesure fortement imprécise.

L'interprétation des images de l'activité cérébrale peut être exprimée par une modélisation computationnelle de l'activité cognitive (i.e. modélisation formelle, générant des modèles pouvant être mis en œuvre sur ordinateur, Mesulam 1990). Un modèle peut représenter le rôle supposé des différentes aires du réseau ainsi que des connexions les unissant. Une simulation informatique peut

aider à visualiser le comportement global du système et à retrouver, ou pas, l'activation qui était observée chez le vivant.

Cette interprétation peut s'exprimer en terme d'*hypothèses* sur les relations entre les aires activées ainsi que sur le fonctionnement propre à chacune d'elles (Pastor *et al.* 1997). C'est en réalisant des modèles de la propagation de l'information, que nous proposons d'exprimer ces hypothèses. Ces hypothèses s'appuient sur les observations et les connaissances de la structure cérébrale. Elle ne sont pas forcément compatibles entre elles, mais toutes peuvent s'exprimer en termes de traitement et de propagation dynamiques. Pour pouvoir comprendre et valider ces hypothèses, le formalisme sous-jacent doit permettre de les exprimer de façon explicite. Il faut que l'on puisse les identifier dans le modèle.

On peut définir deux sortes d'hypothèses : les hypothèses portant sur l'existence et les modalités de relations privilégiées entre les aires et les hypothèses définissant le fonctionnement spécifique de chaque aire. Les premières s'appuient sur les connaissances anatomiques dont elles dépendent. Nous les nommerons *hypothèses structurelles*. Nous devons de ce fait conserver une certaine plausibilité neuro-anatomique dans nos modèles pour pouvoir exprimer de telles hypothèses. Les secondes portent sur les fonctionnalités propres de chacune des aires cérébrales. Nous les nommerons *hypothèses fonctionnelles*. On doit pouvoir exprimer ces deux types d'hypothèses dans un même formalisme, de manière à conserver une homogénéité de représentation.

Nous verrons dans le chapitre suivant les différents formalismes communément utilisés pour représenter des modèles computationnels de l'activité cérébrale. Nous discuterons de leur aptitude à répondre aux considérations que nous venons d'énumérer.

MODELES COMPUTATIONNELS DE L'ACTIVITE CEREBRALE

1. INTRODUCTION

Les modèles computationnels de l'activité cérébrale ont suivi le développement des ordinateurs, du moins avec les travaux de Alan Turing dans les années 1950 (Pylyshyn 1989). A la même époque, John Von Neumann et Norbert Wiener posent les bases de la cybernétique, marquant ainsi le départ d'une nouvelle voie pour la modélisation du comportement cognitif, en s'inspirant du modèle du neurone proposé en 1943 par Warren McCulloch et Walter Pitts. Dès lors, l'approche computationnelle de la modélisation cognitive ne va cesser de se diversifier. Les premiers modèles du fonctionnement de l'esprit humain étaient fondés sur une comparaison littérale avec la machine de Turing. Ils sont maintenant abandonnés au profit de modèles plus proches de la réalité, mais ils restent encore une métaphore didactique.

Abandonné comme source directe de modèles, l'ordinateur reste un outil privilégié pour simuler le fonctionnement du cerveau. Ceci n'est que la manifestation de la propriété remarquable des ordinateurs à reproduire des systèmes dont l'architecture n'a rien de commun avec la leur (Rumelhart 1989). C'est pourquoi nous retiendrons sous le terme modèles computationnels tant les modèles qui tirent une analogie entre le fonctionnement d'un ordinateur et du cerveau que les modèles plus « mathématiques » suffisamment spécifiés pour pouvoir produire des simulations grâce à un ordinateur (Dayan 1994).

Les modèles computationnels de la cognition sont un champ tout à fait propice aux échanges interdisciplinaires. Ils sont, en effet, motivés par des recherches que l'on peut disposer sur deux axes croisés. Le premier axe regroupe un ensemble d'études de neurosciences dont le but est de comprendre le fonctionnement du cerveau et qui, pour y parvenir, utilise l'ordinateur. On les appelle de facto *neurosciences computationnelles* (McClelland & Plaut 1993, Anderson 1989). Sur le second axe, on peut placer des études dont le but est de faire résoudre, à des machines, des problèmes de la manière dont l'humain les résout (Alliot & Schiex 1994). C'est l'*approche cognitive de*

l'intelligence artificielle. En fait, cette dernière et l'approche computationnelle des neurosciences peuvent être entendues comme un point de jonction entre les disciplines auxquelles elles se rapportent. On en vient d'ailleurs souvent à les confondre, le seul trait les démarquant réellement étant celui qui nous a permis de les définir, à savoir que l'une cherche à étudier le fonctionnement du cerveau en tant que finalité alors que la seconde cherche à appliquer ce fonctionnement à un problème précis. En dehors de cette divergence, il est bien difficile d'observer d'autres traits distinctifs, c'est pourquoi les échanges interdisciplinaires sont ici très fréquents.

Un brillant exemple d'échanges interdisciplinaires est incarné par les réseaux de neurones formels (RNF). Ce formalisme est fondé sur un modèle du neurone réel (tel qu'il est décrit au chapitre 1, section 1.2) développé en 1943 par un neurophysiologiste, McCulloch, et un logicien, Pitts. Ce formalisme a été développé pour la cybernétique, puis il a été utilisé en intelligence artificielle. Cette utilisation a mis en exergue la faculté d'apprentissage de tels réseaux, notamment grâce à un algorithme dit de rétro-propagation². Puis, les réseaux de neurones ont été utilisés pour bâtir des modèles très proches des performances humaines. Actuellement, ce formalisme est le plus utilisé pour la modélisation cognitive, mais cela ne l'a pas empêché d'être critiqué, notamment sur le plan de la plausibilité neurologique des mécanismes mis en jeu par la rétro-propagation (Crick 1989). En effet, cet algorithme n'est pas fondé sur des connaissances physiologiques ; il est totalement abstrait. Ainsi, il ne peut être mis en avant qu'un réseau utilisant un tel algorithme d'apprentissage puisse apprendre comme un réseau biologique, et ceci même s'il présente les mêmes facultés. Pour y remédier des algorithmes biologiquement plausibles ont été développés (comme le « wake-sleep algorithm », Hinton *et al.* 1995, que nous verrons plus loin). Cette mésaventure illustre en fait une divergence entre les manières d'approcher le problème. La première manière consiste à faire *émerger* un fonctionnement humainement plausible en se basant donc sur des manifestations de ce fonctionnement, alors que la seconde manière cherche à *expliquer* le fonctionnement dans le modèle pour le reproduire. Dans ce sens, l'algorithme de rétro-propagation convient parfaitement à ceux qui veulent seulement faire émerger un comportement, mais pas à ceux qui veulent expliciter ce comportement. Pour nous y retrouver parmi les divers modèles et approches proposées, nous avons choisi de les séparer en deux familles, *l'approche fonctionnelle* et *l'approche structurelle*.

L'approche fonctionnelle, telle que nous la définissons ici, a pour but de reproduire le fonctionnement du cerveau humain au travers de ses propriétés fonctionnelles et caractéristiques

² Back-propagation en anglais.

intrinsèques. Pour y parvenir, les modèles incarnent des processus cognitifs et doivent reproduire les performances, ou déficits, propres au cerveau humain (ou animal) dans la résolution de la même tâche (McClelland & Plaut 1993, Rumelhart 1989). On met généralement en évidence les éléments fonctionnels qui composent la fonction cognitive et les interactions qui les caractérisent. Cette approche est une approche computationnelle assez proche de la psychologie.

L'approche structurelle s'appuie, quant à elle, sur l'organisation fonctionnelle du cerveau comme on peut l'observer, par exemple, par les moyens de la neuro-imagerie fonctionnelle (voir Chapitre 1 partie 2). Les fonctions cognitives doivent être reproduites et contraintes par une architecture biologiquement plausible (Koch 1993, Arbib 1985). On y met en interaction plusieurs éléments fonctionnels situés anatomiquement (e.g. des aires du cortex), dans le but de faire émerger la fonction cognitive étudiée (Laplaine 1994, McClelland & Plaut 1993, Arbib 1985). Cette approche s'appuie sur des observations anatomiques, ainsi que sur des résultats d'études d'activation, et est donc assez proche de la neuropsychologie.

On peut considérer que les deux approches, telles qu'elles sont définies ici, peuvent être complémentaires. Comme nous le verrons, il peut exister un point de jonction entre un modèle fonctionnel et un modèle structurel lorsque, par exemple, les éléments fonctionnels du premier peuvent être localisés sur les éléments structurels du second. Dans ce cas on a mis en évidence le lien entre structure et fonction.

Au vu des contraintes énoncées au chapitre précédent, on peut d'ores et déjà avancer que notre problématique se situerait naturellement dans une approche structurelle. Mais comme nous venons de le souligner, l'approche fonctionnelle est complémentaire de l'approche structurelle et elle présente, comme nous le verrons plus loin, certains aspects qui se rapprochent de notre problématique.

Le but de ce chapitre n'est pas de faire le tour des différents modèles proposés. Il s'agit essentiellement de mettre en évidence les apports des différentes approches et notions fondatrices des formalismes qui ont été développés pour la modélisation cognitive computationnelle. Nous étudierons donc les différentes propriétés des formalismes employés dans les deux approches sus-définies et nous illustrerons notre propos par quelques exemples précis de modélisation. Puis nous synthétiserons notre discours, pour situer le contexte de notre travail en déterminant, d'une part, dans quelle approche il s'inscrit et, d'autre part, quel formalisme répondrait le mieux à nos objectifs.

2. MODELISATION FONCTIONNELLE

L'approche fonctionnelle de la modélisation de l'activité cérébrale (i.e. la modélisation fonctionnelle) diffère, dans un sens, de la modélisation structurelle parce qu'elle fait une totale abstraction du substrat cérébral. Elle cherche une réponse au niveau du programme (la fonction) plutôt qu'au niveau de l'implémentation (Rumelhart 1989). C'est, sans doute, pourquoi la modélisation fonctionnelle s'est développée bien avant la modélisation structurelle, à une époque où les connaissances sur l'organisation fonctionnelle du cerveau étaient très limitées. Les premiers modèles fonctionnels datent des années 1950, et sont des travaux inspirés de la machine de Von Neumann.

La modélisation fonctionnelle est tributaire du découpage de l'activité cognitive en fonctions distinctes proposées par les études psychologiques. Elle n'est pas tributaire d'observations anatomiques ou neuro-fonctionnelles. Dans une approche fonctionnelle, on valide les modèles en comparant leurs performances à celles de l'homme. En d'autres mots, le modèle doit être consistant avec les capacités humaines dans la résolution d'un problème, tant dans ses qualités que dans ses défauts (Laplaine 1994, Reeke & Sporns 1993).

Ce découpage en fonctions cognitives n'interdit pas d'étudier les interactions entre les diverses fonctions mais il permet avant tout de concentrer un modèle sur une fonction unique. Il est donc possible de classer les modèles fonctionnels suivant la fonction cognitive qui les concerne (e.g. acquisition du langage, reconnaissance et production de mots, mémorisation et apprentissage, reconnaissance de formes, programmation et exécution de mouvements...). Il convient ici de s'intéresser surtout aux formalismes employés pour ce genre de modélisation. De ce point de vue, on peut aisément convenir d'une scission en deux groupes : les modèles symboliques et les modèles connexionnistes. Ces deux approches sont très différentes tant dans leur mise en œuvre, que dans les modèles qu'elles proposent. Elles ont toutes deux été critiquées et largement utilisées dans la modélisation de diverses fonctions cognitives.

2.1. Approche symbolique

L'approche symbolique est une approche de haut niveau conceptuel. Les symboles sont utilisés pour représenter les différents traits des objets du monde réel. Ce sont donc des entités interprétables sémantiquement. Sous ce point de vue, les processus cognitifs peuvent être associés aux règles formelles qui manipulent ces *symboles*. Ainsi, l'on parle de l'approche symbolique comme

d'une approche de haute transparence sémantique. On définit la transparence sémantique comme suit :

Définition 1 : Transparence Sémantique

On dira d'un système qu'il est transparent sémantiquement s'il est possible de décrire une correspondance précise entre une interprétation sémantique de la représentation interne (au modèle) et une description sémantique du comportement du système modélisé (Wallace *et al.* 1993, Clark 1989).

2.1.1. Les premiers modèles

Historiquement, on peut considérer le *general problem solver* (GPS) de Simon et Newell (Newell & Simon 1963) comme l'ancêtre des modèles symboliques computationnels. Selon cette théorie, les processus impliqués dans les tâches cognitives sont des processus élémentaires assemblés et organisés hiérarchiquement et exécutés séquentiellement. Les GPS étaient motivés par la résolution de problèmes dont on ne connaissait pas de solution algorithmique convenable. C'est la même motivation qui a conduit à l'élaboration des *systèmes experts*, quelques années plus tard. On peut considérer les systèmes experts comme une nouvelle génération de GPS, dont ils pallient la relative inefficacité.

Les systèmes experts illustrent, en intelligence artificielle, la différence entre une approche pragmatique et cognitive de l'intelligence. Etymologiquement un système expert cherche à reproduire les performances de l'expert (humain). Pour y parvenir, le système doit représenter un savoir souvent incomplet, et y appliquer des méthodes de résolution propres au savoir-faire de l'expert (Alliot & Schiex 1994). Le plus souvent, on choisit de représenter formellement le savoir par des symboles et le savoir-faire par des règles manipulant ces symboles. L'application du savoir-faire demande un *moteur d'inférences* qui permet, schématiquement, de produire un nouveau savoir. Cette manière de procéder est considérée comme une approche cognitive lorsque les représentations du savoir et du savoir-faire sont compatibles avec ce que l'on connaît de la représentation humaine de ces mêmes concepts. En fait, le modèle symbolique cognitif sera validé par comparaison à des observations psychologiques, en analysant ses performances et celles des humains (Pylyshyn 1989). Par exemple, un système expert utilisant une représentation fondée sur un formalisme logique classique (e.g. calcul des propositions ou logique des prédicats) ne peut être considéré comme cognitif car ses performances sont trop éloignées de celles d'un humain (Schank & Abelson 1977).

La recherche de cette *validité* cognitive a fait évoluer les modèles symboliques vers des systèmes incorporant des notions d'ordre psychologique dans leur formalisme même (Newell *et al.* 1989). Cette évolution s'est traduite par le passage de « représenter ce que nous connaissons et comment nous l'acquérons » à « représenter comment s'organise ce que nous connaissons » (Anderson 1989). La nécessité de cette évolution peut être illustrée par le déficit de la logique des prédicats où les connaissances sont exclusivement *déclaratives*. Or, le savoir humain est également caractérisé par l'habilité à savoir comment utiliser les faits observés. Ce savoir est appelé *procédural* (Omar 1994). Les psychologues ont montré que le savoir déclaratif était séparable du savoir procédural et que l'organisation du savoir procédural est différente de celle du savoir déclaratif. Cette différenciation est absente de la logique classique.

De nombreux systèmes ont été proposés, utilisant diverses représentations formelles. Nous citerons, par exemple, le système des Memory Organizer Packets (MOPs) de Roger C. Schank (Schank & Farrell 1988, Schank & Abelson 1977). Ce système, assez proche d'un système expert classique, utilise une représentation fondée sur la théorie de la dépendance conceptuelle (Alliot & Schiex 1994, Schank & Abelson 1977). Cette théorie permet de représenter le sens de phrases sous la forme d'un réseau de dépendances conceptuelles construit à l'aide de 11 primitives. Cette méthode assure l'unicité de la représentation de phrases sémantiquement équivalentes. De plus, les informations implicitement contenues dans la phrase sont explicitées dans le réseau. Les diverses connaissances sont organisées sous la forme de scripts qui permettent de replacer une connaissance dans son contexte. Un script est une séquence d'événements, décrits par des réseaux de dépendances conceptuelles, se rapportant à une scène particulière (e.g. un repas dans un restaurant). La détermination d'un script auquel se rapporte une connaissance partielle permet d'en déduire simplement toutes les connaissances sous-entendues. Les MOPs sont des structures dynamiques qui permettent de reconstruire un script. En s'appuyant sur une architecture de système expert classique, ce système est un bon exemple d'organisation cognitive de la mémoire épisodique.

2.1.2. La théorie ACT*

Nous allons nous intéresser maintenant à un système dont l'architecture est bien plus différente d'un système expert que le système des MOPs. John R. Anderson a proposé une théorie des mécanismes de mémorisation, appelée ACT, pour « Adaptive Control of Thought » (Reeke & Sporns 1993, Anderson 1989, 1983). Elle-même fondée sur le système HAM (pour « Human Associative Memory »), cette théorie a d'abord évolué en ACT* (Anderson 1983), puis a été reprise et étendue par le système PUPS (Anderson 1989).

ACT* est bâtie autour d'un système à base de règles. Son architecture prend en compte deux types différents de mémoire à long terme, une mémoire déclarative et une mémoire procédurale. Ces deux différents types de mémoire interagissent entre eux au travers d'une mémoire de travail, comme l'illustre le schéma ci-dessous (Figure 7).

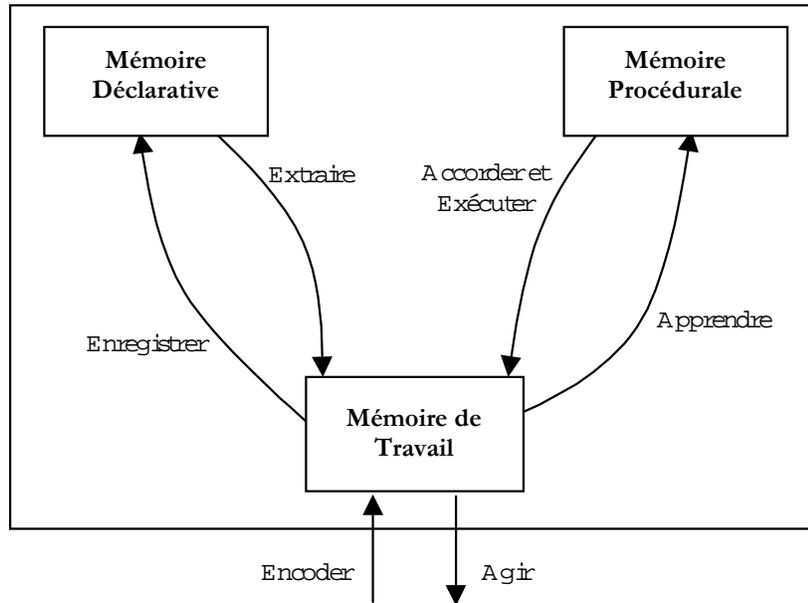


Figure 7: Architecture de la théorie ACT* d'après Anderson (1989).

Les éléments de base manipulés dans cette théorie sont appelés *unités cognitives* (Anderson 1983). Ils représentent, à un niveau conceptuel, les faits qui sont mémorisés par le système. Une unité cognitive est définie comme un nœud et un ensemble d'éléments. Le nœud peut être considéré comme le concept lui-même (e.g. « les chats attrapent les souris »), il est en tout cas le label de l'unité. L'ensemble d'éléments est un ensemble de traits reliés par le concept (e.g. « chat », « attraper », « souris »). Ces unités peuvent être organisés hiérarchiquement en reliant les éléments qui leur sont communs. Elles sont de plus associées à un poids qui décroît dans le temps si l'unité n'est pas réutilisée. La fonction d'atténuation des poids est identique pour chaque unité. Elle correspond, pour un poids P à l'équation suivante (1) :

$$(1) \quad P = \sum_{i=1}^n t_i^{-d}$$

Où n est le nombre de fois où l'unité a été utilisée, t_i est le temps écoulé depuis la $i^{\text{ème}}$ utilisation et d est un coefficient constant compris entre 0 et 1.

Les unités au poids élevé sont préférées aux unités aux poids plus bas. Le poids d'une unité qui est actuellement utilisée vaut $+\infty$ ($t_i = 0$). La courbe d'atténuation tend vers zéro lorsqu'une unité n'est plus utilisée, mais avec une pente de plus en plus faible. De ce fait, pour deux unités de même poids, à un instant donné, celle qui est plus ancienne (i.e. qui a été utilisée il y a longtemps) sera préférée à la plus récente car son poids décroîtra plus lentement.

Anderson différencie quatre sortes de mémorisations élémentaires dans la théorie ACT (Anderson 1989). La première est matérialisée par le renforcement des poids des unités stockées dans les mémoires déclarative et procédurale, lorsqu'elles sont réutilisées. Ce mécanisme ne modifie pas ce qui est globalement mémorisé par le système mais modifie l'ensemble des connaissances qui seront préférentiellement utilisées (i.e. les unités de poids le plus fort).

La deuxième sorte de mémorisation élémentaire est l'enregistrement en mémoire déclarative. Toutes les informations entrant dans le système sont placées dans la mémoire de travail (Figure 7). Les informations déclaratives placées dans la mémoire de travail peuvent être transférées de façon permanente dans une mémoire à long terme, la mémoire déclarative. Les informations stockées dans la mémoire déclarative pourront en être extraites de nouveau pour pouvoir être utilisées ultérieurement. Ce mécanisme de mémorisation, l'enregistrement dans la mémoire déclarative, est gouverné par un processus probabiliste. Toutes les informations ne seront pas mémorisées à long terme. De plus la probabilité d'enregistrement est constante au cours du temps et est égale pour toutes les informations contenues dans la mémoire de travail (Anderson 1983).

La troisième mémorisation élémentaire est un mécanisme de compilation qui permet de transformer une connaissance déclarative en une connaissance procédurale (Omar 1994, Anderson 1989). ACT propose deux manières de compiler les éléments mémorisés, par composition et par *procéduralisation*. La composition permet de comprimer plusieurs règles procédurales en une seule en court-circuitant les étapes intermédiaires. La procéduralisation transforme des règles procédurales générales en des règles spécifiques au problème à résoudre en éliminant les références aux éléments déclaratifs actuellement mémorisés.

Le dernier mécanisme de mémorisation élémentaire est la généralisation des connaissances procédurales, pour en créer de nouvelles. Ce dernier mécanisme a été remplacé, dans le système PUPS, par un mécanisme fonctionnant par analogies (Anderson 1989).

Le point fort de la théorie ACT est que chacune des 4 mémorisations élémentaires trouvent un écho favorable dans diverses études psychologiques (Anderson 1983). C'est cette validation qui

permet d'affirmer que cette théorie est un modèle plausible de la mémoire humaine. Cette théorie est aussi un bon exemple de décomposition d'une fonction cognitive en primitives fonctionnelles. On remarquera que c'est d'ailleurs sur le choix de ce découpage que l'on peut critiquer cette théorie. D'autres systèmes proposent en effet des modèles de la mémoire sans considérer le même découpage. C'est le cas du système symbolique SOAR (Newell *et al.* 1989), ou encore, plus largement, des modèles connexionnistes. Ces systèmes n'utilisent qu'un seul, voire deux mécanismes de mémorisation, soit par exemple, un mécanisme de compilation (uniquement par composition) pour SOAR (Newell *et al.* 1989, Laird *et al.* 1986) soit un mécanisme de renforcement de poids pour les modèles connexionnistes (McClelland 1994, Schneider 1993). Les autres mécanismes de mémorisation *émergent* dans le système.

Pour terminer, on soulignera que, comme tout modèle symbolique, ACT *suppose* que la pensée peut être représentée par un calcul formel sur des symboles. Cette hypothèse, *a priori* aisée à concevoir, est pourtant sujette à caution, dans le sens où la formalisation symbolique peut appauvrir le champ d'application de la pensée (Laplaine 1994, Reeke & Sporns 1993, Parks *et al.* 1991). On ne peut, de ce point de vue, considérer qu'un modèle symbolique modélise l'ensemble de la pensée humaine. Les mécanismes modélisés par les modèles symboliques n'en restent pas moins valides. Il convient donc de considérer les modèles symboliques comme des modèles ciblés sur des fonctions particulières (e.g. la mémoire) qui peuvent être proprement formalisées.

2.2. Approche connexionniste

L'approche connexionniste de la modélisation fonctionnelle a connu son envol assez tardivement, au début des années 1980, avec, par exemple, les travaux de James L. McClelland et David E. Rumelhart (McClelland & Rumelhart 1981). Ces travaux ont été baptisés « approche PDP », pour « Parallel Distributed Processing » (Parks *et al.* 1991), par opposition à l'approche symbolique. Le terme PDP est maintenant largement employé et au delà même des modélisations fonctionnelles. De fait, il devient un synonyme du terme connexionnisme, bien qu'à l'origine il se limitait à un usage de modèles connexionnistes pour la modélisation cognitive.

Cette approche est une modélisation fondée sur la notion des réseaux de neurones formel (McClelland & Plaut 1993, Ohayon 1990). Toutes ces dénominations sont quasiment identiques mais nous retiendrons ici la définition de l'approche PDP donnée par McClelland et Rumelhart (1986) :

Les modèles PDP supposent que le traitement de l'information est accompli au travers d'interactions entre un grand nombre de processeurs élémentaires, appelés unités. Chacune de ces unités envoie aux autres unités des signaux excitateurs ou inhibiteurs.

L'approche PDP utilise une technique de modélisation qui se rapproche du traitement cérébral qui, rappelons-le, est hautement parallèle et exécuté par des unités élémentaires simples fortement interconnectées (voir Chapitre 1, partie 1). Cette recherche d'une *consistance* avec le style de traitement employé par le cerveau est une réponse aux critiques de l'approche symbolique (McClelland & Plaut 1993, Parks *et al.* 1991). Cette approche ne nécessite, en effet, pas de formalisation de la connaissance qui se trouve plus ou moins distribuée dans le réseau, suivant l'architecture de celui-ci. Cette distribution de la connaissance entraîne en contrepartie une perte de transparence sémantique. L'information est bien moins accessible dans un modèle connexionniste que dans un modèle symbolique.

Avant d'entrer dans le détail de l'approche PDP, il convient de se demander dans quelle mesure on peut classer cette approche dans la modélisation fonctionnelle ou dans la modélisation structurelle, telle qu'elles ont été définies en introduction de ce chapitre. Comme nous le verrons dans la deuxième partie de ce chapitre, les modélisations structurelles font, elles aussi, largement appel aux réseaux de neurones formels. Cependant, même si le formalisme de base est le même, la méthodologie reste totalement différente. Il n'y a, a priori, aucune contrainte de plausibilité neurobiologique dans un modèle connexionniste fonctionnel. A posteriori, on peut toujours essayer de rapprocher les éléments du modèle fonctionnel d'éléments structurels identifiés dans le cerveau (Arbib 1985). Ce travail *a posteriori* sur le modèle fonctionnel est facilité par l'utilisation du PDP. Cette capacité est d'ailleurs souvent citée comme l'une de ses principales qualités (Parks *et al.* 1991). Il est pourtant rare que le lien entre structure et fonction soit directement atteint de cette manière.

2.2.1. Description de base des réseaux de neurones formels

Les RNF sont des graphes orientés connexes. Parmi toutes les unités du réseau, on en distingue de trois sortes : les unités d'entrée qui reçoivent les données externes au système, les unités de sortie qui envoient des données à l'extérieur du système, et les unités cachées qui sont des unités intermédiaires. Ces dernières sont invisibles de l'extérieur du système. Les stimuli fournis au réseau et les réponses de celui-ci sont des vecteurs de valeurs.

Les RNF sont différenciés par leur architecture, c'est à dire par les connexions qui lient les neurones entre eux. On peut, par exemple, citer les réseaux complets et les réseaux en couches. Les réseaux complets sont des réseaux dans lesquels chaque neurone est connecté à tous les autres neurones du réseau. Les réseaux en couches sont les réseaux les plus utilisés. Dans de tels réseaux, les neurones sont rassemblés dans des *couches*, la première couche rassemble les unités d'entrée et la dernière, les unités de sortie. Les couches intermédiaires, si elle existent, ne contiennent que des unités cachées. Tout élément d'une couche donnée ne peut se connecter qu'aux éléments de la même couche ou aux éléments de la couche directement supérieure (Alliot & Schiex 1994, Ohayon 1990). On différencie les réseaux à *deux couches*, ne possédant pas d'unités cachées, et les réseaux *multicouches*, possédant une ou plusieurs couches intermédiaires.

Le comportement du réseau repose sur la définition du neurone formel (Figure 8) qui modélise le comportement d'un neurone naturel (Ohayon 1990). Ainsi, à chaque instant t , l'état du neurone formel est caractérisé par (Alliot & Schiex 1994, Rumelhart 1989) :

- Un ensemble de poids w_i pour chaque connexion qu'il possède avec un neurone en amont dans le réseau.
- Une valeur d'activation $a(t)$. Cette valeur est obtenue par l'application d'une fonction, f_a , appelée règle d'activation.
- Une valeur de sortie, $o(t)$. Cette valeur est obtenue, elle aussi, grâce à une fonction, f_s , appelée fonction de sortie, ou règle de décision.

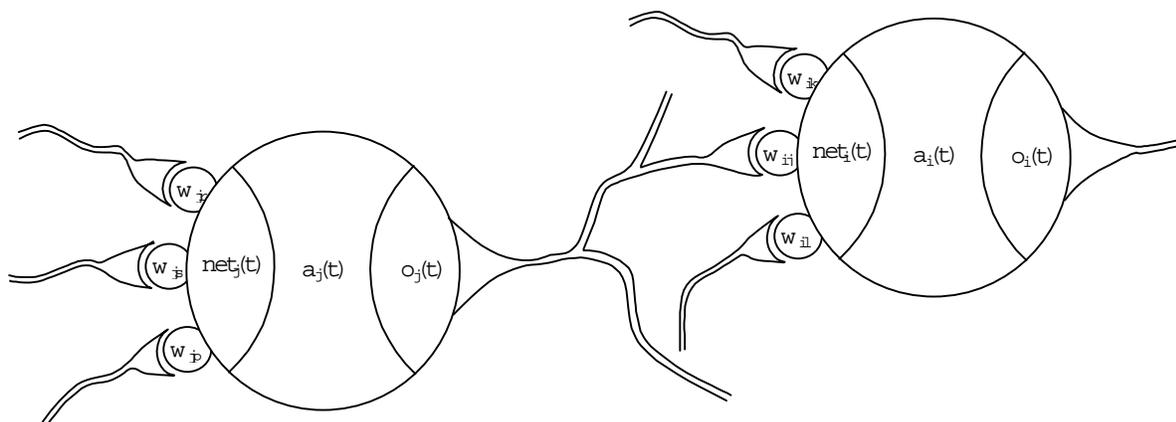


Figure 8 : Neurone formel d'après Rumelhart (1989).

A chaque instant, pour un neurone donné, connecté à n neurones en amont du réseau, on calcule :

- la somme des entrées $net(t)$, pour toutes les valeurs de sortie des neurones afférents, $o_i(t)$ à $o_n(t)$:

$$(2) \quad net(t) = \sum_{i=1}^n w_i \cdot o_i(t)$$

- la valeur d'activation du neurone :

$$(3) \quad a(t) = f_a(a(t-1), net(t))$$

- la valeur de sortie :

$$(4) \quad o(t) = f_s(a(t))$$

On procède de la sorte pour tous les neurones du réseau, des unités d'entrée aux unités de sortie. Le comportement du système réside autant dans le choix des connexions entre les unités, que dans celui des poids des connexions ou des règles d'activation et de décision des unités.

Un poids négatif sur une connexion fait baisser la valeur d'activation. La connexion correspondante agit donc comme une connexion inhibitrice. Au contraire, un poids positif représente une connexion excitatrice. Un poids nul équivaut à supprimer la connexion.

La règle d'activation peut être une fonction identité, dans ce cas $a(t) = net(t)$. Elle peut aussi prendre, par exemple, la forme d'une fonction sigmoïde et ainsi borner les variations de la valeur d'activation. La fonction de sortie peut, elle aussi, prendre la forme d'une fonction identité. Elle peut être une fonction stochastique, la valeur de sortie dépend d'une manière probabiliste du niveau d'activation. Mais elle a généralement la forme d'une fonction linéaire à seuil. Cette grande liberté dans les caractéristiques des neurones du réseau a entraîné la multiplication des architectures durant ces vingt dernières années.

Une architecture très simple, en couches, avec un minimum d'unités, a permis à James L. McClelland et David E. Rumelhart (McClelland *et al.* 1986, McClelland & Rumelhart 1981) de proposer un modèle de la perception de lettres dans les mots. Ils organisent leur modèle en trois couches. Chaque couche est associée à un niveau de représentation (Figure 9). La couche d'entrée est le niveau des traits des lettres, au sens visuel du terme. Par exemple, la lettre H se décompose en 3 traits : $|-|$, 2 traits verticaux et 1 trait horizontal. La couche intermédiaire représente les lettres. A ce niveau, les lettres sont recomposées à partir des traits qui les composent. La couche de sortie

représente le niveau des mots. Chaque unité de chaque couche est associée à un élément du niveau de représentation correspondant. Chaque unité de la couche intermédiaire est associée à une lettre pour chacune des positions que la lettre peut prendre dans le mot. Il y a par exemple une unité associée à la lettre A en première position, comme dans le mot ABLE, une unité associée à A en deuxième position, comme dans le mot TAKE, et ainsi de suite (la figure ci-dessous n'en présente qu'un sous ensemble dans un souci de lisibilité). Les connexions sont soit excitatrices, soit inhibitrices. D'une manière générale, l'inhibition est employée pour limiter l'activation des unités concurrentes, d'où les inhibitions intra-couches. Toutes les unités d'une couche sont connectées à toutes les unités de la couche suivante. Par exemple, l'unité associée à la lettre T en première position du mot, noté T₁ dans la figure ci-dessous, excite toutes les unités « mots » où T apparaît en première position, soit TIME et TAKE, et inhibe tous les autres mots.

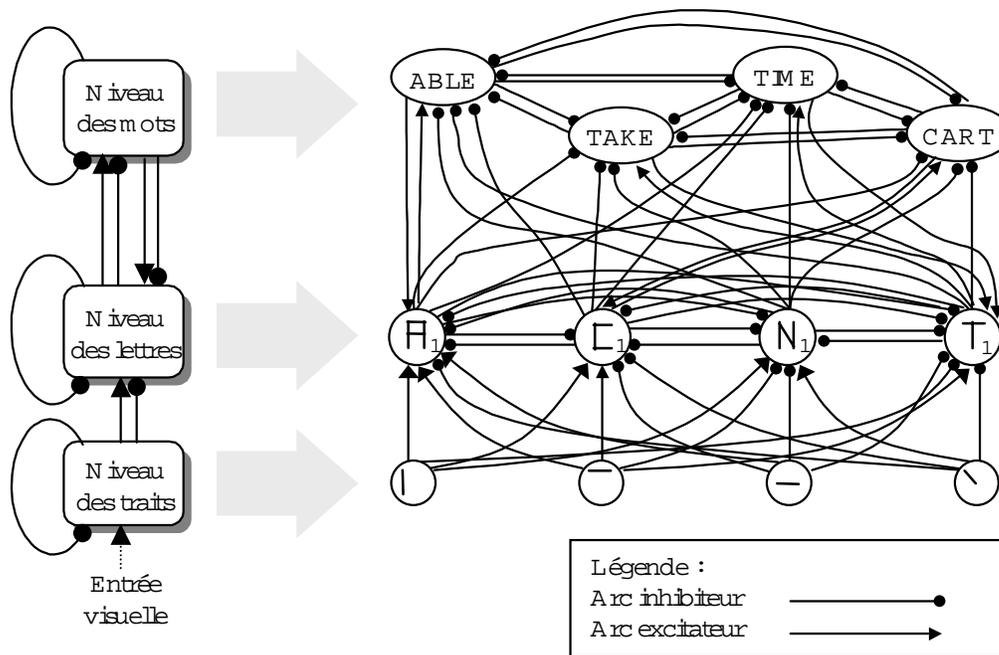


Figure 9 : Schéma simplifié du réseaux en couche de lecture de mots d'après McClelland et Rumelhart (1981).

Cette modélisation reste donc très simple dans son principe. Pourtant, le comportement du modèle reproduit de nombreux phénomènes propres à la lecture de mots, comme par exemple la faculté de complétion de mots partiellement lisibles. De plus le modèle commet des erreurs, qui, elles aussi, se rapprochent des erreurs commises par l'homme (McClelland *et al.* 1986, McClelland & Rumelhart 1981). On peut cependant reprocher à ce modèle le pré-cablage des connexions et l'étiquetage des unités qui limitent l'utilisation de ce modèle à une échelle réaliste. Ce modèle, contrairement aux systèmes symboliques de mémorisation décrits ci-dessus, est incapable

d'apprentissage. Cependant, cette incapacité n'est due qu'à l'objectif de la modélisation (la lecture « routinière »).

2.2.2. Apprentissage

L'une des propriétés des RNF est la possibilité de changer les valeurs des poids des connexions au cours de la simulation. C'est ce seul mécanisme qui est à la base de la faculté d'apprentissage des réseaux de neurones. Là encore, on a le choix entre plusieurs méthodes d'apprentissage.

a) L'apprentissage supervisé par rétro-propagation des erreurs

La première méthode d'apprentissage, la plus répandue et la plus critiquée (Crick 1989), est la méthode de rétro-propagation des erreurs (Rumelhart *et al.* 1986). Cette méthode est une méthode supervisée qui consiste à soumettre des stimuli au réseau et à essayer de faire coïncider sa réponse avec la réponse attendue. Si les deux réponses diffèrent, il faut « remonter » les connexions, de la sortie vers l'entrée du réseau, pour en modifier les poids. Les modifications à apporter sont fournies par une loi d'apprentissage. Cette loi peut être considérée comme une variante du concept d'apprentissage de Hebb (Brown *et al.* 1990, Rumelhart 1989) dont l'idée de base est la suivante. Si un neurone afférent est activé alors que le neurone considéré doit l'être aussi alors on augmente le poids de la connexion. Si le neurone afférent est activé alors que le neurone considéré ne doit pas l'être, alors on réduit le poids de la connexion. La version simplifiée, qui est le plus souvent utilisée, prend la forme suivante (Alliot & Schiex 1994, Rumelhart 1989) :

$$(5) \quad \delta w_i(t) = \varepsilon \cdot a(t-1) \cdot o_i(t-1)$$

Où, δw_i est la variation à appliquer au poids considéré, ε est le coefficient d'apprentissage, a est la valeur d'activation du neurone considéré et o_i est la valeur de sortie du neurone afférent correspondant.

On définit aussi une seconde règle, appelée règle-delta (Alliot & Schiex 1994, Brown *et al.* 1990, Rumelhart 1989), qui permet d'éviter les oscillations. Cette dernière loi ajoute un coefficient de moment, η , à la première règle d'apprentissage :

$$(6) \quad \delta w_i(t) = \varepsilon \cdot a(t-1) \cdot o_i(t-1) + \eta \cdot \delta w_i(t-1)$$

Les deux paramètres, coefficient d'apprentissage et de moment, permettent de faire varier la *plasticité* et la *stabilité* du réseau. Ces deux propriétés s'opposent. Dans le cas où le réseau est stable, il ne peut pas incorporer facilement un nouveau stimulus (non encore mémorisé). Il doit subir une nouvelle phase d'apprentissage par répétition de la présentation du stimulus, au risque d'ailleurs d'oublier des stimuli appris précédemment. Le réseau n'est pas plastique. A l'opposé, si le réseau est plastique, le nouveau stimulus est facilement incorporé. Dans ce cas, les poids des connexions sont rapidement modifiés, et ce à chaque nouvelle présentation. Ces modifications entraînent des changements incessants dans l'organisation du réseau, le réseau est instable.

Le principal avantage de ce mécanisme d'apprentissage reste sa facilité de mise en œuvre. Il a prouvé son efficacité dans de nombreuses modélisations (Rumelhart 1989, McClelland & Elman 1986, Rumelhart *et al.* 1986) et reste encore aujourd'hui très utilisé. Il soulève toutefois des questions concernant la compatibilité de ce mécanisme avec le fonctionnement de l'esprit humain. La rétro-propagation du réseau pose tout d'abord le problème de l'existence de son implémentation dans le cerveau humain (Crick 1989). Ce problème est tout à fait secondaire pour une modélisation fonctionnelle (i.e. non contrainte par la structure du cerveau). Du point de vue purement fonctionnel, l'aspect le plus embarrassant de la rétro-propagation est son fonctionnement supervisé. La supervision est indispensable dans le mécanisme de rétro-propagation pour diriger l'apprentissage du réseau. On ne peut donc pas considérer que l'apprentissage du réseau soit proche de l'apprentissage humain qui est essentiellement spontané.

En réponse à ces observations, divers mécanismes d'apprentissage auto-organiseurs et non supervisés ont été proposés. Nous en commenterons trois, les réseaux de Kohonen, les réseaux ART et l'algorithme « wake-sleep ». Ces mécanismes sont autant de preuves de la grande liberté laissée au concepteur de réseaux de neurones par le formalisme de base des RNF.

b) Les réseaux de Kohonen

Les réseaux de Kohonen (Figure 10) sont des réseaux à deux couches. L'apprentissage s'y faisant par auto-adaptation aux stimuli, il n'est pas nécessaire de présenter un ensemble d'exemples pour que le réseau apprenne. Chaque unité d'entrée est connectée à toutes les unités de sortie. Les unités de sortie ne sont pas toutes connectées entre elles, de manière à pouvoir définir une distance inter-unités (i.e. la norme choisie est la longueur du chemin le plus court entre deux unités).

Pour chaque stimulus présenté, une seule unité de sortie s'active. Le principe est de classifier les entrées en activant la même sortie pour les entrées suffisamment semblables. L'apprentissage

consiste alors à améliorer la discrimination du réseau. La loi d'apprentissage renforce les poids des connexions dirigées vers les unités de sortie proches de l'unité activée, et atténue ceux des unités éloignées (Alliot & Schiex 1994). Au fur et à mesure de la présentation de nouveaux stimuli, une classification va apparaître : une même unité de sortie va s'activer pour plusieurs stimuli différents appartenant à la même classe. Ce qui est intéressant, mais dangereux aussi, c'est que la classification émergente est autonome. Elle peut donc ne pas correspondre à ce que l'on pouvait attendre (Ohayon 1990).

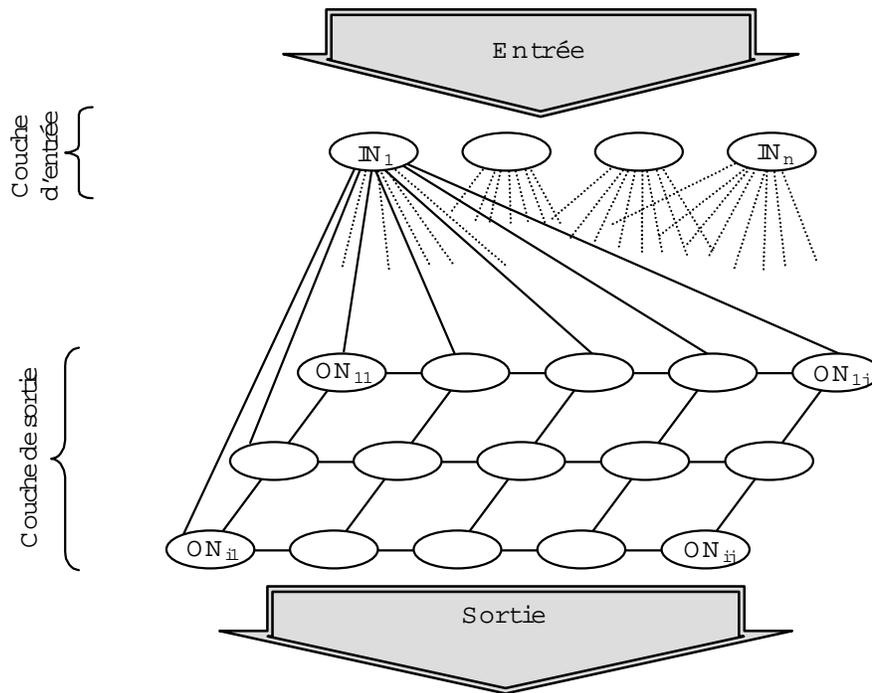


Figure 10 : Réseau de Kohonen, d'après Alliot et Schiex (1994).

c) Les réseaux ART

Les réseaux ART, pour « Adaptive Resonance Theory », ont été développés par Carpenter et Grossberg (1993). L'apprentissage est obtenu par un mécanisme de résonance entre la couche d'entrée et la couche de sortie. En effet, chaque unité de la couche d'entrée est connectée à toutes les unités de la couche de sortie et vice versa. De plus, chaque unité de sortie est connectée à toutes les autres unités de sortie par des connexions inhibitrices. Le renforcement des poids ne concerne que les connexions liant des unités activées. Quand une unité de sortie est activée, elle inhibe ses voisines, les empêchant ainsi de s'activer, et renvoie une activation vers la couche d'entrée pour renforcer son activation lors de la présentation d'un stimulus semblable à celui qui est présenté actuellement. Outre le fait que le mécanisme de résonance soit non supervisé, il permet aussi de

résoudre le problème de la plasticité et de la stabilité. En effet, et pour schématiser, l'activation dans le sens entrées-sorties permet la mémorisation du stimulus (plasticité). L'activation dans le sens sorties-entrées renforce l'activation de l'unité d'entrée lors de la prochaine présentation du même stimulus (stabilité).

d) L'algorithme Wake-sleep

Le dernier mécanisme d'apprentissage non supervisé que nous survolerons a été baptisé « Wake-sleep algorithm » (Hinton *et al.* 1995). Cet algorithme est destiné à des réseaux à multicouches, contrairement aux deux mécanismes d'apprentissage précédents. Dans un apprentissage par rétro-propagation, le superviseur définit la configuration de sortie du réseau. Cette configuration, représentant le but à atteindre, permet de déterminer si la réponse effective est erronée ou non et donc de déterminer les modifications à apporter aux poids des connexions. Ici, le but à atteindre, déterminé a priori, n'est pas exprimé comme une configuration de sortie, mais plutôt comme un ensemble de contraintes telles que : « apprendre les représentations dont la description est la plus économique mais permettant de reconstruite parfaitement le stimulus » (Hinton *et al.* 1995).

Comme pour les réseaux ART, un réseau utilisant l'algorithme wake-sleep doit posséder des connexions croisées entre les différentes couches. Le réseau fonctionne en deux phases. La première phase est une phase ascendante (de l'entrée vers la sortie), où les connexions impliquées sont appelées connexions de reconnaissance. C'est la phase « Wake ». Le but de cette phase est de reconstruire une représentation du stimulus au travers des couches d'unités cachées. La seconde phase est une phase descendante, où les connexions impliquées sont appelées connexions génératrices. C'est la phase « sleep ». Le but de cette phase est de reconstruire une approximation du stimulus à partir de sa représentation dans les couches cachées.

Ces trois derniers mécanismes d'apprentissage non supervisés partagent avec les méthodes supervisées la caractéristique de contraindre l'architecture du modèle, mais ils restent plus complexes à mettre en œuvre. Il faut cependant remarquer que ils peuvent être considérés comme plus biologiquement plausibles que l'apprentissage par rétro-propagation. Ils se soustraient donc aux critiques énoncées ci-dessus et ils sont d'ailleurs utilisés dans des modélisations structurelles que nous verrons dans la seconde partie de ce chapitre.

2.2.3. Transparence sémantique et RNF

Tout comme les modèles symboliques, les modèles connexionnistes ne peuvent être considérés comme des modèles globaux du fonctionnement de l'esprit humain. Ils sont ciblés sur une fonction spécifique. Ce qui fait leur particularité, c'est leur faculté à faire émerger des comportements semblables au comportement humain tant sur les performances que sur les erreurs (Hinton *et al.* 1993, McClelland & Rumelhart 1981). Une autre caractéristique importante est la robustesse des RNF. Il est possible d'endommager un réseau en détruisant certaines de ses unités. Malgré cette « lésion », le réseau continuera à fonctionner mais certainement de façon altérée. Ce qui est intéressant ici c'est qu'une fois endommagé, le réseau fera émerger des aberrations semblables à celles que l'on pourrait observer chez un patient possédant un déficit de la fonction cognitive modélisée par le réseau de neurones (Dehaene & Changeux 1997, Litvan *et al.* 1996, Hinton *et al.* 1993).

Avant de conclure la partie de ce chapitre traitant des formalismes utilisés dans la modélisation fonctionnelle, nous allons revenir un instant sur la transparence sémantique des réseaux de neurones. On parle souvent des réseaux de neurones comme des systèmes « boîte noire », c'est à dire de systèmes dont on ne sait comment le traitement est réalisé en interne. On peut imaginer la difficulté à établir un lien entre le concept et la représentation interne en prenant l'exemple du codage des stimuli. Les données en entrée (input) d'un RNF sont des vecteurs de valeurs numériques représentant la valeur d'activation de toutes les unités d'entrée. Ce vecteur prend souvent la forme d'un vecteur de valeurs binaires (0 pour non activé, 1 pour activé). Suivant les connexions des unités d'entrée, les différents traits représentés dans le vecteur vont se distribuer dans le réseau. Moins l'architecture sera structurée et moins il sera évident de suivre le cheminement de l'information.

Les réseaux les moins structurés sont les réseaux de Hopfield. Un tel réseau est un graphe complet, chaque unité est connectée à toutes les autres. Il n'y a plus aucun câblage spécifique et toute l'architecture du réseau est conditionnée par son apprentissage. Les réseaux de Hopfield sont vraisemblablement les modèles possédant le moins de transparence sémantique.

A l'opposé, les modèles comme celui utilisé par McClelland et Rumelhart pour la perception des lettres possèdent une architecture très contrainte et il est possible d'identifier le rôle de chacune des couches (Figure 9), et même de chacune des unités. D'une manière générale, plus le modèle est contraint dans son architecture plus il est possible d'identifier ses composantes fonctionnelles. Il est même possible de construire un modèle connexionniste composé de plusieurs RNF identifiés et

interagissant entre eux (au travers de connexions de neurones formels). C'est, par exemple, une telle organisation que propose Walter Schneider pour modéliser l'organisation de la mémoire de travail (Schneider 1993). Chaque réseau est organisé en couches et est formellement assigné à un type de mémoire particulier (e.g. mémoire lexicale, sémantique, visuelle...).

2.3. Conclusion

Pour conclure sur la modélisation fonctionnelle, nous pouvons mettre l'accent sur la complémentarité des approches symboliques et connexionnistes. La première s'appuie sur l'hypothèse que la connaissance peut être représentée par des symboles. La seconde propose une distribution de la connaissance, une certaine diffusion parmi de nombreux processeurs. Il en résulte que l'approche symbolique produit des modèles fonctionnels, d'une haute transparence sémantique, où les fonctionnalités sont représentées explicitement. L'approche connexionniste propose des modélisations où les propriétés sont plus ou moins implicites et émergent du modèle. Si les modèles symboliques offrent indéniablement une plus grande transparence sémantique, les modèles connexionnistes n'en offrent pas moins une plus large gamme. En fait, on peut difficilement modifier la transparence sémantique d'un modèle symbolique, alors que l'on peut augmenter ou réduire celle d'un réseau de neurones en contraignant son architecture (Wallace *et al.* 1993).

On peut ajouter que les modèles symboliques supposent que le but de l'apprentissage est, en quelque sorte, de formuler des règles. Cette hypothèse sous-entend que le savoir est exprimable par des symboles, mais elle permet grâce au sens associé aux symboles de produire des généralisations très puissantes. Ceci limite les modèles symboliques à être utilisés dans des domaines où l'on sait expressément représenter la connaissance par des symboles, soit, pour des fonctions cognitives de haut niveau, comme le langage (e.g. l'organisation de phrases dans un discours).

Les modèles connexionnistes supposent que le but de l'apprentissage est de permettre au modèle d'agir comme s'il connaissait les règles et ce sans avoir à les formuler explicitement (McClelland *et al.* 1986). Les règles émergent du comportement du modèle mais n'y sont pas représentées. Cela permet aux modèles connexionnistes de représenter des fonctions cognitives dont on ne sait pas exprimer les mécanismes formellement, comme par exemple des fonctions de plus bas niveau (e.g. la lecture des caractères composant un mot). Mais, comme l'étendue des connaissances représentables est proportionnelle au nombre d'unités du RNF, ces modèles deviennent très complexes lorsqu'on veut représenter une large gamme de connaissances.

L'aspect complémentaire des deux approches a été utilisé par J. G. Wallace qui propose un formalisme hybride de la modélisation fonctionnelle (Wallace *et al.* 1993). L'idée de base est ici d'organiser la modélisation sous la forme d'un réseau sémantique où les nœuds possèdent une représentation symbolique et une valeur d'activation. Les valeurs en entrée du réseau sont des symboles construits à partir des valeurs numériques fournies par les capteurs. La valeur d'activation permet d'activer des chemins sémantiques de manière préférentielle et de représenter des phénomènes d'attention. L'activation des nœuds est conduite par un réseau non symbolique appelé réseau de motivation (Wallace *et al.* 1993). L'union des deux formalismes est ici un compromis entre le caractère explicite de la représentation symbolique et l'émergence fonctionnelle des modèles connexionnistes.

3. MODELISATION STRUCTURELLE

La modélisation structurelle est plus récente que la modélisation fonctionnelle. Ceci est essentiellement dû au fait que cette démarche est inspirée par des connaissances poussées sur l'organisation du cerveau. La validation du modèle dans ce type de modélisation est directement issue d'études d'activation ou de connaissances neuroanatomiques ou neurophysiologiques fines issues de l'expérimentation animale. Le développement des modèles structurels a donc suivi celui des connaissances sur la structure du cerveau (voir Chapitre 1).

Dans une modélisation purement fonctionnelle, le choix d'un traitement semblable à celui du cerveau n'est pas indispensable, comme le prouve l'approche symbolique. Ici, il est inconcevable de conserver une organisation qui ne soit pas distribuée. Les RNF se sont donc naturellement imposés dès que des contraintes biologiques ont été posées. Une grande majorité des modèles structurels font appel à ce formalisme. Cependant, alors que les critiques sur la rétro-propagation pouvaient être considérées comme secondaires dans une optique fonctionnelle (Crick 1989), elles limitent sérieusement l'utilisation de ce mécanisme d'apprentissage dans un modèle structurel. On y préfère donc des mécanismes comme, par exemple, les réseaux ART (Carpenter & Grossberg 1993).

Pour pouvoir approcher la complexité du cerveau humain, d'autres formalismes connexionnistes ont été mis en œuvre. Ils sont plus ou moins proches des RNF décrits dans la partie précédente. Avant de décrire différents formalismes développés, nous allons définir quatre niveaux d'organisation du cerveau dans lesquelles nous pourrions classer les diverses modélisations structurelles.

3.1. Différents niveaux d'organisation du cortex

Il est possible de distinguer quatre niveaux d'organisation dans le cortex cérébral (Burnod 1991, Ohayon 1990). Les modélisations structurelles peuvent être distinguées suivant le niveau organisationnel qui les concerne. À chacun de ces quatre niveaux correspond une approche différente de la modélisation structurelle. Les niveaux organisationnels sont présentés ici par ordre croissant d'intégration. Les modèles associés cherchent à reproduire le comportement correspondant à la structure considérée (Burnod 1991, Arbib 1985).

3.1.1. Le niveau cellulaire

Ce premier niveau est celui des neurones, dont les modèles formels sont des simplifications poussées, en particulier en ce qui concerne les capacités de mémorisation. Selon les neurobiologistes, les mécanismes de mémorisation au niveau du neurone naturel peuvent se décomposer en 4 processus (Burnod 1991) :

- (i) Le délai de transmission de l'information le long de l'axone et le délai de transmission synaptique représentent déjà une mémorisation de l'information. La durée de cette mémorisation est très brève, de l'ordre d'une à cent millisecondes.
- (ii) L'accumulation de calcium modifie la transmission en fonction de l'activité antérieure du neurone. L'activation du neurone est alors facilitée. Ce mécanisme permet une mémorisation à court terme de l'information, par une activité répétitive, de l'ordre de quelques secondes à quelques minutes.
- (iii) Une modification des récepteurs de la membrane de la cellule et des canaux ioniques permet une mémorisation à plus long terme, un jour ou plus.
- (iv) De fortes co-activations répétitives et la maturation cérébrale (stabilisation des contacts synaptiques) permettent une mémorisation permanente.

De ces quatre processus, les neurones formels ne retiennent que l'idée de modification de l'efficacité de la transmission synaptique et le rôle de la co-activation dans cette modification (Ohayon 1990). De plus, le fonctionnement des synapses est bien plus complexe que le rôle de simple poids d'une connexion. Des modèles plus complets du neurone peuvent donc être utiles pour comprendre le rôle fonctionnel des divers aspects de ces cellules. On peut, par exemple, rendre compte des mécanismes d'intégration synaptique des arbres dendritiques en prenant en

compte des réactions électro-biochimiques (Mel 1993). Cependant, à ce niveau d'organisation on ne peut pas parler d'une véritable modélisation cognitive mais plutôt de modélisation physiologique.

Les modèles cellulaires sont des modèles numériques construits le plus souvent sur la base d'études animales (Schutter 1992, Segev 1992). La complexité mathématique de ces modèles ne permet pas de les utiliser comme base d'une modélisation à un niveau plus intégré. Un modèle qualitatif du neurone a été proposé dans le but de diminuer la complexité inhérente aux modèles numériques (Hunter *et al.* 1997). La mise en œuvre de ces modèles comme unité d'un modèle connexionniste entraînerait toutefois un accroissement de la complexité des modèles, par rapport aux neurones formels *classiques*.

3.1.2. Le niveau modulaire

Le deuxième niveau est celui des modules, populations neuronales élémentaires ayant un rôle fonctionnel bien déterminé (Arbib 1985). Le niveau modulaire le plus souvent pris en compte est celui des colonnes corticales. Les neurones du cortex sont organisés en six couches morphologiquement distinctes (voir Brodmann, Chapitre 1, partie 1.1), numérotées de I à VI. Une colonne corticale est une structure perpendiculaire à la surface du cortex, de surface très réduite et comprenant dans sa hauteur les neurones des six couches. Les fonctions et les densités relatives de ces colonnes varient d'un endroit à l'autre du cortex. Mais certaines propriétés sont récurrentes, comme des relations inter-colonnes, excitatrices ou inhibitrices, une activité motivée par la recherche d'un état de stabilité ou des mécanismes d'amplification (Guigon *et al.* 1994, Burnod 1991). Toutes ces propriétés sont implémentées dans les relations que les différentes couches entretiennent entre elles et avec l'extérieur (i.e. des colonnes voisines ou appartenant à d'autres zones du cerveau plus lointaines).

De nombreux modèles de colonnes corticales ont été construits (Frezza-Buet & Alexandre 1998, Guigon *et al.* 1994). Emmanuel Guigon a notamment proposé un modèle où les différentes couches des colonnes corticales sont représentées par des réseaux de neurones formels. Ce modèle reprend l'organisation en couches de la colonne, et représente les connexions inter-couches (intégration verticale) et intra-couches (interaction horizontale de filtrage). Les interactions intra-couches réalisent un traitement en entrée de la colonne. Les interactions inter-couches interviennent ensuite pour réaliser l'intégration de l'information entre les différentes couches de la colonne. Cette intégration est la combinaison de deux processus, la combinaison des résultats de la première étape et l'interaction non linéaire entre les paires de couches. Ce modèle propose deux

méthodes d'apprentissages (Guigon *et al.* 1994) et a été utilisé pour modéliser des architectures de niveaux supérieurs (décrits ci-après). Les règles d'apprentissage permettent à un réseau de *colonnes formelles* de converger vers un but à atteindre, ou d'acquérir une mémoire de travail pour coordonner une séquence d'actions.

3.1.3. Le niveau régional

Le troisième niveau est le niveau régional (ou niveau du tissu). Les interconnexions entre les colonnes corticales permettent de les intégrer dans un cadre plus large en considérant les réseaux de colonnes comme des zones corticales (Burnod 1991). Les délimitations de ces régions sont assez imprécises. Ce niveau de représentation peut aussi intégrer les amas de neurones extra-corticaux (e.g. ganglions et thalami). De plus, les fonctions de chacune de ces aires restent mal connues, seules certaines sont bien définies (e.g. les cortex primaires). Mais d'un point de vue général, on peut dire que les régions s'organisent par un traitement distribué qui permet un apprentissage structuré.

Des modèles spécifiques de *certaines* régions ont été construits. C'est le cas par exemple des aires visuelles primaires qui ont été très bien étudiées chez l'animal (Koch 1993). I. J. Mitchell a proposé un modèle fonctionnel des noyaux de la base (Mitchell *et al.* 1991). Il insiste sur le fait que la fonction précise des ganglions de la base est relativement peu connue. La structure anatomique et la nature de l'information traitée au sein des noyaux sont elles aussi peu connues. L'approche utilisée est donc une approche connexionniste (i.e. RNF à apprentissage supervisé), ce qui permet de ne pas spécifier explicitement le rôle des unités du modèle. En réalité, le modèle est composé de trois RNF identifiés structurellement parmi les ganglions. Les RNF ont une architecture en couches, ils subissent une phase d'apprentissage. Le comportement du modèle peut être observé dans le but d'établir de nouvelles hypothèses sur le fonctionnement des ganglions.

3.1.4. Le niveau global

Le niveau global considère l'organisation coopérative entre les différentes régions du cortex et les autres tissus neuronaux (e.g. cervelet, thalami, ganglions...) à l'échelle de l'ensemble de l'encéphale. On parle alors de réseaux neuronaux à grande échelle. C'est à ce niveau que s'implémentent les fonctions de plus haut niveau cognitif. C'est aussi ce niveau organisationnel qui est mis en évidence par les études d'activation.

Les modèles concernant le niveau global représentent des traitements souvent simplifiés à l'extrême, pour en réduire la complexité (Dehaene *et al.* 1998, Monchi & Taylor 1998). On peut citer en exemple l'utilisation d'un réseau ART pour modéliser des interactions entre le cortex temporal inférieur et l'hippocampe³ (Carpenter & Grossberg 1993). Les interactions des deux structures sont impliquées dans diverses fonctions cognitives comme l'apprentissage, la reconnaissance des formes et la mémorisation. Les propriétés d'apprentissage non supervisé du réseau ART sont donc ici indispensables à la modélisation. On observe que si l'on endommage la partie du modèle associée à l'hippocampe, alors le modèle reproduit certains symptômes de patients amnésiques ayant subi une lésion dans le lobe temporal médian.

3.2. RNF et modélisation structurelle

Tous les modèles énoncés précédemment doivent faire face à la dualité suivante. Aux niveaux cellulaire et modulaire, si la structure est bien connue notamment grâce aux expériences menées sur l'animal, peu de données existent sur la fonction, ou alors cette dernière est décrite en termes biologiques et non cognitifs. Dans l'impossibilité de construire un modèle réaliste, les RNF peuvent être utilisés en tant que simplification extrême des cellules. Aux niveaux supérieurs (i.e. régional et global), l'activité est plus facilement observable mais le système devient trop complexe d'un point de vue structurel. A ces niveaux organisationnels, une modélisation structurelle, fondée sur un modèle individuel du neurone, pourrait nécessiter un nombre d'unités comparable au nombre de neurones de la structure cérébrale modélisée. Ceci représenterait, pour une colonne corticale de 1 mm² de surface, quelques 100.000 unités (Sakai *et al.* 1997). Quand bien même un tel modèle pourrait être traité par un ordinateur, les données anatomiques sur les connexions, ainsi que les données physiologiques sur l'ensemble des neurones eux-mêmes, ne seraient pas suffisantes pour spécifier les propriétés de toutes les unités (Malot & Giannakopoulos 1996).

On peut aussi soulever une seconde objection à l'utilisation des RNF tels qu'ils ont été définis dans la partie précédente de ce chapitre. Cette objection concerne encore une simplification de la formalisation. En effet, les neurones formels ne représentent pas l'axone du neurone réel. L'axone n'implique pas seulement un délai à la transmission, mais il indique aussi la direction de l'ensemble des connexions du neurone. L'arbre des terminaisons axonales privilégie les connexions de proximité. C'est à dire que, si un neurone A possède une connexion vers un neurone B, alors il y a une forte chance que les voisins du neurone A soient aussi connectés à B. De telles connexions sont

³ Partie du lobe temporal médian

très nombreuses et jouent un rôle important dans les colonnes corticales (Malot & Giannakopoulos 1996). Pour représenter ces connexions dans un RNF classique, il faut lier manuellement tous les neurones formels qui sont considérés comme proches, ce qui augmente encore la complexité du modèle.

3.3. Formalismes dédiés à la modélisation structurelle

On peut évoquer différentes sortes de solutions aux problèmes décrits ci-dessus. La première solution consiste à réduire la représentation de l'information à son strict minimum, en la codant, par exemple, dans un espace réduit de valeurs. De cette manière, l'information est traitable par un nombre beaucoup plus réduit d'unités (Dehaene *et al.* 1998, Monchi & Taylor 1998). On peut alors utiliser des neurones formels classiques, puisque l'on a ramené la complexité du traitement à réaliser au niveau de simplification leur correspondant.

L'autre solution consiste à faire évoluer le formalisme connexionniste pour qu'il puisse se rapprocher de la réalité. Cela peut se traduire, dans un premier temps, par une augmentation de la capacité du système informatique implémentant le modèle. C'est ce que propose le système NEXUS (Sakai *et al.* 1997).

L'idée de base de ce système est d'implémenter un modèle connexionniste sur un ensemble d'ordinateurs interconnectés entre eux, via un réseau local (LAN). Chaque ordinateur prend à sa charge la modélisation d'un *module* du modèle (e.g. une zone corticale). Pour compléter le manque de données sur les paramètres du modèle, le système fait des approximations topographiques avec des relations de proximité. Pour organiser de tels réseaux, on peut vouloir tester la faisabilité de l'approche sur un modèle à une échelle moins grande. Dans ce modèle *test*, on pourra, par exemple, remplacer les modules par des RNF de petite taille. Dans le système NEXUS, grâce aux approximations topologiques (Figure 11), le passage à la modélisation d'envergure est quasi automatique (Sakai *et al.* 1997). Les unités sont regroupées, selon leur configuration de connexions et leur fonction, en groupes, appelés plans (maps). Les connexions entre les plans sont représentées par des projections topographiques qui permettent, avec un minimum de paramètres, de représenter l'ensemble des connexions entre les unités des deux plans.

Cette approche peut être utilisée comme une prolongation d'une modélisation simplifiée. Elle apparaît en définitive comme une augmentation des RNF classiques puisqu'elle permet, à partir

d'un RNF, d'obtenir un modèle qui prend en compte les connexions de proximité, grâce aux approximations topologiques.

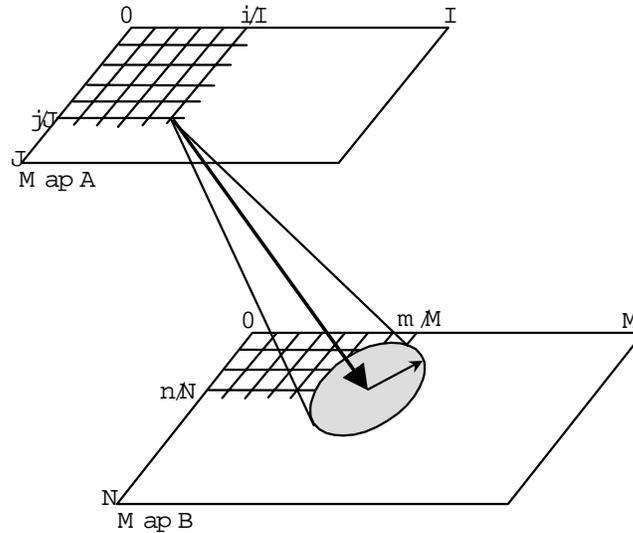


Figure 11 : Connexion entre deux plans définie par une projection topographique, d'après Sakai et al. (1997).

On peut s'approcher davantage d'une modélisation semblable à celle du cerveau, tout en limitant la complexité des modèles. L'approche consiste ici à ne plus modéliser des neurones individuels mais des *populations de neurones* (Malot & Giannakopoulos 1996).

La formalisation des populations de neurones considère une telle population comme un plan orthonormé (map) dans lequel un neurone est localisé par ses coordonnées. Les plans ont une taille limitée mais on les définit sur \mathbb{R}^2 . On considère de ce fait une certaine continuité entre les divers neurones de la population, prenant ainsi en compte les relations de proximité.

Dans le cas d'une connexion binaire d'un plan \mathcal{A} vers un plan \mathcal{B} , la somme des entrées pour un neurone x du plan \mathcal{B} , à l'instant t , devient, $net(x)$:

$$(7) \quad net(x) = \int_{\mathcal{A}} W(x, x') o(x') dx'$$

Où : \mathcal{A} est le plan afférent au plan \mathcal{B} , $o(x')$ est la valeur de sortie du point x' du plan \mathcal{A} , et W est appelé fonction *noyau*. Elle tient lieu de poids de la connexion entre \mathcal{A} et \mathcal{B} . Cette fonction prend souvent la forme suivante :

$$(8) \quad W(\mathbf{x}, \mathbf{x}') = w \cdot (\mathbf{x} - \mathbf{x}')$$

Où w est un paramètre constant correspondant au poids de la connexion. Dans ce cas, l'équation (7) devient une convolution, la connectivité de tous les neurones du plan A est uniforme.

Dans le cas où les fonctions noyau prennent en compte une variance spatiale et où les connexions font intervenir plusieurs plans afférents, les équations deviennent plus complexes. Cependant, un tel formalisme permet des traitements analytiques massifs (de plan à plan). Il réduit le nombre de spécifications avec une seule équation pour tout le plan. De plus, il tient compte des relations de proximité entre les neurones d'un même plan. Il convient donc parfaitement à la modélisation de zones corticales où les couches neuronales sont très homogènes et où la propagation de proximité est très développée.

Un autre formalisme de population de neurones propose une approche sensiblement différente. Ce formalisme est appelé « *Wave theory* » (Koch & Leisman 1996). Il utilise aussi une représentation continue des couches corticales. L'élément différentiel de ce continuum ne représente pas un neurone mais un ensemble aggloméré de plusieurs neurones. A ce niveau de représentation, les neurones peuvent être remplacés par des cellules figuratives purement excitatrice (e-cell) ou inhibitrices (i-cell). Dès lors, il existe 4 sortes de connexions : e-cell vers e-cell (e-e), i-cell vers e-cell (i-e), etc. (e-i et i-i). Ces connexions sont probabilistes. La probabilité de chaque type de connexion décroît avec la distance entre les cellules considérées. Ceci néglige l'existence de connexions anatomiques entre des aires distales du cortex. Les signaux sont représentés sous forme de vagues (waves) se propageant le long de la couche et au travers de connexions avec d'autres couches. Les vagues peuvent entrer en résonance, créant ainsi une sorte de mécanisme de mémorisation. Ce concept de vague est intéressant car il rappelle les signaux observés en EEG (voir chapitre 1, partie 2.1). Un modèle de base du cortex cérébral a été construit à partir de cette théorie (Koch & Leisman 1996). La totalité du cortex est représentée par seulement deux couches, étendues sur la totalité de sa surface. Dans ce modèle, les régions du cerveau ne sont pas clairement délimitées (elles n'ont pas de frontières). Elles sont caractérisées par des paramètres de propagation différents (Koch & Leisman 1996). Leur localisation est donc peu précise (comme en EEG) et difficilement interprétable sur le plan fonctionnel.

4. CONCLUSION

Représenter des aires cérébrales activées dans la résolution d'une tâche, à la manière dont une étude d'activation permet de l'observer, classe notre problématique dans le cadre d'une modélisation structurelle. En effet, notre premier souci est de contraindre notre modèle à représenter un réseau d'aires cérébrales, ce qui correspond au niveau organisationnel global (quatrième niveau selon (Burnod 1991)). De ce point de vue, le traitement doit nécessairement être parallèle et distribué. Ce type de modélisation devrait donc nous amener naturellement à une approche connexionniste.

Cependant, le but de notre formalisme est de construire des modèles dans lesquels les hypothèses fonctionnelles sont exprimées explicitement et où elles peuvent être identifiées et individualisées. Une modélisation explicite revient alors à une modélisation de haute transparence sémantique. La transparence sémantique traduit en effet le degré d'énonciation des propriétés des modèles. Ceci nous conduirait donc à une approche symbolique.

Ces observations peuvent être synthétisées en considérant un point qui n'a pas été explicitement discuté jusqu'à présent : la représentation de l'information. A la manière des niveaux organisationnels, il est possible de définir un degré d'abstraction de l'information cérébrale dans les modèles computationnels. Ce degré d'abstraction suit très exactement le degré de transparence sémantique du modèle. Dans le cerveau, l'information est représentée par l'état biochimique et électrique de quelques dizaines de milliards de cellules. Dans un modèle connexionniste, il prend la forme d'une matrice de poids de connexions et d'un vecteur de valeurs d'activation. C'est un degré d'abstraction déjà plus élevé et, de fait, les réseaux de neurones formels s'adaptent mieux au niveau modulaire qu'au niveau cellulaire. Dans un modèle de population de neurones, la représentation prend une forme continue. Les matrices et vecteurs sont représentés par des plans de neurones (maps). Cette représentation convient donc à la modélisation d'un signal se propageant à la surface d'une structure cérébrale continue, tel qu'il est recueilli par l'EEG. Cependant, ce signal est difficilement interprétable en terme d'information puisqu'on ne peut lui attribuer de contenu sémantique, exprimable par la catégorie de cette information (e.g. la configuration de neurones qui déchargent pour un stimulus donné). Dans un modèle symbolique, en revanche, on peut représenter l'information par un symbole. L'abstraction est alors plus élevée, mais elle s'accommode mal du caractère continu et imprécis des variations de l'énergie de l'information cérébrale (i.e. le nombre de neurones qui déchargent pour un stimulus donné).

La représentation de l'information est conditionnée par le degré de transparence sémantique que l'on souhaite obtenir. Plus la transparence se voudra élevée, plus l'information devra être accessible pour pouvoir être manipulée par les divers éléments du modèle. Il paraît en effet difficile d'exprimer des hypothèses de façon explicite sans pouvoir accéder à l'information qu'elles manipulent.

L'exigence d'un haut degré d'abstraction de l'information cérébrale (i.e. une bonne transparence sémantique) nous empêche de nous satisfaire d'une approche connexionniste classique. Il nous appartient donc d'unifier un traitement parallèle et distribué, à une représentation explicite de l'information. Pour ce faire, nous devons nous intéresser à des formalismes de haute transparence sémantique manipulant une information assimilable à l'information cérébrale, c'est à dire une information ayant à la fois un aspect continu et imprécis (l'énergie) et un niveau purement symbolique (la catégorie).

FORMALISMES REPRESENTANT DES CONNAISSANCES IMPRECISES ET
APPROCHE CAUSALE.

1. INTRODUCTION

Compte tenu des observations du chapitre précédent, les choix méthodologiques qui s'imposent à nous peuvent se résumer à un compromis entre un haut niveau de transparence sémantique et une plausibilité neurobiologique convenable. La plausibilité est nécessaire pour exprimer des hypothèses issues de l'observation de la structure fonctionnelle du cerveau. Elle nous contraint à un traitement de l'information physiologiquement plausible, c'est-à-dire un traitement parallèle, distribué et non supervisé. Ceci nous interdit une approche fonctionnelle symbolique classique telle que celles que nous avons décrites au chapitre précédent. La transparence sémantique est, elle, nécessaire pour exprimer explicitement les hypothèses modélisées. Sans une représentation explicite, il est impossible de les identifier et de les contrôler. Cette transparence ne nous permet pas d'opter pour une approche connexionniste classique. Une dernière contrainte vient de la nécessité de représenter l'imprécision de l'information, entraînée, d'une part, par l'incertitude du traitement neuronal (i.e. décharges probabilistes) et, d'autre part, par la faiblesse des mesures de cette activité.

Dans ce chapitre, nous verrons des formalismes de haute transparence sémantique et qui permettent de représenter des modèles de connaissances incomplètes. Par connaissance incomplète, on entend que l'information manipulée est partiellement connue ou que ses transformations ne dépendent pas nécessairement de lois précisément définies. Nous présentons les principes de trois grandes familles de formalismes permettant de modéliser de telles connaissances, le calcul des probabilités, la logique floue et le raisonnement qualitatif. Nous présenterons ensuite, d'une manière transversale, l'approche causale de la modélisation qui offre, dans chacun de ces formalismes, les capacités nécessaires à une représentation distribuée du traitement de l'information.

2. LE CALCUL DES PROBABILITES

Historiquement le calcul des probabilités s'intéresse à la représentation des chances d'obtention d'un événement incertain (Arnold & Guessarian 1994). Cette théorie décrit le modèle mathématique d'expériences aléatoires dont le résultat détermine l'événement produit.

2.1. Evénement probabiliste

Formellement, un événement est un ensemble de résultats d'une expérience particulière. Considérant une expérience E , on peut donc définir un événement A comme :

$$A = \{ w / A \text{ est réalisé si } w \text{ est le résultat de l'expérience } E \}$$

L'espace des résultats possibles d'une expérience est appelé *espace des épreuves*, noté Ω . Par exemple, considérons l'expérience qui consiste à lancer un dé à 6 faces et à regarder quelle face est obtenue quand le dé s'immobilise. L'espace des épreuves de cette expérience est $\Omega = \{\text{face 1, face 2, face 3, face 4, face 5, face 6}\}$. Et l'on peut définir les événements : « la face 1 est obtenue », soit $\{\text{face 1}\}$ ou « la face obtenue est inférieure à 4 », soit $\{\text{face 1, face 2, face 3}\}$. On note T , l'ensemble de tous les événements. L'événement $A = \Omega$ est appelé événement certain.

On peut composer les événements au moyen d'opérations ensemblistes. Ainsi, une conjonction d'événements, pour une même expérience, est obtenue par leur intersection (\cap). De même une disjonction est obtenue par l'union des événements (\cup). L'événement complémentaire d'un événement A , noté \bar{A} , est l'événement se réalisant lorsque A ne se réalise pas. Il est défini comme : $\bar{A} = \Omega - A$.

2.2. Probabilité

La notion de probabilité correspond à la notion intuitive de fréquence de réalisation (Arnold & Guessarian 1994). Une probabilité exprime en quelque sorte le degré de certitude que l'on a de voir un événement s'accomplir. Une probabilité P est une application de T vers $[0, 1]$ telle que :

$$P(\Omega) = 1$$

$$\forall A \in \mathcal{T}, 0 \leq P(A) \leq 1$$

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \text{ si et seulement si } A_1 \cap A_2 = \emptyset$$

On déduit de cette définition le principe du tiers exclu, spécifiant que $P(\bar{A}) + P(A) = 1$.

Cette définition de la probabilité est suffisamment contrainte pour définir, sous certaines conditions, la probabilité d'un événement lorsqu'un autre événement est connu. C'est-à-dire, $P(A/B)$, la probabilité conditionnelle de l'événement A sachant l'événement B.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

2.3. Modèle probabiliste

Dans un modèle probabiliste, on associe aux variables du système des probabilités en fonction de la valeur que peuvent prendre les variables. On appelle cette fonction, distribution de probabilités. Pour une variable X, on note $P_X(x)$ ou $P(X = x)$, la probabilité que la variable X prenne la valeur x.

Dans un modèle à n variables, X_1 à X_n , indépendantes les unes des autres (i.e. telles que l'observation de l'une d'entre elles n'apporte aucune information sur la probabilité des autres), la probabilité d'obtenir un état du système dans lequel $X_i = x_i$ est égale à :

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$

que l'on notera en abrégé : $P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i)$

Mais il est rare que toutes les variables d'un système soit indépendantes. On ne peut alors exprimer la probabilité globale sans connaître les dépendances des variables. La dépendance d'une variable X par rapport à une variable Y peut se traduire par des probabilités conditionnelles $P(X = x / Y = y)$. Si le système comporte n variables, la dépendance devient pour X_i : $P(X_i = x_i / X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \dots, X_n = x_n)$. S'il existe un ordonnancement des variables tel que toute variable ne dépende que d'un sous ensemble de ses prédécesseurs, alors on peut réduire la dépendance des

variables du système. Les éléments de cet ensemble sont appelés *parents markoviens*. La théorie des chaînes de Markov (Arnold & Guessarian 1994) permet de déterminer la probabilité d'un événement lié à une variable en ne connaissant que ses parents markoviens. On note PA_i l'ensemble des parents markoviens de la variable X_i . On a la distribution suivante :

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i / pa_i)$$

Cette théorie est à la base des réseaux bayésiens (Koller & Pfeffer 1997, Pearl 1995) qui représentent graphiquement les dépendances markoviennes. On utilise le calcul des probabilités pour bâtir des modèles statistiques permettant le traitement et l'interprétation de données, mais aussi des modèles stochastiques permettant de représenter des systèmes dynamiques soumis à des bruits et des perturbations inconnues. La théorie des chaînes de Markov permet de prédire la probabilité de réalisation d'un état du système avec un minimum de connaissances sur les probabilités des dépendances internes. L'avantage de cette théorie est qu'elle peut être appliquée à tout type de variables (numérique ou symbolique). Cependant, son principal inconvénient est la détermination des distributions des probabilités qui ne peuvent pas être déduites (i.e. les $P(x_i/pa_i)$ dans les réseaux bayésiens). Ces distributions doivent associer un nombre précis entre 0 et 1 pour toutes les valeurs que peuvent prendre les variables associées. Par exemple, si A et B sont les parents markoviens de C et que ces trois variables sont numériques, il faudra avoir des observations suffisamment importantes en nombre, pour pouvoir estimer les différentes valeurs de $P(C/A,B)$, c'est-à-dire en général, déterminer le type de la distribution et établir la fonction qui exprime les paramètres de $P(C/A,B)$ en fonction des valeurs de A et B.

3. LOGIQUE FLOUE

La *logique floue* permet de formaliser des raisonnements fondés sur des connaissances incertaines et imprécises (Alliot & Schiex 1994, Kosko & Isaka 1993). En logique classique une connaissance (e.g. « Albert Einstein est né le 14 mars 1879 ») a une valeur booléenne, vrai ou fausse. Une connaissance incertaine est une connaissance dont la valeur de vérité n'est ni vraie, ni fausse. L'incertitude traduit un manque de connaissance (e.g. « Albert Einstein a dû fumer sa première pipe le 11 novembre 1907 » est une connaissance incertaine). Une connaissance imprécise est aussi une connaissance incomplète, mais le manque de connaissance ne se situe pas dans le degré de vérité à accorder à la connaissance mais dans la formulation même de la connaissance (e.g. « Albert Einstein

est né vers la fin du siècle dernier » est une connaissance certaine (cf. exemple précédent), mais elle est imprécise). Les notions d'imprécis et d'incertain sont complémentaires dans le sens où une connaissance imprécise entraîne souvent un raisonnement incertain (Kosko & Isaka 1993, Dubois & Prade 1987).

3.1. Logiques multivaluées

Avant de présenter la logique floue, nous allons nous intéresser aux logiques multivaluées (ou logiques multivalentes suivant les auteurs). Il existe en effet une grande variété de logiques permettant de raisonner avec des connaissances incertaines. La logique floue n'est que l'une d'entre elles. Hormis cette dernière, que nous présenterons plus loin, les logiques multivaluées ne diffèrent que légèrement de l'une à l'autre, le principe restant invariable. Nous ne présenterons que les logiques de Jan Lukasiewicz (Iturrioz 1990, Lukasiewicz 1970).

3.1.1. Fondement des logiques multivaluées

Le but des logiques multivaluées est de raisonner avec des connaissances qui ne sont ni vraies ni fausses, des connaissances incertaines. Les premières logiques multivaluées ont été développées, au début du vingtième siècle, autour, notamment, du débat philosophique portant sur les paradoxes logiques. Les paradoxes sont fondés sur une réfutation du *principe de non-contradiction* d'Aristote qui n'est autre que le principe de tiers exclu. Ce principe stipule que : « Il est impossible que le même attribut appartienne et n'appartienne pas en même temps au même sujet et sous le même rapport ».

On peut par exemple citer le Paradoxe énoncé en 1901 par le philosophe Bertrand Russell. Ce paradoxe est fondé sur la notion d'ensemble (Alliot & Schiex 1994) :

Soit S l'ensemble contenant tous les ensembles qui ne sont pas éléments d'eux-mêmes. Si S est un élément de lui-même, alors, par définition, il ne peut se contenir et donc ne peut être un élément de lui-même. De même, si S n'est pas un élément de lui-même alors, par définition, il doit se contenir et doit donc être un élément de lui-même.

En d'autres termes, la proposition « $S \notin S$ » n'est ni vraie ni fausse. Pour pallier ce défaut de la logique classique, différents systèmes de logiques multivaluées ont été énoncés.

La logique classique repose sur une représentation booléenne de la vérité. Vrai et Faux sont les deux seules valeurs possibles. Une logique multivaluée possède plusieurs valeurs de vérité (au moins

trois). Les différentes logiques multivaluées peuvent être différenciées en fonction du nombre de valeurs de vérité qu'elles admettent et sur le sens accordé à ces valeurs. Nous ne verrons en détail que deux logiques multivaluées, parmi les logiques de Lukasiewicz, que nous noterons L_3 et L_{N1} (Iturrioz 1990).

3.1.2. Logique L_3

L_3 est une logique reposant sur 3 valeurs de vérité, Vrai, Faux et une troisième valeur, intermédiaire, que l'on peut nommer Indéterminé, Possible ou bien Incertain. Lukasiewicz représente ces trois valeurs par des réels 1 pour Vrai, 0 pour Faux et $\frac{1}{2}$ pour Indéterminé. Il définit ensuite l'implication, \rightarrow , et la négation, \neg , comme suit :

$x \rightarrow y$		y	
	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
x $\frac{1}{2}$	1	1	$\frac{1}{2}$
0	1	1	1

x	$\neg x$
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

On peut remarquer que $\neg \frac{1}{2} = \frac{1}{2}$ et que, comme en logique classique, $x \rightarrow x = 1$ quel que soit $x \in \{1, \frac{1}{2}, 0\}$. On peut définir la conjonction, \wedge , ou la disjonction, \vee , à partir des deux définitions précédentes.

$$\forall x, y \in \{1, \frac{1}{2}, 0\}, x \vee y = (x \rightarrow y) \rightarrow y \text{ et } x \wedge y = \neg(\neg x \vee \neg y).$$

On remarque que contrairement à la logique classique, l'implication n'est pas *l'implication matérielle*, c'est à dire que $x \rightarrow y \neq \neg x \vee y$. Il existe d'autres logiques multivaluées qui conservent l'implication matérielle de la logique classique, c'est par exemple le cas de la logique de Kleene (Iturrioz 1990).

Cette logique a permis à Lukasiewicz d'exprimer le concept de possibilité. La possibilité d'une proposition p est une proposition, notée $\diamond p$, dont la valeur de vérité est Vrai s'il est possible que p soit vraie, et Faux dans le cas contraire. Si x est la valeur de vérité de p et $\diamond x$ celle de $\diamond p$ alors $\diamond x = \neg x \rightarrow x$. On définit la nécessité de p , noté $\circ p$, comme dans la logique modale de Lewis (Alliot & Schiex 1994), c'est-à-dire que $\circ x = \neg \diamond \neg x$. On peut remarquer que la valeur de possibilité et de nécessité ont des valeurs de vérité booléenne (1 ou 0), que l'on peut résumer dans le tableau suivant :

x	$\diamond x$	$\circ x$
1	1	1
$\frac{1}{2}$	1	0
0	0	0

L'introduction de l'incertitude dans le domaine des valeurs de vérité entraîne la perte d'un certain nombre d'axiomes de la logique classique. C'est, par exemple, le cas du principe du tiers exclu qui stipule que $p \vee \neg p$ est toujours vrai. Ce principe qui est valide dans la logique classique ne l'est pas dans L_3 où l'on a $\frac{1}{2} \vee \neg \frac{1}{2} = \frac{1}{2}$. Ceci permet de statuer sur des paradoxes (e.g. la proposition « $S \subset S$ » a pour valeur de vérité $\frac{1}{2}$).

3.1.3. Logique L_{N1}

La logique L_{N1} est une généralisation extrême de la logique L_3 . Elle possède un ensemble infini de valeurs de vérité, correspondant à l'ensemble des nombre réels de l'intervalle $[0, 1]$. On y définit l'implication et la négation comme (Alliot & Schiex 1994, Iturrioz 1990) :

$$(9) \quad x \rightarrow y = \min(1, 1 - x + y)$$

$$(10) \quad \neg x = 1 - x.$$

où $x, y \in [0, 1]$.

Cette logique est très proche, dans l'idée, de L_3 . La seule différence est la graduation de l'incertain engendrée par la multiplication des valeurs intermédiaires entre Vrai (1) et Faux (0). Cette graduation permet de représenter des quantificateurs tels que « généralement » ou « certains ». Cette représentation reste arbitraire, et spécifique au domaine représenté. Il est, en effet, difficile de déterminer *a priori* la valeur de vérité d'une proposition débutant par le quantificateur « généralement ». D'autres formalismes sont plus adaptés à cette représentation, comme par exemple la logique des défauts de Reiter (Alliot & Schiex 1994).

La graduation des valeurs de vérité n'est pas sans rappeler la mesure d'événements probabilistes. La mesure de probabilité d'un événement A, $P(A)$, peut être perçue comme une interprétation de la valeur de vérité de la proposition affirmant le même événement. Cependant, le non respect du principe du tiers exclu arrête là la comparaison entre les deux formalismes. En effet, selon la théorie des probabilités, la disjonction d'un événement et de son contraire ($p \vee \neg p$) a une probabilité égale à 1, alors que ce n'est pas le cas dans L_3 et L_{N1} .

3.2. Théorie des ensembles flous

La théorie des ensembles flous s'inspire de la conception de l'incertain telle qu'elle a été exprimée dans les logiques multivaluées (i.e. la logique floue) pour représenter la notion de « connaissance imprécise ». L'imprécision est introduite au travers de propositions vagues. Ces propositions sont des propositions graduelles, c'est à dire que leur valeur de vérité peut varier graduellement suivant la situation à laquelle elles sont appliquées (Dubois & Prade 1987). Considérons, par exemple, la proposition « Jacques est une personne d'âgée », suivant l'âge réel de Jacques, cette proposition sera plus ou moins vérifiée. Si Jacques a 20 ans elle sera fausse, s'il a 70 ans elle sera vraie, mais, s'il a 40 ans, il sera plus difficile de statuer sur sa valeur de vérité !

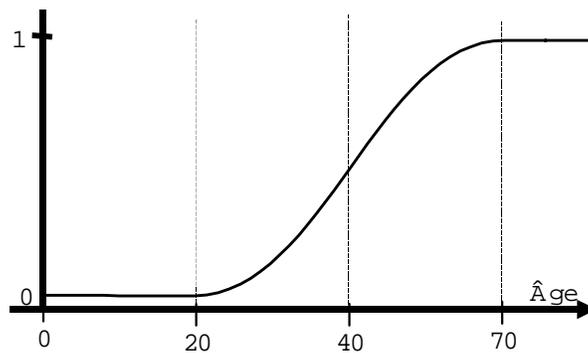


Figure 12: Fonction d'appartenance de l'ensemble flou associé au prédicat « est une personne âgée ».

Les propositions vagues peuvent être représentées par des ensembles flous. Un tel ensemble est défini par une fonction d'appartenance variant de manière continue entre 0 et 1 (0 indiquant que l'élément considéré n'appartient pas à l'ensemble et 1 indiquant qu'il y appartient). Pour revenir à notre exemple (Figure 12), l'ensemble flou associé au prédicat « est une personne âgée » posséderait une fonction d'appartenance qui aurait une valeur nulle pour tous les âges compris entre 0 et 20 ans et elle vaudrait 1 pour tous les âges supérieurs à 70, entre les deux sa valeur augmente progressivement. De telles fonctions d'appartenance sont dites nuancées (Dubois & Prade 1987).

On définit un ensemble flou F par la fonction d'appartenance μ_F . On peut définir des opérations sur les ensembles flous à partir de généralisations des opérations sur les ensembles classiques (i.e. union, intersection, inclusion, égalité). On peut de plus représenter un ensemble flou par une famille d'ensembles classiques associés à différents niveaux du seuil d'appartenance. On appelle ces ensembles des *coupes de niveau*. On définit une coupe de niveau α d'un ensemble flou F , par l'ensemble classique F_α tel que :

$$(11) \quad F_\alpha = \{x \mid \mu_F(x) \geq \alpha\}$$

On appelle *noyau* de l'ensemble flou la coupe de niveau 1 et *support* la coupe de niveau α telle que α tend vers 0 mais reste supérieur à 0. (e.g dans notre exemple, le noyau de l'ensemble flou est l'intervalle $[70, +\infty[$, le support l'intervalle $[20, +\infty[$).

Dans de nombreux systèmes où les valeurs sont des valeurs réelles, on représente les ensembles flous par une fonction d'appartenance trapézoïdale. L'ensemble peut alors être défini par 4 réels $[a, b, \alpha, \beta]$ tels que $a, b, \alpha, \beta \in \mathbb{R}$ et $\alpha \geq 0, \beta \geq 0$. L'intervalle $[a, b]$ représente le noyau de l'ensemble flou et l'intervalle $[a - \alpha, b + \beta]$ représente son support (Figure 13).

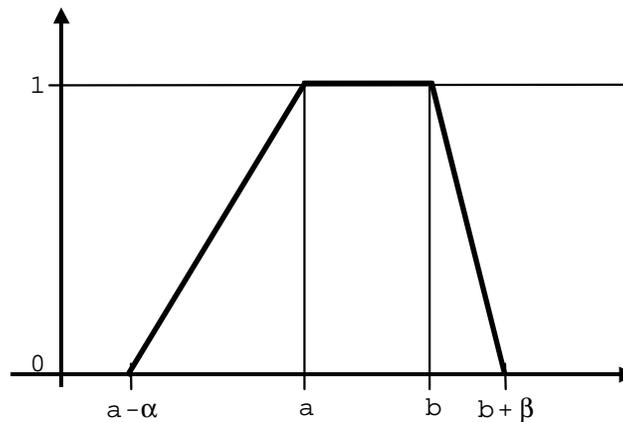


Figure 13: Fonction d'appartenance trapézoïdale.

On associe la conjonction des propositions vagues à l'intersection des ensembles flous et la disjonction à l'union. Les opérations intersection et union sont souvent définies par :

$$(12) \quad \mu_{F \cap G}(u) = \min(\mu_F(u), \mu_G(u))$$

$$(13) \quad \mu_{F \cup G}(u) = \max(\mu_F(u), \mu_G(u))$$

Il y a d'autres manières de définir l'intersection et l'union d'ensembles flous. C'est d'ailleurs l'un des intérêts de la logique floue qui permet ainsi de définir des connecteurs logiques particuliers en plus de ceux définis ci-dessus (Dubois & Prade 1987). Cette observation est aussi valable pour l'implication, que l'on peut définir en s'inspirant de l'implication matérielle ($a \rightarrow b = \neg a \vee b$) ou pas. Tous ces connecteurs peuvent cohabiter dans un même système mais, comme dans les

logiques multivaluées, aucun de ces connecteurs ne permet à lui seul de retrouver toutes les propriétés de la logique classique.

La logique floue est largement utilisée pour la modélisation et le contrôle des systèmes automatiques, les systèmes étant gérés par des déductions au travers de règles floues. La règle d'inférence est appelée *modus ponens généralisé*. D'une manière générale, les règles floues représentent des relations floues, définies par des fonctions d'appartenance, par exemple, notées $\mu_R(u,v)$ pour une relation binaire R.

La logique floue permet de représenter des connaissances de complexité élevée de manière synthétique (Kosko & Isaka 1993). Elle permet aussi de traiter les paradoxes de la logique classique. Pour reprendre le paradoxe de Russell, l'ensemble S est un ensemble flou dont la fonction d'appartenance μ_S est telle que $\mu_S(S) = 1/2$.

Pour pallier le problème de la spécification des fonctions d'appartenance des ensemble flous, on peut associer une représentation floue à un autre formalisme, comme par exemple des réseaux de neurones (Pedrycz *et al.* 1995, Kosko & Isaka 1993). Ceci permet un apprentissage sur les fonctions d'appartenance. Cette association permet ainsi de limiter une contraignante expertise, quoique l'apprentissage soit souvent supervisé. Les unités d'un réseau de neurones flous possèdent des valeurs d'activation floues (Pedrycz *et al.* 1995). Les entrées du neurone sont combinées par des opérations sur des ensembles flous et peuvent alors représenter des conjonctions, disjonctions ou tout autres connecteurs logiques (Pal & Konar 1996). Cette organisation permet de représenter les règles floues de manière distribuée.

3.3. Logique possibiliste

La logique possibiliste est un formalisme proche de la logique floue (Dubois *et al.* 1994). Elle considère le degré d'appartenance, $\mu_F(u)$, d'un élément u à un ensemble flou F, comme un degré d'incertitude, ou de possibilité suivant les cas (Andres *et al.* 1990, Dubois & Prade 1987). La fonction d'appartenance μ_F est donc considérée comme une distribution de possibilités, notée ici π_F . Cette distribution permet de définir la mesure de possibilité, Π , et de nécessité, N , d'un événement. Soit F un ensemble possibiliste, défini par la distribution π_F , alors la mesure de possibilité et de nécessité, selon F , de l'ensemble de valeurs X sont égales à :

$$(14) \quad \Pi(X) = \sup_{x \in X} \pi_F(x)$$

$$(15) \quad N(X) = \inf_{x \in X} (1 - \pi_F(x))$$

Si $\Pi(X) = 0$ alors X est faux. On peut dire aussi que X est impossible. Si $N(X) = 1$ alors X est vrai, ou autrement dit, X est nécessaire. Si $\Pi(X) = 1$ et $N(X) = 0$ alors X est totalement possible mais absolument pas nécessaire, ce qui traduit l'ignorance totale sur l'événement X .

On observe que la dualité de la mesure de possibilité et de nécessité est vérifiée, soit $\Pi(X) = 1 - N(\bar{X})$ où \bar{X} est le complémentaire de X . Ces deux mesures sont complémentaires. Cette représentation granulaire et duale fait toute la richesse de la logique possibiliste. Comme dans la logique floue, on peut définir des connecteurs logiques au moyen d'opérations sur les ensembles. Quel que soit P et Q on a :

$$(16a) \quad \Pi(P \vee Q) = \max(\Pi(P), \Pi(Q))$$

$$(16b) \quad N(P \wedge Q) = \min(N(P), N(Q))$$

Le fait de considérer la fonction d'appartenance comme une distribution de possibilités rapproche la logique possibiliste de la théorie des probabilités. Cependant, la logique floue est moins contrainte que la théorie des probabilités. On peut noter une différence fondamentale dans ce que ces deux théories représentent. La logique floue représente l'ignorance sur une connaissance, alors que la théorie des probabilités exprime la certitude d'accomplissement de la connaissance. La différence est subtile mais elle peut être illustrée par l'impossibilité de représenter l'ignorance absolue sur une information avec une seule distribution de probabilité (Dubois & Prade 1987). La probabilité de l'information en question sera en effet dépendante des différents choix possibles et variera selon leur nombre (e.g. si il y a deux événements possibles chacun a une probabilité de $\frac{1}{2}$, si il y en a trois elle passe à $\frac{1}{3}$) alors que dans la logique possibiliste l'ignorance sera exprimé par une possibilité totale et une nécessité nulle, comme on l'a vu ci-dessus. De manière consensuelle, on peut considérer les mesures de possibilité et de nécessité comme une enveloppe de la probabilité de réalisation de l'événement considéré (Dubois & Prade 1987).

4. RAISONNEMENT QUALITATIF

Le raisonnement qualitatif trouve, en majeure partie, ses origines dans les travaux de Patrick J. Hayes sur la physique naïve (Hayes 1985). L'objectif de ses travaux était de construire une modélisation de notre perception *de sens commun* du monde physique, et ainsi passer outre la complexité des modèles mathématiques. Hayes proposait une axiomatisation en logique du premier ordre du comportement intuitif des liquides, dont la modélisation soulève de nombreux problèmes dans une approche classique (Hayes 1985).

Depuis ces travaux, la physique qualitative, qui est devenue le raisonnement qualitatif, a beaucoup évolué et divers formalismes originaux ont été proposés. D'une manière très générale, le raisonnement qualitatif a pour vocation de combler l'insuffisance des méthodes numériques dans des domaines où les connaissances sont peu formalisées et/ou difficilement quantifiables (Travé-Massuyès *et al.* 1997). Ce type de raisonnement est fondé sur une description comportementale de l'état des *variables* du modèle en considérant les *qualités* qui les caractérisent (de Kleer & Brown 1983).

Prosaïquement, les « qualités » sont des symboles plus ou moins ordonnés. On peut, par exemple, considérer que les qualités associées à une variable représentant la « température de l'eau » peuvent correspondre aux trois états : « glacé », « liquide » et « gazeux » de l'eau. Dans ce cas, on peut ordonner les qualités : glacé < liquide < gazeux. On peut ainsi raisonner sans connaître la véritable valeur de cette température. On appelle l'ensemble des qualités représentant la valeur d'une variable son *espace des qualités*. La finesse du raisonnement est proportionnelle au nombre de qualités considérées. Les *algèbres qualitatifs* qui permettent de manipuler directement les qualités comme des entités algébriques.

Dans un raisonnement qualitatif, la notion d'imprécision est *implicite* dans la représentation des données (i.e. implicite dans la notion même de qualité). Les algèbres qualitatives, ainsi que les formalismes qui les utilisent, ne manipulent donc pas explicitement l'imprécision des valeurs, contrairement à la théorie des ensembles flous. On peut cependant coupler un formalisme qualitatif et cette théorie (Dubois & Prade 1989) en offrant ainsi deux niveaux d'abstraction complémentaires. Cette association permet, suivant l'orientation de la démarche, de guider la spécification des ensembles flous associés aux règles du système (Bellazzi *et al.* 1999, Yao 1998, Foulloy 1993) ou, à l'inverse, d'affiner la spécification des relations qualitatives (Bousson & Travé-Massuyès 1993, Dubois & Prade 1989).

4.1. L'algèbre des signes

L'algèbre des signes est un formalisme datant du début des années 1980. L'espace des quantités y est réduit à un ensemble de 4 valeurs symboliques, ce qui simplifie énormément les calculs dans les modèles (Travé-Massuyès *et al.* 1997). Le raisonnement sur les signes est souvent suffisant pour déterminer l'évolution qualitative d'un système. On raisonne ainsi sur les variations croissantes, décroissantes ou nulles, des variables d'état du système.

L'espace des qualités, que nous notons S , est égal à l'ensemble $\{-, 0, +, ?\}$ où $-$ représente la qualité « négatif », 0 , la qualité « nul », $+$ la qualité « positif » et $?$ la qualité « indéterminé ». Cette dernière qualité a été ajoutée pour rendre les opérations complètes sur l'espace S . On définit deux opérations sur S , la somme \oplus et la multiplication \otimes , par les tables suivantes :

\oplus	$-$	0	$+$	$?$
$-$	$-$	$-$	$?$	$?$
0	$-$	0	$+$	$?$
$+$	$?$	$+$	$+$	$?$
$?$	$?$	$?$	$?$	$?$

\otimes	$-$	0	$+$	$?$
$-$	$+$	0	$-$	$?$
0	0	0	0	0
$+$	$-$	0	$+$	$?$
$?$	$?$	0	$?$	$?$

Puis on peut, par exemple, définir la soustraction, représentée par $a \ominus b$, comme équivalente à l'expression $a \oplus (- \otimes b)$. On définit aussi une égalité qualitative sur S , noté \approx . Elle peut être définie par la table de vérité suivante :

\approx	$-$	0	$+$	$?$
$-$	Vrai	Faux	Faux	Vrai
0	Faux	Vrai	Faux	Vrai
$+$	Faux	Faux	Vrai	Vrai
$?$	Vrai	Vrai	Vrai	Vrai

Bien qu'appelée « égalité », cette relation ne définit pas une classe d'équivalence car elle n'est pas transitive ($- \approx ?$, $? \approx +$ mais on n'a pas $- \approx +$). On définit naturellement une relation d'ordre strict partiel sur S , notée $>$, telle que $+ > 0$, $0 > -$ et $+ > -$.

Les propriétés de cette algèbre sont assez proches de l'algèbre linéaire classique. Cette proximité, alliée à la simplification de la spécification des valeurs du système, permet de raisonner dans des modèles physiques. Pour ce faire, il suffit de remplacer les valeurs réelles par des valeurs qualitatives. En exprimant les dérivées d'une variable, on peut donner une indication sur sa

variation et de cette façon préciser l'état du système. La facilité de mise en œuvre de l'algèbre des signes permet son utilisation pour des analyses de tendance ou des vérifications rapides de cohérence d'un modèle. Cependant, sa faiblesse de description entraîne beaucoup d'indétermination et n'en fait donc pas, à lui seul, un formalisme adéquat pour la description de systèmes complexes (Travé-Massuyès *et al.* 1997).

4.2. Les ordres de grandeur

On peut être plus précis, dans la description des qualités d'une variable, que ne le permet l'algèbre des signes, suivant la connaissance que l'on possède du système à modéliser. Ainsi, dans l'exemple de description qualitative de la température de l'eau, on peut remplacer l'espace de qualités {glacé, liquide, gazeux} par des qualités un peu plus *précises*, comme {glacé, froid, frais, tiède, chaud, bouillant, gazeux}. Un bon compromis doit être atteint entre une représentation trop synthétique et une représentation trop précise, de manière à pouvoir représenter tous les états « utiles » du système et ainsi obtenir un comportement correct du modèle qualitatif. Ce compromis est souvent dépendant de l'application considérée.

Le raisonnement sur les ordres de grandeur regroupe deux styles de raisonnement. Le raisonnement sur les ordres de grandeur absolus s'appuie sur une partition de l'ensemble des réels plus fine que celle de l'algèbre des signes. Le raisonnement sur les ordres de grandeur relatifs définit des relations d'ordre entre les qualités (Travé-Massuyès *et al.* 1997).

4.2.1. Ordres de grandeur absolus

Ce formalisme est bâti sur un espace des qualités représentant une partition de l'ensemble des réels (\mathbb{R}). Cette partition est naturellement ordonnée par la relation supérieur, $>$, définie sur \mathbb{R} . L'espace des qualités associé à cette partition est alors considéré comme une échelle de qualités ; nous le notons S_1 .

A partir des symboles de S_1 , on peut bâtir de nouveaux symboles. On les note $[x, y]$, où $x \in S_1$, $y \in S_1$ et l'élément de la partition associé à x est strictement inférieur à celui associé à y . Ces nouveaux symboles représentent la réunion de x , y et de tous les éléments les séparant (i.e. au sens de la relation supérieur). On appelle univers de description complet, l'ensemble S^* , composé des symboles de S_1 et des nouveaux symboles.

On peut, par exemple, considérer la partition de \mathbb{R} inspirée de l'espace de l'algèbre des signes mais où l'on distingue les réels proches et éloignés de 0. On représente cette partition par l'espace des qualités $S_1 = \{NG, NP, 0, PP, PG\}$, où NG représente les réels négatifs grands en valeur absolue, NP les réels négatifs petits en valeur absolue, etc. On obtient donc l'échelle : $NG < NP < 0 < PP < PG$.

On peut alors bâtir de nouveaux symboles comme, par exemple, $[NG PP]$ représentant la réunion de NG, NP, 0 et PP. On remarque que $[NG NP]$ représente les réels négatifs et que $[PP PG]$ les réels positifs. De même $[NG PG]$ représente l'ensemble de tous les réels et peut donc être comparé à la qualité ? de l'algèbre des signes. On peut remarquer que S^* , bâti à partir de cet ensemble, contient l'espace des qualités S de l'algèbre des signes (i.e. l'ensemble $\{[NG NP], 0, [PP PG], [NG PG]\}$).

On définit, enfin, une relation d'ordre sur S^* , notée \angle , telle que $x \angle y$ si et seulement si les éléments de la partition associés à x sont inclus dans ceux associés à y. Cette relation donne à S^* une structure de demi-treillis.

Dans l'exemple que nous venons de citer on a, par exemple, $[NP 0] \angle [NP PP] \angle [NG PG]$, ce dernier symbole étant le sommet de notre treillis. On fournit à S^* une relation appelée égalité qualitative, notée \approx , construite comme une généralisation de l'égalité de l'algèbre des signes et définie comme :

Soit $a, b \in S^*$, $a \approx b$ si et seulement si $\exists z \in S^*$ tel que $z \angle a$ et $z \angle b$.

Muni de ces relations, on peut construire des abstractions des fonctions à valeurs dans \mathbb{R} . Ces abstractions appelées *fonctions qualitatives associées*, doivent conserver la même sémantique que les fonctions qu'elles représentent, c'est-à-dire que l'image d'une valeur qualitative Q est le plus petit élément de S^* contenant l'image des réels sémantiquement associés à Q. Les opérations d'addition et de multiplication qualitatives sont définies comme les fonctions qualitatives associées aux opérations + et \times définies sur \mathbb{R} . On peut les construire grâce à des tables semblables à celles de l'algèbre des signes (Travé-Massuyès *et al.* 1997, Parsons 1991).

Les calculs fondés sur cette algèbre possèdent sensiblement les mêmes avantages que ceux fondés sur l'algèbre des signes. Les propriétés des opérateurs sont toutefois plus faibles (e.g. on ne conserve pas l'associativité et la distributivité stricte des opérateurs somme et produit, Travé-

Massuyès *et al.* 1997). Cependant, la partition n'est pas choisie *a priori* et, de ce fait, la finesse des résultats peut être adaptée en fonction du système modélisé.

4.2.2. Ordres de grandeur relatifs

Dans un système aux ordres de grandeurs relatifs, on ne considère plus une partition absolue de l'ensemble des réels mais on raisonne par comparaisons relatives entre les qualités elles-mêmes. Les comparaisons se font via des relations binaires. Ce sont ces relations qui vont formaliser le raisonnement sur les qualités. S'il existe plusieurs formalismes, proposant différentes relations, nous ne discuterons, dans le cadre de notre présentation, que du formalisme $O(M)$ de Mavrovouniotis et Stephanopoulos (Travé-Massuyès *et al.* 1997, Dubois & Prade 1989, Mavrovouniotis & Stephanopoulos 1988, 1987).

$O(M)$ propose 7 relations sur les ordres de grandeurs, notées : \gg , $>-$, $>\sim$, $==$, $\sim<$, $-<$ et \ll . Elles représentent les assertions qualitatives suivantes (Mavrovouniotis & Stephanopoulos 1987) :

$A \gg B$ est équivalent à dire que A est grandement supérieur à B.

$A >- B$ est équivalent à dire que A est modérément supérieur à B.

$A >\sim B$ est équivalent à dire que A est légèrement supérieur à B.

$A == B$ est équivalent à dire que A est identique à B.

$A \sim< B$, $A -< B$ et $A \ll B$ sont les relations symétriques à $A >\sim B$, $A >- B$ et $A \gg B$.

On peut combiner (par disjonction) ces 7 relations pour obtenir un total de 21 relations non équivalentes. On peut par exemple considérer la disjonction de $>\sim$, $>-$ et \gg , que l'on note $>\sim..>>$ et qui n'est autre que la relation supérieur (Mavrovouniotis & Stephanopoulos 1987).

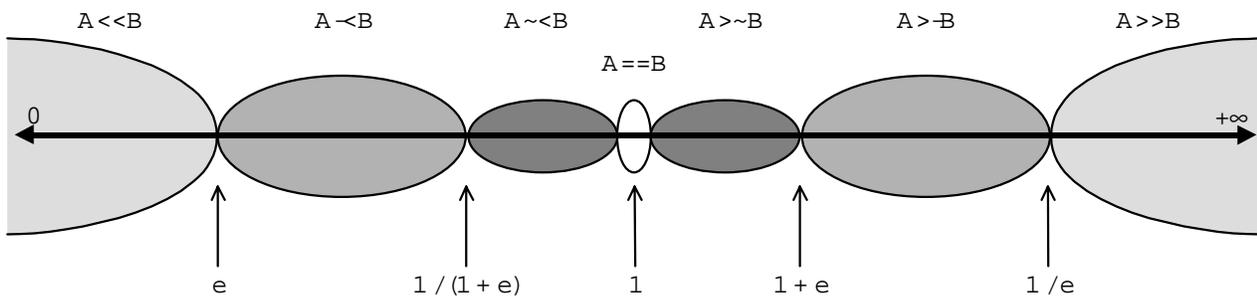


Figure 14 : Interprétation contrainte de $A r B$ en fonction du rapport A/B dans $O(M)$, d'après Mavrovouniotis et Stephanopoulos (1988).

Le raisonnement se fait, soit par déduction de nouvelles relations à partir d'un ensemble de relations exprimées sur les variables du système, soit directement par le calcul du rapport entre les valeurs de 2 variables. En effet, toute relation de $O(M)$ du style $A r B$ est équivalente à $A/B r 1$, avec $r \in \{>>, >-, >\sim, ==, \sim<, -<, <<\}$. Dès lors, et d'après les propriétés des relations et de différents postulats (e.g. $(A - B) << B \rightarrow A >\sim B$), on montre que l'on peut déterminer la relation entre A et B si l'on connaît le rapport A/B comme le montre la figure ci-dessus (Figure 14).

On peut directement utiliser cette partition pour interpréter la relation qui lie deux variables. Si, par exemple, $A = 10$, $B = 2$ et que l'on a fixé $e = 0.1$, on obtient $A >- B$. Mais on peut aussi utiliser cette partition pour inférer de nouvelles qualités. Admettons que l'on sache $C \sim< A$ et $A \sim< B$ sans connaître la valeur de A, B ou C. A partir de ces deux relations on peut dire :

$$(1+e)^{-1} < C/A < 1 \quad \text{et} \quad (1+e)^{-1} < A/B < 1.$$

D'où l'on peut déduire en multipliant ces deux équations :

$$(1+e)^{-2} < C/B < 1$$

Et ayant fixé $e = 0.1$, on peut inférer que $C -< B$ ou $C \sim< B$, ce que l'on note $C -<..\sim< B$. Il est cependant possible et souhaitable d'affiner cette inférence pour limiter l'accroissement de l'imprécision (Travé-Massuyès *et al.* 1997, Parsons 1991, Mavrovouniotis & Stephanopoulos 1988).

Comme on le voit, la détermination de la relation associée à un rapport A/C est dépendante de la valeur du paramètre e . Fixer ce paramètre équivaut au choix d'une partition avec des ordres de grandeurs absolus. On constate qu'ici, grâce aux postulats et propriétés des relations de $O(M)$, il n'y a qu'un seul paramètre à choisir. Il représente en fait un réel positif considéré comme *très inférieur* à 1, soit $e << 1$. Ce système est plus simple à spécifier qu'un système aux ordres de grandeur absolus, puisqu'il n'y a qu'une seule valeur à fixer. Cependant les résultats restent toujours dépendants du choix de e et on ne peut définir les qualités des valeurs de manière indépendante.

4.3. Les intervalles

Le raisonnement sur les intervalles peut être considéré comme une extension du raisonnement sur les ordres de grandeur. On peut en effet considérer l'espace des intervalles réels comme l'univers de description complet associé à une partition infinie dont chaque élément de base serait un réel (Travé-Massuyès *et al.* 1997). Il en découle deux des principales caractéristiques de ce formalisme.

Premièrement, il permet de faire cohabiter des quantités (les réels) et des qualités (les intervalles). De plus, l'espace des qualités contenant un ensemble infini de qualités, aucune spécification *a priori* sur les valeurs qualitatives des variables n'est nécessaire.

En d'autres mots, nous sommes dispensés de définir une partition de \mathbb{R} . Définir l'espace des quantités d'une variable peut se réduire à donner les bornes entre lesquelles la variable varie. On parle dans ce cas de *domaine* et non d'espace des qualités (Travé-Massuyès *et al.* 1997). Un autre effet de bord, dû au nombre infini de qualités, est la possibilité de *glisser* d'une qualité vers une autre, c'est-à-dire que les variations des valeurs ne se font pas par bond d'un élément de la partition à un autre. La variation s'exprime par un changement progressif de l'intervalle réel, ce qui représente un réel avantage pour les applications dynamiques. En effet, la notion d'*événement qualitatif* (i.e. lorsqu'une valeur d'une variable change de qualité) devient dynamique et on peut alors représenter tous les états intermédiaires de ce changement de valeur. L'évolution de cette valeur peut être représentée de manière continue.

Formellement, on définit un intervalle $[\alpha, \beta]$ comme l'ensemble des réels x tels que $\alpha \leq x \leq \beta$, α est appelé borne inférieure et β borne supérieure de l'intervalle (Travé-Massuyès *et al.* 1997, Sunaga 1958). Cette définition ne considère que les intervalles réels fermés. On ne peut donc pas construire de partition de l'ensemble des réels avec ces intervalles et, de ce fait, l'algèbre des intervalles reste différent d'un raisonnement sur les ordres de grandeurs. Cependant, comme dans le raisonnement sur les ordres de grandeur absolus, on peut définir une relation d'égalité qualitative, noté \approx , pour les intervalles. On la bâtit comme une extension de la précédente, sur la base de l'inclusion des intervalles :

Soit A et B deux intervalles, on dit que $A \approx B$ si et seulement si il existe un réel x tel que $x \in A$ et $x \in B$.

En d'autres termes, deux intervalles sont qualitativement égaux si et seulement si ils possèdent une intersection non vide. Cette relation n'est pas transitive, et l'on remarque que deux intervalles contigus sur l'ensemble des réels ne sont pas disjoints et donc ils sont qualitativement égaux (e.g. $[1, 2] \approx [2, 3]$).

On peut aussi définir les opérations d'addition et multiplication sur les intervalles comme suit :

Si $\diamond \in \{+, \times\}$ et A et B sont deux intervalles alors $A \diamond B$ est l'intervalle $\{x \diamond y / x \in A \text{ et } y \in B\}$.

L'opposé d'un intervalle A est l'intervalle noté $-A = \{x / -x \in A\}$, et l'inverse d'un intervalle A et noté A^{-1} , est défini, si et seulement si $0 \notin A$, par $A^{-1} = \{x / x^{-1} \in A\}$.

Ces opérations peuvent être réalisées directement sur les bornes des intervalles. On a, par exemple, $[a, b] + [c, d] = [a+c, b+d]$. Les calculs numériques effectués sur les bornes des intervalles facilitent le traitement de l'information. Cependant, cet aspect numérique place ce formalisme à la limite du raisonnement qualitatif. C'est pourquoi on le désigne souvent sous le terme de raisonnement *semi-qualitatif*.

4.4. La simulation qualitative

L'abondance des critères de modélisation a donné naissance à une grande variété de systèmes de simulation qualitative. Tous ces formalismes ont été appliqués à divers domaines, notamment à la supervision et au diagnostic des systèmes physiques complexes (Iwasaki 1997, Travé-Massuyès *et al.* 1997, Vescovi *et al.* 1997, MQ&D group 1996). Tous ces formalismes reposent sur les *algèbres qualitatives* que nous venons de présenter. De nombreux formalismes permettent de combiner des calculs qualitatifs à des calculs purement symboliques et/ou numériques, le but étant à chaque fois de pouvoir représenter l'ensemble le plus vaste de connaissances sur un système.

On peut caractériser les modèles qualitatifs par le fait qu'ils expriment *explicitement*, ou pas, la causalité entre les variables du système (Travé-Massuyès *et al.* 1997, 1993, Leyval *et al.* 1994). Lorsque la causalité n'est pas représentée, les modèles prennent alors la forme de systèmes d'équations qualitatives qui peuvent, par exemple, être considérés comme l'abstraction de systèmes d'équations algébriques. Dans le cas particulier des modèles dynamiques, les équations sont des équations différentielles. Les modèles causaux prennent souvent la forme de graphes d'influences qui expriment des relations orientées (i.e. de cause à effet) entre les variables.

Lorsque le modèle qualitatif ne peut représenter l'évolution du système qu'au travers de ses états d'équilibre, on dit qu'il est statique. Mais le modèle peut vouloir représenter explicitement l'évolution du système entre les états d'équilibre, tenant ainsi compte de la dynamique propre du système. Le modèle est alors dit dynamique. Le critère statique/dynamique permet de caractériser la simulation qualitative du modèle considéré. Les modèles statiques s'expriment sous la forme de relations fonctionnelles entre les valeurs des variables du système (i.e. dans le cas de modèles sans causalité explicite, les relations prennent la forme d'équations linéaires qualitatives ou d'équations algébriques).

Dans le cas des modèles dynamiques, la simulation consiste à déterminer les évolutions possible du système à partir d'un état initial, qui n'est pas forcément un état d'équilibre. On peut, globalement, distinguer deux principes de simulation. Le premier principe consiste à déterminer les états successeurs possibles du système en fonction des contraintes connues et de l'état à l'instant courant. On détermine ainsi un nouvel instant correspondant au moment où ce système change d'état. Si plusieurs états successeurs différents sont prévisibles, on considère séparément les différents états et on réitère le procédé pour chacun. De cette façon, on obtient un arbre d'états possibles dont la racine est l'état initial. Ce principe est, par exemple, celui du simulateur qualitatif QSIM (Travé-Massuyès *et al.* 1997, Kuipers 1993). Selon ce procédé, l'échelle chronologique entre les différents instants est variable. A l'inverse, d'autres simulateurs déterminent l'évolution de l'état du système en effectuant des calculs régulièrement échelonnés dans le temps. Ils se munissent alors d'une horloge discrète, dont l'unité détermine le *pas d'échantillonnage* du système. Cette constante permet d'obtenir une évolution précise du système, si elle est suffisamment petite face à sa dynamique propre. Ce second procédé ne permet pas, pour une raison de coût, de déterminer l'arbre de tous les états possibles du système à partir d'un état initial. Cependant, lorsque le pas d'échantillonnage est suffisamment petit, on peut de cette manière obtenir une simulation très précise en fonction du temps.

Le simulateur Ca~En (Bousson 1993) est un logiciel de simulation qualitative qui utilise une horloge discrète. Il est de plus caractérisé par une représentation mixte des connaissances. Ce simulateur fait cohabiter une représentation causale, appelée *niveau local*, et une représentation non causale, appelée *niveau global* (Bousson *et al.* 1994). Ce dernier niveau est utilisé pour exercer une contrainte sur le niveau local en filtrant les valeurs des variables de ce niveau. Les contraintes globales représentent des connaissances exprimées sous la forme d'équations numériques qualitatives (i.e. manipulant des intervalles), comme par exemple des lois physiques (e.g. la loi d'Ohm : $U = R \cdot I$). Destiné à la supervision de systèmes physiques complexes (Milne *et al.* 1996), ce simulateur propose, en plus du mécanisme précédent, des algorithmes de diagnostic de fautes par calcul des divergences entre valeurs mesurées et valeurs simulées (Pons *et al.* 1999, Travé-Massuyès *et al.* 1997). Les connaissances sur les systèmes dynamiques physiques se ramenant souvent à des équations différentielles, le simulateur Ca~En effectue des calculs fondés sur des fonctions de transfert exprimables par des équations différentielles linéaires du premier ordre. Ceci revient à raisonner sur des variations de valeurs et permet une discrétisation de la simulation.

5. L'APPROCHE CAUSALE

L'expression des relations de cause à effet, ou causalités, est un principe fondamental du raisonnement. L'approche causale est une méthode de modélisation qui est, de ce fait, largement utilisée dans les sciences en général (notamment dans les sciences sociales) et en intelligence artificielle en particulier. Il existe notamment des approches causales pour chacun des formalismes présentés dans ce chapitre. On peut, par exemple, utiliser des réseaux bayésiens comme support de modèles causaux probabilistes (Pearl 1999, 1995), les règles floues sont naturellement causales (Dubois & Prade 1987), et les modèles causaux sont très répandus dans tous les formalismes du raisonnement qualitatif (Travé-Massuyès *et al.* 1993, Kuipers 1984, de Kleer & Brown 1983).

La définition la plus prosaïque de la causalité est de la considérer comme une relation spatio-temporelle orientée qui unit deux éléments (e.g. des variables d'états d'un modèle), la cause et son effet. Les causalités sont des relations *localisées* à l'ensemble des éléments qu'elles concernent. Les modèles causaux présentent donc l'avantage d'explicitement les relations et les interactions qui lient tous les éléments modélisés, en les représentant individuellement. La représentation des relations entre causes et effets soulève le problème des causes premières (i.e. les causes ont des causes). Communément, un modèle causal définit un ensemble de causes exogènes, qui seront considérées comme externes au modèle et en seront les causes premières. Les causes endogènes sont celles qui sont représentées dans le modèle.

Cette approche focalise la représentation sur chaque élément (ou variable) par rapport aux relations qui le lient à ses causes. Cette représentation orientée élément-relation prédispose les modèles causaux à une représentation distribuée. On peut, en effet, identifier et localiser tout élément au travers des relations qui l'unissent aux autres éléments du système. Cette structure peut être représentée par un graphe orienté où les éléments sont les nœuds et les relations causales sont les arcs du graphe. Le graphe orienté associé à un modèle causal est appelé *graphe causal*. Une telle représentation graphique facilite la description des modèles.

Pour formaliser la notion de modèle causal, nous allons tout d'abord reprendre la définition proposée par Judea Pearl (1999). Cette définition représente l'effet d'une causalité comme la fonction de l'ensemble des causes dont il dépend. On peut donc mathématiquement définir l'architecture d'un modèle causal, très simplement, comme suit :

Un modèle causal est un triplet : $\langle U, V, F \rangle$, où :

- U est l'ensemble fini des variables exogènes du modèle. Ces variables sont celles dont les valeurs dépendent de facteurs externes au modèle.
- V est l'ensemble fini des variable endogènes du modèle. Les valeurs de ces variables sont déterminées par les autres variables du modèle ($U \cup V$).
- F est un ensemble fini de fonctions. Chaque fonction f_i , de F, est associée à une variable V_i , de V, et est une projection de $U \cup V - \{V_i\}$ sur $\{V_i\}$. La valeur de V_i , noté v_i , est obtenue par l'application de f_i à l'ensemble des valeurs w_1 à w_n des n variables, W_1 à W_n , appartenant à $U \cup V - \{V_i\}$:

$$(17) \quad v_i = f_i(w_1, \dots, w_n)$$

Le graphe causal est obtenu en associant un nœud à chaque variable du modèle et en liant par un arc les variables qui agissent effectivement sur les autres variables au travers des fonctions de F (i.e. les variables W_1 à W_n). Il n'y a qu'une fonction par variable, elle représente donc toute la causalité produisant la variable considérée. Cette représentation, très puissante, n'identifie pas explicitement les dépendances entre les variables. D'autres définitions de modèles causaux sont organisées autour de la notion d'*influence*. Une influence est une relation qui lie une cause à un effet. Elle exprime un lien causal direct. Dans ce cas, les arcs du graphe causal représentent les influences du modèle. Sur la base de cette définition, on peut définir différents types d'influences en les associant à diverses opérations de transformations, comme des fonctions de transfert (Travé-Massuyès *et al.* 1997).

De plus, la notion d'influence pose le problème des influences convergentes et de leur combinaison. Si une variable du modèle est l'effet de deux influences, il faut pour pouvoir obtenir la valeur de cette variable en combinant les deux effets pour n'en obtenir qu'un. La combinaison peut être obtenue en appliquant le principe de superposition des influences (i.e. on somme les influences). On suppose alors que les influences afférentes décrivent des mécanismes autonomes et qu'elles sont donc indépendantes les unes des autres. Cette notion est matérialisée par la représentation d'influences marginales (i.e. définie par la notion de variation marginale par Bousson (1993)). Cette notion est clairement définie dans les formalismes qualitatifs du simulateur Ca~En.

On peut alors exprimer une nouvelle définition d'un modèle causal sous la forme d'un quadruplet $\langle U, V, M, C \rangle$ où :

- U et V restent les ensembles des variables exogènes et endogènes.

- M est l'ensemble des fonctions associées aux influences. Pour toute variable V_j influençant une variable V_i on définit une fonction M_{ij} . La variation marginale de V_i , noté δv_{ij} et telle que $\delta v_{ij} = M_{ij}(v_j)$, est définie comme la variation de V_i si V_j était la seule variable l'influençant.
- C est l'ensemble de fonctions ξ , appelées fonctions de combinaison. La variation réelle de V_i est obtenue en appliquant cette fonction aux variations marginales, $\delta v_i = \xi(\delta v_{i1}, \dots, \delta v_{in})$.

Pour être valide une fonction de combinaison doit cependant respecter deux propriétés (Bousson 1993) :

Propriété 1 : Consistance

$$\forall j \in \{1, \dots, n\}, \xi_i(0, \dots, \delta v_{ij}, \dots, 0) = \delta v_{ij}.$$

Propriété 2 : Multi-Symétrie

$$\forall j, k \in \{1, \dots, n\} \text{ et } j \neq k, \xi_i(\delta v_{i1}, \dots, \delta v_{ij}, \dots, \delta v_{ik}, \dots, \delta v_{in}) = \xi_i(\delta v_{i1}, \dots, \delta v_{ik}, \dots, \delta v_{ij}, \dots, \delta v_{in}).$$

On peut facilement montrer que tout modèle défini par $\langle U, V, M, C \rangle$ peut se ramener à un triplet $\langle U, V, F \rangle$ (e.g. $f_i(v_1, \dots, v_n) = \xi_i(M_{i1}^p(v_1), \dots, M_{in}^p(v_n))$, où $M_{ij}^p(v_j) = M_{ij}(v_j)$ si V_j influence V_i , sinon $M_{ij}^p(v_j) = 0$).

Les deux définitions de modèles causaux que nous venons de donner ne sont pas dépendantes d'une représentation particulière des valeurs des données. Rien n'interdit d'ailleurs que la représentation de l'information ne change d'une variable à l'autre à l'intérieur d'un même modèle. Ces deux définitions mettent en avant le caractère distribué et non supervisé de la propagation de l'information dans un modèle causal. Chaque variable et influence représente une partie de la connaissance du modèle. Le traitement de l'information est local, réalisé par les fonctions de l'ensemble F (ou C et M, suivant la définition). Dans les définitions que nous venons de donner, le traitement ne dépend que des variables considérées et de la fonction associée. Il n'est pas dirigé par un système superviseur qui exercerait une contrainte sur le propagation des valeurs par les relations de causalité. On peut d'ailleurs remarquer à ce sujet que le système Ca~En fait exception à cette règle puisqu'il permet de construire des modèles mixtes, représentant des connaissances causales et non-causales, ces dernières supervisant la propagation causale (Travé-Massuyès *et al.* 1997).

6. CONCLUSION

Comme nous l'avons souligné, les modèles causaux sont largement utilisés dans tous les domaines car ils représentent un type de raisonnement naturel pour l'humain. Néanmoins, la relation de causalité s'applique tout particulièrement au cerveau humain. En effet, le fonctionnement du cerveau est fondamentalement causal, comme l'illustre l'activité neuronale. L'activation d'une certaine population de neurones va *causer* par l'intermédiaire de la propagation électrique axonale et des échanges chimiques synaptiques, l'activation des neurones qui lui sont connectés, et ainsi de suite.

Le fonctionnement neuronal implique de même l'existence d'une chronologie des événements cérébraux. Cette chronologie peut être déterminée par les techniques d'imagerie de surface (i.e. l'E.E.G. et la M.E.G., voir page 20). Un modèle de ce fonctionnement doit donc prendre en compte la temporalité des relations et le formalisme dans lequel le modèle est représenté doit permettre la manipulation explicite du temps.

Pour résumer, on peut énoncer les contraintes que notre formalisme devra satisfaire : une représentation de la causalité, une prise en compte explicite du temps, une haute transparence sémantique et une représentation de l'imprécision.

FORMALISATION DE L'ACTIVATION FONCTIONNELLE

Ce chapitre décrit le langage formel que nous avons élaboré pour le développement de modèles qualitatifs distribués de l'activation fonctionnelle du cerveau humain. Il reprend à son compte les observations qui se sont accumulées précédemment. Le formalisme que nous allons bâtir sera utilisé par le simulateur que nous proposerons au chapitre suivant. L'intégrité des définitions que nous allons formuler ici est donc primordiale pour la suite de notre travail.

Nous allons, dans un premier temps, synthétiser les notions fondamentales qui régissent notre formalisation. Ces notions fondamentales nous serviront à formaliser notre approche de manière à satisfaire à nos soucis de plausibilité et de transparence sémantique. Elles dirigeront toute la suite de nos travaux. Après ce premier volet, nous entrerons dans le détail de notre formalisme en décrivant l'architecture de nos modèles, la représentation de l'information et finalement comment propager cette information au travers de cette architecture.

1. NOTIONS FONDAMENTALES

Comme nous l'avons déjà dit, l'objectif premier du formalisme que nous allons présenter est d'exprimer la plus grande gamme possible d'hypothèses sur la propagation de l'information cérébrale. Ceci sous-entend que le formalisme doit être bâti sur un nombre minimal de postulats. Moins le formalisme aura été contraint, plus le nombre d'hypothèses explicitement exprimables par les modèles sera grand.

Nous allons donc distinguer deux sortes d'hypothèses. Premièrement, les hypothèses fondamentales (ou *hypothèses fortes*) sur lesquelles notre formalisme est bâti et, secondement, les hypothèses exploratoires (ou *hypothèses faibles*) qui seront explicitement représentées par les modèles. Les hypothèses exploratoires doivent être compatibles avec les hypothèses fondamentales que nous allons formuler.

Nous conservons le terme d'hypothèses pour les deux classes. Car, bien que nos hypothèses fondamentales fassent l'objet d'un large consensus, elles sont fortement liées à l'état actuel des connaissances en neuropsychologie.

1.1. Processus cognitifs modélisés

Notre objectif est de proposer un formalisme pour modéliser les processus cognitifs complexes mis en jeu lors de la réalisation d'une tâche particulière. Ces processus ont pour substrat par des réseaux de régions cérébrales fonctionnelles. Ils mettent en œuvre des mécanismes de traitement automatiques, pilotés par les données, ainsi que des mécanismes contrôlés, pilotés par les buts (Pastor 1999, Pastor *et al.* 1997). La toute première des hypothèses fondamentales de notre approche est née de l'observation de la rapidité du traitement de l'information quand le cerveau est soumis à un stimulus. Nous supposons que l'efficacité et la vitesse de ce traitement sont très probablement une manifestation des mécanismes automatiques, dirigés par les stimuli (i.e. stimuli-dépendants), prépondérants dans l'activité cérébrale (Pastor 1999). C'est pourquoi nous focaliserons nos modèles sur l'activité produite en réponse à un stimulus de l'environnement :

(Hypothèse I) L'activité représentée est celle directement produite par le traitement des stimuli.

C'est parce qu'ils représentent une grande partie de l'activité cérébrale, qu'il est intéressant de s'intéresser à la modélisation de ces processus. De plus, s'il sont réellement à la source de l'efficacité du fonctionnement cérébral, leur représentation devient indispensable dans un modèle fonctionnel physiologiquement plausible. Ces processus possèdent aussi l'avantage d'être relativement mieux connus que les processus contrôlés, ce qui facilitera sans doute leur modélisation. Se concentrer sur le traitement piloté par les stimuli permet d'étudier le cerveau comme un système automatique dynamique (Pastor *et al.* 1997, Lacotte 1996) dont la mise en œuvre produit une activité qui peut être mesurée par les techniques d'imagerie neuro-fonctionnelle.

1.2. Architecture des modèles

Comme nous l'avons déjà vu, notre approche se situe à un niveau organisationnel global (c.f. Chapitre 2, page 50), c'est-à-dire au niveau correspondant le plus à celui des observations obtenues par les techniques d'imagerie fonctionnelle. A ce niveau, l'activité cérébrale se manifeste par

l'activation de régions du cerveau de tailles et formes variables. A ce niveau de détail, il paraît raisonnable d'adopter les trois hypothèses fondamentales suivantes :

(Hypothèse II) Une région cérébrale est une partie connexe du cerveau, localisable géographiquement (i.e. identifiable même si elle ne peut être située avec précision). Certaines régions peuvent être décomposables en un ensemble dénombrable de sous-régions (i.e. des régions de taille plus réduite et considérées comme géographiquement contenues dans la région qu'elles composent).

(Hypothèse III) Le traitement de l'information est réalisé par les régions cérébrales, les liens anatomiques qui les unissent régissent le transfert de l'information à traiter.

(Hypothèse IV) Les liens anatomiques peuvent se ramener à deux types de connexions : les connexions distales et les connexions de proximité. Les connexions distales relient les régions par des faisceaux d'axones, les connexions de proximité sont plus diffuses. Cependant, elles sont toutes deux orientées et impliquent de nombreux axones. Les délais de transmission de l'information le long de ces connexions sont variables en fonction de la connexion considérée.

Les régions cérébrales considérées peuvent, par exemple, être les aires fonctionnelles de Brodmann. Ou bien, elles peuvent être des formations de plus grande surface comme des gyri, voire des lobes du cortex, ou au contraire des formations plus fines, tels que les modules cérébraux (e.g. les colonnes corticales). Le choix des régions à représenter est dépendant de l'activité à modéliser, mais aussi de la précision de l'observation qui aura permis de bâtir le modèle.

L'hypothèse II stipule que les aires sont localisables anatomiquement. Nous restons cependant très vagues quant à cette localisation, compte tenu de la précision des observations que l'on peut obtenir et de la variabilité interindividuelle. Pour formaliser la notion d'unité de traitement de l'information dans le cerveau nous allons définir la notion de *zone fonctionnelle cérébrale* (Pastor 1999) :

Définition 3 : Zone fonctionnelle

Une zone fonctionnelle cérébrale⁴ est une région connexe du cerveau fonctionnellement cohérente et topographiquement délimitée par l'ensemble des connexions qui l'unissent avec les autres zones fonctionnelles du cerveau.

On ne s'intéresse pas à la structure anatomique propre d'une zone. Les seules informations topographiques que l'on a besoin de connaître sur une zone sont les connexions qui l'unissent aux

autres zones, car, au sens de l'hypothèse IV, les connexions sont seules indispensables à la propagation de l'information. On ne se soucie pas, par exemple, de la taille ou de tout autre aspect morphométrique des zones.

La cohérence fonctionnelle est une notion complexe, difficile à formaliser. Elle permet de bâtir un modèle purement fonctionnel pour une zone. Ce modèle décrit la manière de traiter l'information qui lui parvient. Nous allons donc définir deux niveaux de modélisation, un *niveau structurel*, définissant le réseau de zones impliqué dans le traitement, et un *niveau fonctionnel* définissant, pour chaque zone son mode opératoire.

1.2.1. Modélisation à 2 niveaux

Les deux niveaux de modélisation sont clairement distincts de par les éléments qu'ils représentent. On peut cependant relever certaines propriétés communes. L'élément central de notre approche est la zone. Les hypothèses concernant l'identité des zones et les connexions inter-zones relèvent du niveau structurel. Les hypothèses concernant la fonction d'une zone, les transformations internes de l'information et la manifestation locale (le niveau d'activité) de ces transformations, relèvent du niveau fonctionnel.

Définition 4 : Modèle structurel

Un modèle structurel décrit le réseau de zones que l'on suppose impliquées dans la résolution de la tâche cognitive que l'on souhaite modéliser. Chaque zone du réseau correspond à une région cérébrale.

Cependant, suivant les hypothèses que l'on souhaite exprimer dans le modèle et suivant les observations que l'on peut obtenir sur cette région, on peut soit garder la correspondance zone/région, soit décomposer la région en sous-régions (hypothèse II). Si l'on souhaite exprimer des *hypothèses structurelles* sur le fonctionnement interne de la région considérée, on doit la représenter par un modèle structurel décrivant la connectique des régions plus élémentaires qui composent la région. Dans le modèle structurel original, la zone correspondante sera donc remplacée par un sous-réseau de zones. Si l'on souhaite exprimer des *hypothèses fonctionnelles* sur le fonctionnement interne de la région considérée, on la représentera par un modèle purement fonctionnel.

⁴ On emploiera, pour simplifier, le terme de zone.

Cette décomposition structurelle introduit une organisation de complexité décroissante du modèle structurel. Cette organisation est utile pour simplifier la description de la structure du modèle. En effet, plus le nombre de zones fonctionnelles est important, plus la complexité du modèle structurel est élevée. Pour réduire le nombre d'éléments d'un modèle structurel, on peut ainsi identifier des sous-réseaux pour les remplacer chacun par un seul élément.

Définition 5 : Modèle fonctionnel

Le modèle fonctionnel d'une zone décrit le traitement de l'information propre à cette zone. Il n'y apparaît aucune hypothèse structurelle, c'est-à-dire aucune décomposition anatomique.

La division du modèle en deux niveaux distincts nous permet d'utiliser, au niveau fonctionnel, un formalisme de haute transparence sémantique. Au niveau structurel, chaque zone est connectée à d'autres zones par des connexions orientées (hypothèse IV). Le but du modèle fonctionnel est de décrire la manière dont l'information entrant dans une zone donnée est traitée et de préciser quelle sera l'information que la zone transmettra à son tour. Donc, le fonctionnement interne de la zone n'est contraint que par les connexions définies au niveau structurel. Les hypothèses fonctionnelles ne portent que sur les transformations que l'on appliquera à l'information.

Comme dans une approche fonctionnelle classique (Chapitre 2, partie 1), la fonction d'une zone peut être décomposée en fonctions plus simples, ou *primitives fonctionnelles*. Les primitives fonctionnelles interagissent et manipulent l'information entrante, pour produire l'information qui sera propagée en aval. Chaque primitive fonctionnelle est représentée par un *processeur élémentaire*. Une fonction n'est une primitive qu'à un niveau de représentation fonctionnelle donné. Une meilleure connaissance en neurophysiologie peut amener à faire perdre le statut de primitive à une fonction en la décomposant en primitives fonctionnelles plus simples. Inversement, on peut, pour obtenir un modèle plus simple, réunir plusieurs primitives (e.g. mémoire de travail, mémoire à court terme et mémoire à long terme) en une primitive moins précise (e.g. mémoire).

Cette décomposition des primitives fonctionnelles introduit une organisation hiérarchique, de complexité décroissante du modèle fonctionnel. Cette décomposition n'est pas sans rappeler celle du niveau structurel. Ces mécanismes de décomposition sont indépendants du formalisme choisi pour représenter l'information. Les modélisations hiérarchiques sont d'ailleurs de plus en plus souvent proposées comme une approche efficace de la représentation des systèmes complexes et ce, quel que soit le système en question ou le formalisme employé (Koller & Pfeffer 1997, Kuipers 1996, Okuda & Ushio 1991, Zeigler 1990). Outre une vision descendante de la complexité d'un

système, une approche hiérarchique offre aussi une grande modularité aux modèles. Cette modularité permet de réutiliser à moindre coût une partie d'un modèle dans un autre modèle. Cette réutilisation peut être réalisée au niveau structurel comme au niveau fonctionnel.

1.2.2. Graphes d'influences

Les trois hypothèses fondamentales (II), (III) et (IV) nous conduisent à un formalisme fondé sur la notion de graphe orienté. En effet, les zones sont les processeurs de l'information et peuvent être représentées par les nœuds du graphe (hypothèses II et IV)). Les connexions, identifiées anatomiquement, peuvent être modélisées par des arcs (hypothèses III et IV). Cette représentation structurelle simpliste nous permet de nous détacher de toute localisation, celle-ci étant trop hasardeuse du fait de la variabilité interindividuelle. Mais nous conserverons tous les paramètres essentiels, liés au transfert de l'information comme, par exemple, le délai de transmission de l'information entre zones. De plus, cette représentation a l'avantage d'identifier, de manière individuelle, tous les éléments impliqués dans la propagation (processeurs et chemins). De ce fait, l'expression d'hypothèses sur ces éléments s'en trouvera simplifiée. Une représentation qui n'aurait pas expressément identifié les connexions, n'aurait pas permis une représentation explicite des hypothèses les concernant.

Les graphes orientés peuvent aussi être utilisés au niveau fonctionnel. Ils permettent la représentation des relations causales entre les diverses primitives fonctionnelles qui y seront représentées. Ils facilitent l'identification des primitives fonctionnelles (les nœuds) et leurs interactions (les arcs). De cette façon, la décomposition en sous-niveaux fonctionnels s'en trouve facilitée, tout comme l'expression d'hypothèses, explicitées sous la forme d'interactions entre ces primitives.

Aucune des hypothèses fondamentales citées n'intervient directement sur le niveau fonctionnel. On pourrait donc envisager d'y utiliser un formalisme différent. Cependant, il ne faut pas omettre la dépendance du niveau fonctionnel vis à vis du niveau structurel. En effet, tout modèle fonctionnel est dépendant de la zone auquel il est associé, elle-même connectée dans un graphe orienté. L'information entrant et sortant du modèle fonctionnel doit transiter dans un graphe orienté. Cette structure ne contraint en rien la forme de l'information transmise, mais elle la prédispose à un traitement homogène dans le modèle fonctionnel.

L'utilisation d'influences, pour représenter les relations causales entre les processeurs fonctionnels, se justifie par l'hypothèse I qui stipule que nous ne représentons uniquement des processus

automatique. Une représentation binaire des relations suppose l'absence d'interactions directes entre deux chemins de propagation différents. C'est à dire qu'un processeur donné ne peut exercer de contrôle sur l'information propagée entre deux autres processeurs.

1.2.3. Modèles génériques

La notion de modèle générique est une notion indépendante des deux précédentes. Cette notion ne s'est d'abord appliquée qu'au fonctionnement propre des zones, elle dépend d'une cinquième hypothèse fondamentale qui peut être exprimée comme suit :

(Hypothèse V) Certaines régions cérébrales, aux rôles différents, peuvent avoir des structures fonctionnelles identiques.

Cette hypothèse a été émise d'après des résultats expérimentaux montrant la plasticité cérébrale et la faculté de réorganisation corticale après lésion. On peut en déduire que certaines régions possèdent des capacités fonctionnelles semblables sans pour autant être totalement identiques. On peut ainsi définir des *classes fonctionnelles* réunissant plusieurs zones. Cela pourrait, par exemple, être le cas des cortex primaires.

Un modèle fonctionnel générique définit un modèle partiellement spécifié. Ce modèle ne précise que les éléments qui sont communs à la classe fonctionnelle correspondante. Ce modèle définit en quelque sorte un *modèle* de modèle fonctionnel. Un modèle fonctionnel générique ne peut être utilisé directement comme modèle de zone. Pour ce faire, il faut terminer la spécification. Le modèle alors obtenu est appelé *instanciation* du modèle générique. Le niveau de généralité est défini par le nombre de fonctionnalités partagées par les zones d'une même classe : moins il y aura de fonctionnalités communes, et plus le modèle sera générique. En effet, un modèle générique spécifiant peu de fonctionnalités sera plus réutilisable. Il conviendra à une classe regroupant un plus grand nombre de zones différentes.

La notion de généralité peut être étendue au niveau structurel. En effet, prenons l'exemple de deux modèles partageant la même structure mais faisant intervenir des zones fonctionnellement différentes, comme, par exemple, dans une certaine mesure, l'hémisphère droit et l'hémisphère gauche du cortex. Dans ce cas, les modèles fonctionnels seront différents mais les modèles structurels seront identiques. On pourra alors définir un modèle structurel générique en ne spécifiant pas les zones impliquées dans la structure décrite. Cette spécification sera achevée lors de l'instanciation du modèle structurel.

L'utilisation de modèles génériques augmente significativement la réutilisabilité des modèles. Cela diminue d'autant la complexité de l'élaboration des modèles d'une même classe, puisqu'il suffira de définir un seul modèle pour toute la classe, puis de l'instancier pour chaque élément.

1.3. Représentation de l'information

Dans le cerveau, au niveau organisationnel global, l'information est issue du signal produit par l'activité intégrée de plusieurs millions de neurones. Ainsi, à un instant donné, l'information transmise dans un faisceau d'axones reliant deux régions corticales est la composition de différents facteurs. Par exemple, on peut citer le nombre de neurones déchargeant à un instant donné, la fréquence de leurs décharges, la nature inhibitrice ou excitatrice des synapses concernées, mais aussi la chronologie des décharges des neurones du faisceau et la configuration des neurones qui déchargent. Quelle que soit la nature réelle de l'information cérébrale, il n'existe à l'heure actuelle aucun moyen non invasif pour l'observer dans son ensemble. Les mesures obtenues par les techniques d'imagerie fonctionnelle sont des mesures très intégrées, ne donnant qu'une idée approximative de la *localisation* et de la *chronologie* de l'activité. De plus, les mesures réalisées lors d'études d'activation sont des *mesures relatives* obtenues par rapport au « bruit » de l'activité globale, non spécifiquement dédiée à la résolution de la tâche, ou sont des mesures soustraites à l'activité générée par la résolution d'une tâche de référence.

Notre objectif étant de formuler des modèles construits en partie sur des observations d'images fonctionnelles, il nous faut conserver dans la représentation de l'information, une notion plus ou moins directe de volume d'activité. Ce volume d'activité est en effet hautement significatif de certains aspects du traitement de l'information cérébrale. Mais cette mesure ne peut suffire à caractériser cette information. Deux activités de même volume peuvent différer selon les neurones effectivement impliqués. Cette observation illustre la double dimension de l'activité liée au traitement d'une information. En plus du volume, mesurable, l'activité est aussi caractérisée par certaines propriétés de la population des neurones qui s'active spécifiquement pour l'information traitée et non directement mesurables chez l'humain.

Nous considérerons, dans nos modèles, l'information comme une abstraction du signal cérébral. Une modélisation explicite de la propagation de l'information suppose que l'information soit accessible dans une forme suffisamment formelle et compacte pour pouvoir être directement manipulée. Le degré d'abstraction de l'information et sa représentation formelle sont contraints par l'architecture que nous avons définie dans la première partie de cette présentation.

Nous avons choisi de représenter l'information comme un couple dont les parties sont une abstraction du volume d'activité et de la catégorie du signal cérébral. Pour ce faire, nous posons l'hypothèse suivante :

(Hypothèse VI) Au niveau de l'activité globale du cerveau, le signal cérébral peut se réduire à l'expression duale d'une quantité d'énergie propre et d'une représentation catégorielle.

Nous définissons une représentation formelle de l'information en deux parties complémentaires : *l'amplitude* et le *type* de l'information, se rapportant chacune respectivement à l'énergie et à la catégorie de l'information (Lacotte 1996).

1.3.1. Amplitude de l'information

L'amplitude de l'information (Pastor *et al.* 2000, Pastor 1999, Lafon *et al.* 1998) est considérée comme une abstraction du volume de l'activité engendrée par le signal cérébral. Cette partie de l'information est au niveau biologique, une abstraction du nombre et du taux de décharge des neurones activés. Elle est assez proche de *l'énergie* observée par les techniques d'imagerie cérébrale. Elle ne représente pas directement une variation de débit sanguin ni une valeur du gradient d'un champ électrique. Cependant, elle peut être interprétée comme s'y rapportant, ou du moins être comparée aux mesures de l'imagerie.

Le fait que l'amplitude soit directement représentée dans le modèle permet d'exprimer des hypothèses en terme d'activation de zone dans nos modèles. On peut aussi appliquer des transformations à l'amplitude pour la confronter aux mesures de l'imagerie. Un nœud du modèle fonctionnel d'une zone pourra, par exemple, exprimer la variation du débit sanguin produit par le traitement réalisé par la zone et être assimilé aux mesures tomographiques obtenues par TEP ou IRMf.

Le caractère numérique de la valeur des mesures de l'activité cérébrale est indéniable. Toutefois, les mesures de l'imagerie fonctionnelle sont assez imprécises. Elles sont relativement éloignées de l'activité neuronale. Dans la majorité des cas, le rapport qui les lie à cette activité ne peut être considéré que qualitativement. Par exemple, outre les diverses étapes de traitement du signal, la méconnaissance du filtre métabolique entre l'activité neuronale et les variations du débit sanguin en IRMf et TEP, ou bien l'intégration spatiale du signal en EGG, ne permettent pas de considérer les mesures obtenues comme des mesures quantitatives de l'activité cérébrale.

La représentation la plus simple et la plus neutre que l'on puisse associer à la représentation qualitative d'une valeur numérique est un intervalle réel. Les intervalles permettent de représenter simplement une information mal connue. De plus, les relations qualitatives peuvent facilement être exprimées sous la forme d'influences causales (Travé-Massuyès *et al.* 1997, MQ&D group 1996) ce qui nous convient tout à fait.

1.3.2. Type de l'information

Le type de l'information représente une abstraction de la « catégorie » du signal cérébral (Pastor *et al.* 2000, Pastor 1999, Lafon *et al.* 1998), c'est-à-dire de propriétés telles que la nature et la distribution de la population de neurones qui s'active pour l'information en question. La correspondance entre catégorie d'information et population de neurones est illustrée par l'organisation des aires primaires en somatotopie, tonotopie ou rétinitopie (cf. Chapitre 1, page 15, Arbib *et al.* 1998). Cette notion est cohérente avec la faculté de catégorisation propre au cerveau.

Certaines études neuropsychologiques mettent en évidence l'existence d'une telle organisation au delà des aires primaires, notamment en ce qui concerne la motricité, dans les aires motrices secondaires et dans les régions sous-corticales (Alexander *et al.* 1992). À l'évidence, la catégorie de l'information propagée ne peut être directement mesurée et il est difficile de définir exactement sa valeur. Nous avons décidé de la considérer comme un ensemble de symboles. Pour pouvoir manipuler aisément la catégorie, nous représenterons cet ensemble par un ensemble flou, le type de l'information. Un type symbolique est une forme tout à fait apte à représenter des catégories (Rosch 1978) et il satisfait à notre exigence de transparence sémantique.

L'adoption d'une représentation symbolique, nous rapproche du postulat de l'approche fonctionnelle symbolique classique, considérant que toute pensée est exprimable dans un langage formel. Cependant, nous ne considérons cette expression que comme un simple élément du couple constituant l'information et donc insuffisante à la caractériser entièrement. Même si nous réduisons la portée de ce postulat, il reste indispensable car, si l'on veut augmenter la transparence sémantique de nos modèles, l'information doit être accessible sous une forme compacte et explicite.

De même que le volume d'activité est capable de croître au cours d'une tâche, le type de l'information peut aussi évoluer. En l'absence de connaissances précises, nous avons défini des opérateurs basiques sur les types, comme la *combinaison*, utilisée pour produire de nouveaux types à partir de types différents. On peut aussi représenter des opérations plus physiologiquement fondées, comme le filtrage, qui représente un mécanisme des aires corticales, communément admis,

le seuillage d'entrée (input gating). Ce dernier mécanisme provoque l'*atténuation* de tout ou partie d'un type.

1.3.3. Conclusion

Amplitude et type sont deux notions complémentaires, qui reflètent l'énergie nécessaire au traitement de l'information et sa catégorie. On peut faire une correspondance directe entre l'information et toute population de neurones pouvant la traiter. L'amplitude peut en effet être rapprochée du nombre de neurones activés à un instant donné. Le type peut être rapproché, par exemple, de l'organisation des neurones effectivement impliqués dans l'activité qui a produit cette amplitude. La représentation de l'information sous la forme d'un couple amplitude et type est conservée dans la propagation de l'information entre deux zones.

La figure ci-dessous (Figure 15, page 91) peut être une illustration d'un échange de deux informations différentes entre deux zones cérébrales, A et B. Si l'on était capable de l'obtenir, on pourrait comparer la coupe des faisceaux d'axones reliant les deux zones. A deux instants différents, on observerait une même amplitude, (12 neurones déchargent) mais deux types différents (ce ne sont pas les mêmes neurones qui déchargent).

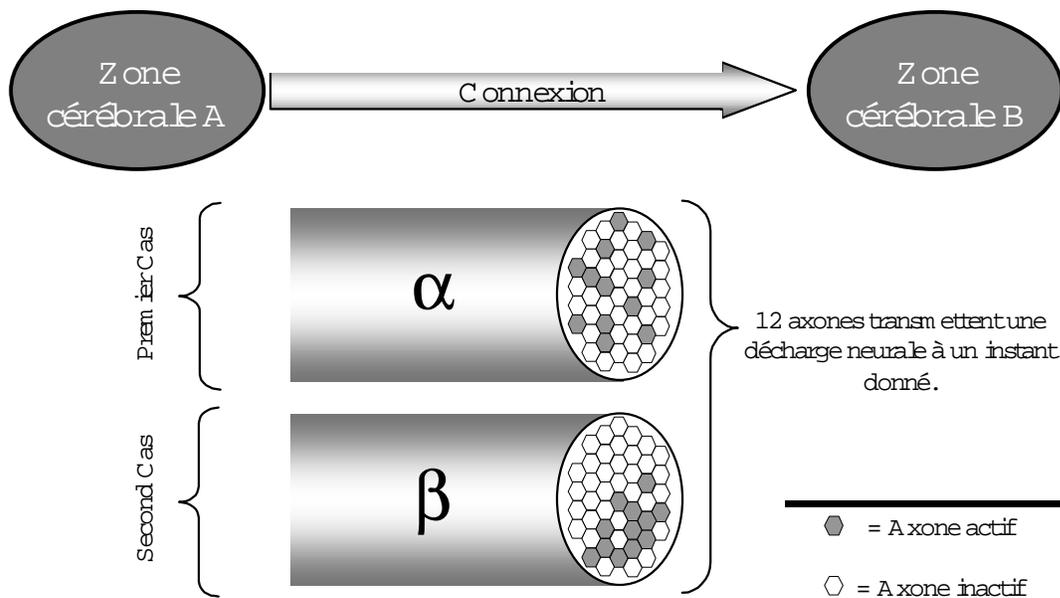


Figure 15: Illustration neuronale du couple Amplitude /Type.

Si amplitude et type sont nécessaires à la représentation de l'information, ils jouent un rôle différent dans la validation des modèles. Contrairement à l'amplitude, il n'est pas possible d'obtenir chez

l'Homme des mesures de la catégorie du signal cérébral et donc d'y confronter les valeurs des types propagées dans nos modèles. Cette constatation renforce l'utilité de l'amplitude, car elle est la seule partie de l'information qui soit mesurable. L'amplitude reste donc le seul élément qui puisse permettre la validation d'un modèle par confrontation aux données de l'imagerie fonctionnelle.

Pour conclure, on peut remarquer que cette représentation en couple (amplitude, type) peut aussi s'appliquer aux stimuli en entrée des réseaux cérébraux. On peut, par exemple, caractériser un *ton pur* par une intensité (amplitude) et une fréquence (type). Cette représentation est cohérente avec l'organisation tonotopique des aires auditives primaires dont on sait qu'elles possèdent une organisation en sous régions dont chacune est spécifiquement dédiée à la détection d'une plage de fréquences sonores particulières.

2. ARCHITECTURE DES MODELES

Pour représenter tous les niveaux de nos modèles avec des graphes, nous définissons ici une représentation orientée composant, permettant de décomposer un nœud d'un graphe en un sous-graphe. Cette architecture est commune aux niveaux structurel et fonctionnel. Au premier de ces niveaux, les éléments des graphes représentent les zones cérébrales et les connexions anatomiques. Au second niveau, les éléments des graphes représentent les primitives fonctionnelles et les influences causales entre ces primitives.

2.1. Les graphes causaux de processeurs

Nous avons déjà donné la définition d'un graphe causal au chapitre précédent (pages 77 et suivantes). Dans un tel graphe, les arcs matérialisent les influences de cause à effet et les nœuds représentent les variables d'état du système. Dans notre application, les nœuds sont les *processeurs fonctionnels*, qui traitent et transforment l'information. Les arcs sont les *connexions*, qui permettent la propagation de l'information.

Définition 6 : Graphe causal

Un graphe causal, G , est représenté par un couple $\langle N, C \rangle$, où N est l'ensemble des nœuds, les processeurs, et C l'ensemble des connexions orientées entre les processeurs.

Encore une fois, cette représentation est valide avec les deux niveaux de notre modélisation et nous ne distinguerons donc pas formellement le niveau structurel du niveau fonctionnel. De cette manière, les notions que nous aborderons s'en trouveront grandement simplifiées et seront valides pour les deux niveaux de modélisation.

2.1.1. Les nœuds du graphe

Pour reprendre ce qui a été dit précédemment, l'idée maîtresse qui nous a permis d'élaborer notre architecture est l'idée selon laquelle le nœud d'un graphe (une zone ou une primitive fonctionnelle) peut, en cas de trop grande *complexité*, être décomposé en éléments plus *simples*. Ces éléments peuvent aussi être décrits par un graphe dont ils seront à leur tour les nœuds. C'est cette idée d'association des nœuds d'un graphe à d'autres graphes que reprend la notion de *composant*, directement inspirée de la modélisation objet (Zeigler 1990).

Définition 7 : Composant

Un composant est un nœud d'un graphe, *défini* par un autre graphe, appelé graphe du composant. Ce second graphe décrit les éléments qui *constituent* le composant et les relations qui les unissent.

Définition 8 : Atome

Un atome est un nœud représentant un processeur indivisible, il est associée à une valeur qui varie au court du temps.

Un atome est donc un « nœud classique », représentant un processeur élémentaire. Il n'est donc pas associé à un graphe. Les nœuds de nos graphes sont donc soit des atomes, soit des composants. Le graphe définissant un composant peut, à son tour, contenir des composants, suivant la complexité des éléments qu'il doit décrire. On peut ainsi construire tout un emboîtement de graphes. Ce mécanisme simplifie la définition d'un modèle car il permet de bénéficier d'une grande *réutilisabilité*. En effet, il est possible de définir deux composants en utilisant le même graphe. Les deux composants, bien que différents, auront le même comportement, comme deux copies d'un même objet.

2.1.2. Les connexions du graphe

Les connexions sont des arcs orientés. Elles possèdent donc une *origine* et une *destination*. Elles sont identifiées par les nœuds correspondants. Si A et B sont deux nœuds d'un graphe G, on notera $(A \rightarrow B)$, la connexion dirigée du nœud A vers le nœud B. On dit aussi que A est l'*antécédent* de B

dans le graphe G. Il existe en tout quatre sortes de connexions, suivant que la source et la destination de la connexion sont des atomes ou des composants. On réserve le terme d'influence à la connexion de base qui relie deux atomes.

Définition 9 : Influence

On appelle influence une connexion dont la source et la destination sont des atomes.

Seuls les atomes possèdent une valeur à proprement parler. Comme dans un graphe causal classique, la valeur d'un atome est calculée en fonction des connexions dont l'atome est la destination. Donc, on utilise, pour calculer la valeur d'un atome, les valeurs de ses antécédents. La difficulté est d'obtenir la valeur à utiliser lorsque l'antécédent n'est pas un atome mais un composant.

Dans la figure ci-dessous (Figure 16), le graphe G illustre différentes sortes de connexions. La connexion (A1 → A2) est la seule influence du graphe. Deux connexions étant dirigées vers A2 (i.e. (A1 → A2) et (C1 → A2)), pour calculer la valeur de cet atome, on utilisera la valeur de l'atome A1, ainsi que la valeur issue du composant C1. La valeur de l'atome A1 sera utilisée pour calculer la valeur soumise en entrée du composant C1, compte tenu de la connexion (A1 → C1).

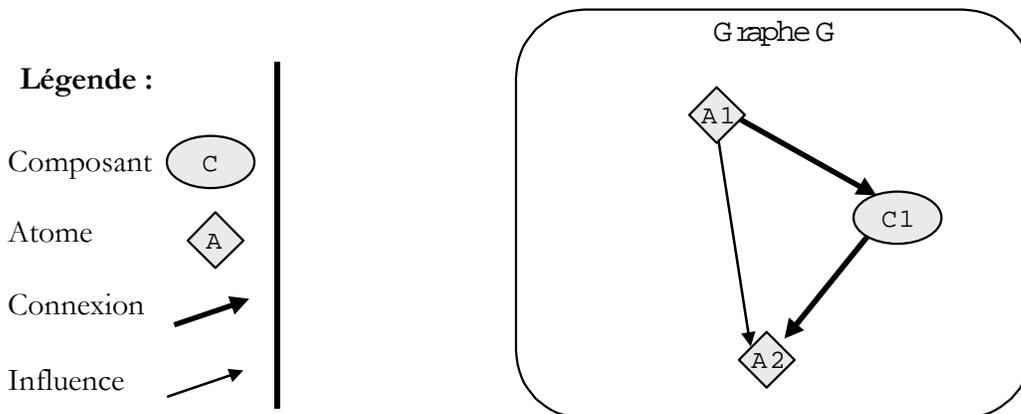


Figure 16 : Différentes sortes de connexions.

Il y a deux façons de décrire l'implication d'un composant dans une connexion. La façon la plus simple consiste à déterminer ses *points de connexion*. La seconde façon, plus complète, réside dans la définition de ses *schémas de connexion*. Chaque moyen peut être utilisé en entrée et en sortie d'un composant.

Définition 10 : Point de connexion

Un point de connexion est un nœud du graphe du modèle du composant. On distingue les *points d'entrée* des *points de sortie*. Les premiers peuvent être utilisés pour définir une connexion dirigée vers le composant. Les seconds sont utilisés par les connexions dont le composant est la source.

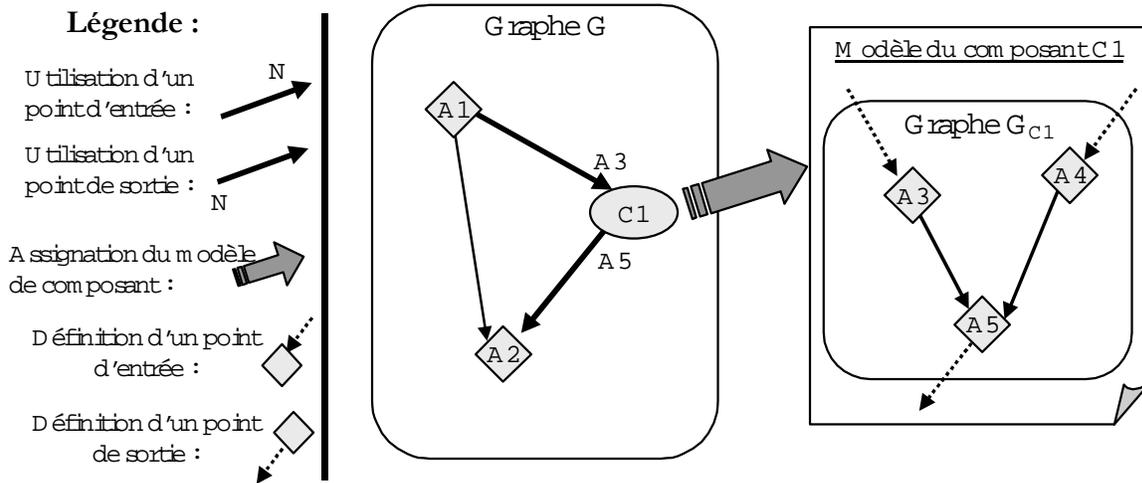


Figure 17: Points de connexion d'un composant.

Les points d'entrée et de sortie sont fixés à la définition du composant. Chacun de ces points peut être utilisé zéro, une ou plusieurs fois. En reprenant l'exemple précédent, chaque connexion utilisant un composant doit faire un choix suivant le nombre de points de connexion définis par le modèle du composant (Figure 17). Si le modèle du composant C1 spécifie deux points d'entrées, A3 et A4, et un seul point de sortie, A5, la connexion (A1 \rightarrow C1) du graphe G doit préciser quel point d'entrée elle utilise, soit ici le point A3 (indiqué, sur la figure, à la pointe de la flèche). De même, la connexion (C1 \rightarrow A2) doit désigner le point de sortie utilisé, soit l'atome A5. L'atome A4, qui est un point d'entrée possible du composant C1, n'est pas utilisé dans le graphe G. Il ne reçoit donc aucune influence.

L'utilisation de points de connexion suppose que le mode opératoire de la connexion est défini à l'extérieur du composant. Lorsque le mode opératoire est défini dans le composant, on fait appel aux schémas de connexion.

Définition 11 : Schéma de connexion

Un schéma de connexion d'entrée d'un composant est identifié par un nom et est défini par un ensemble de connexions dont le nœud source n'est pas spécifié. Respectivement, un schéma de connexion de sortie, ne spécifie pas les nœuds de destination des connexions qui le définissent.

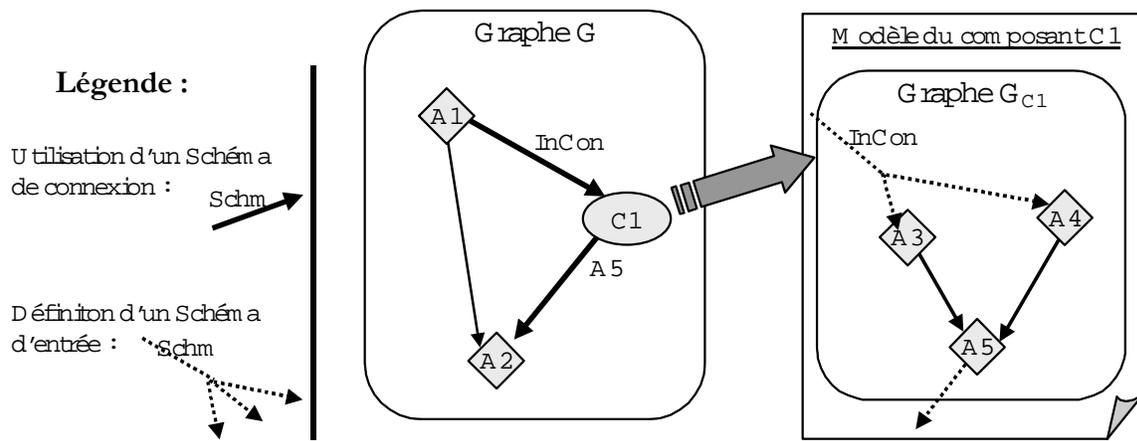


Figure 18 : Schémas de connexions.

Reprenons l'exemple précédent (Figure 16), mais admettons maintenant que le composant C1 définisse un schéma de connexion que nous nommerons, par exemple, *INCON*. Ce schéma est défini par l'ensemble $\{(\rightarrow A3), (\rightarrow A4)\}$ représentant deux connexions ne spécifiant pas leur nœud source. Toute connexion dirigée vers le composant C1 et utilisant ce schéma sera considérée à l'intérieur du composant comme une double connexion vers A3 et A4. Imaginons maintenant que la connexion $(A1 \rightarrow C1)$ utilise ce schéma de connexion (Figure 18, l'utilisation du schéma est révélée par l'indication de son nom, ici *INCON*). Pour C1 cette connexion représentera un couple de connexions, $(A1 \rightarrow A3)$ et $(A1 \rightarrow A4)$. L'utilisation de schémas de connexions nous permet donc de définir simplement des connexions qui en réalité peuvent être beaucoup plus complexes, en termes de nombre de points de connexion. Cette technique de connexion suppose donc que les modes opératoires sont dépendants du modèle du composant.

Définition 12 : Interface de connexion

Une interface de connexion est divisée en deux parties, l'interface d'entrée et l'interface de sortie. L'interface d'entrée, respectivement l'interface de sortie, peut spécifier zéro, un ou plusieurs points d'entrée, respectivement de sortie, dans le graphe définissant le composant. Ces points doivent être des nœuds de ce graphe. Elle peut aussi spécifier zéro, un ou plusieurs schémas d'entrée, respectivement de sortie. L'interface doit obligatoirement définir au moins un point, ou schéma, d'entrée et de sortie.

2.2. Modèle de composant

Pour être tout à fait précis, nous devons revenir sur la définition des composants, à laquelle il faut ajouter maintenant une interface de connexion.

Définition 13 : Modèle de composant

Un modèle de composant est un couple $\langle G, I \rangle$, où G est un graphe et I une interface de connexion.

Définition 14 : Composant (2)

Un composant est un nœud d'un graphe. Un composant est *défini* par son modèle de composant.

La notion de modèle de composant est très utile car elle augmente la réutilisabilité des graphes de composant, déjà évoquée en début de cette partie de chapitre. En effet, deux composants d'un même graphe peuvent utiliser le même modèle de composant. On considère alors qu'ils en sont deux copies indépendantes, qui sont deux *instances* du modèle de composant.

Cela permet donc, par exemple, de ne définir qu'une seule fois une fonctionnalité très courante et de la réutiliser dans divers modèles en réalisant autant d'instances que nécessaire. On définit cette fonctionnalité par un modèle de composant. Si on modifie sa définition (e.g. en modifiant le graphe qui la compose), on modifie alors automatiquement le comportement de tous les composants qui l'utilisent. La notion de modèle de composant met aussi l'accent sur la modularité de notre architecture. On peut, par exemple, tester un modèle de composant indépendamment du reste du modèle, de manière à faciliter la vérification de modèles de grande envergure.

D'une manière générale, l'association de graphes aux composants crée ainsi une relation d'ordre partiel entre les divers graphes utilisés, appelée « englobe ». Notée \triangleright , cette relation est définie par :

Définition 15 : Relation « englobe »

On dit qu'un graphe G *englobe* un graphe G' , $G \triangleright G'$, si et seulement si G contient un composant tel que G' soit le graphe du modèle de ce composant.

On peut par exemple illustrer cette relation en considérant un graphe G_1 contenant deux composants C_1 et C_2 (Figure 19). Le composant C_1 est défini comme l'instance du modèle de composant MC_1 et le composant C_2 est défini comme l'instance du modèle MC_2 . Le modèle MC_1 est composé d'une interface définissant un schéma de connexion en entrée et un point de sortie. Le

graphe de ce modèle, le graphe G2, contient un composant qui est, à son tour, défini comme l'instance du modèle MC3. On a donc $G1 \triangleright G2 \triangleright G4$ et $G1 \triangleright G3$.

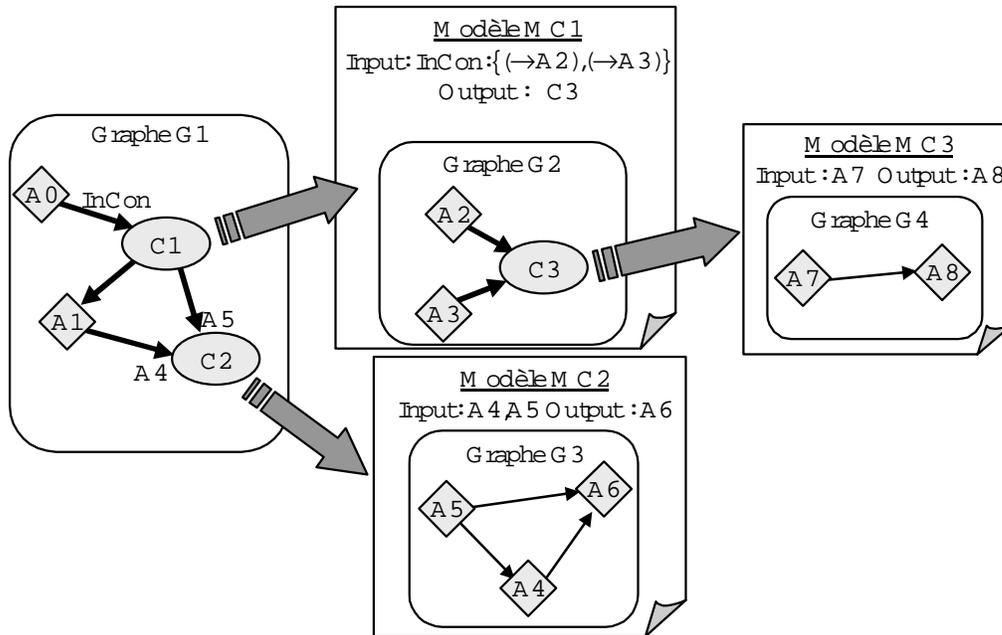


Figure 19: Exemple de la définition d'un graphe.

Selon notre approche de la modélisation orientée composant, un composant représente un élément de *complexité* inférieure à l'élément représenté par le graphe qui le contient. La notion de complexité change suivant le niveau de modélisation auquel on se situe, mais le principe reste le même. Par exemple, au niveau fonctionnel, un modèle fonctionnel d'une zone décrit la fonction réalisée par la zone, alors que tous les éléments qui le composent représentent des primitives fonctionnelles d'un degré de complexité inférieur (voir partie 1.2.1 de ce chapitre, page 85).

Dans notre approche, toute entité ne peut donc faire appel qu'à des éléments de complexité plus faible. Ceci nous conduit à interdire les définitions de graphes récursives. Si, pour définir un graphe G, on utilise, même indirectement, ce même graphe, il existe alors un cycle, selon la relation « englobe », c'est à dire une suite de graphes G_1 à G_n tels que $G \triangleright G_1 \triangleright \dots \triangleright G_n \triangleright G$. Ce cycle sous-entend qu'un composant du graphe G_n se définit grâce au graphe G. Or, cette définition va à l'encontre de notre approche, car le graphe G est forcément d'une complexité supérieure au graphe G_n . Par exemple (Figure 20), l'association du composant C4 au modèle MC1 rompt la décroissance de la complexité en créant un cycle dans la définition. MC1 ne peut être défini sans MC2, qui ne peut être défini sans MC3 qui dépend lui même de MC1.

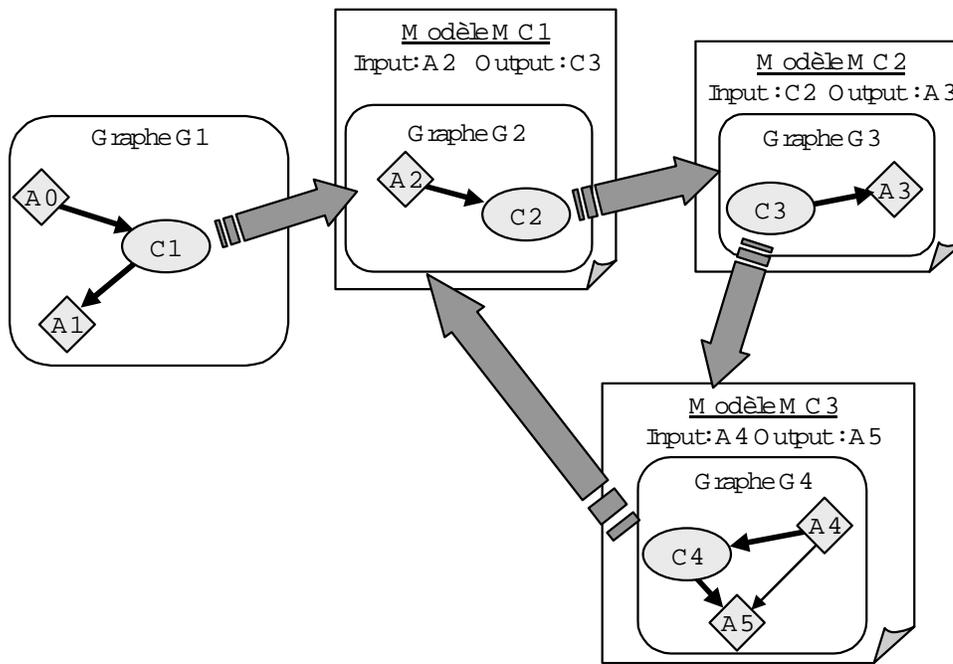


Figure 20 : Exemple d'une définition récursive.

Nous interdirons, de même, les définitions faisant intervenir un nombre infini de graphes, pour des raisons de calculabilité évidentes.

Définition 16 : Validité d'un graphe

On dira qu'un graphe G est valide seulement s'il ne contient aucun composant, ou bien, s'il est défini par un nombre fini de graphes, G_1 à G_n , tels que $G \triangleright G_i \ i \in \{1, \dots, n\}$ et que tous ces graphes sont valides.

On montre qu'un graphe dont la définition est récursive n'est pas valide.

2.3. Modèles de composants génériques

On a déjà souligné qu'un même modèle de composant peut être instancié par plusieurs composants dans un ou plusieurs graphes. Chaque instance est alors considérée comme une copie indépendante. Pour augmenter la réutilisabilité des modèles de composants et pour satisfaire à notre hypothèse fondamentale IV (page 87), nous avons défini la notion de modèle de composant générique. Par abus de langage, on parlera de modèle générique.

Un modèle de composant générique est un triplet $\langle G, I, P \rangle$ où G est un graphe, I est une interface de connexion définie sur G et P un ensemble de paramètres génériques utilisés dans I ou dans G .

Tout élément du graphe (e.g. nœud ou connexion) peut être remplacé par un paramètre générique. Les paramètres sont identifiés par un nom et tiennent la place de ce qu'ils remplacent. Par exemple, si un paramètre générique prend la place d'un nœud il apparaîtra dans le modèle en lieu et place de ce nœud.

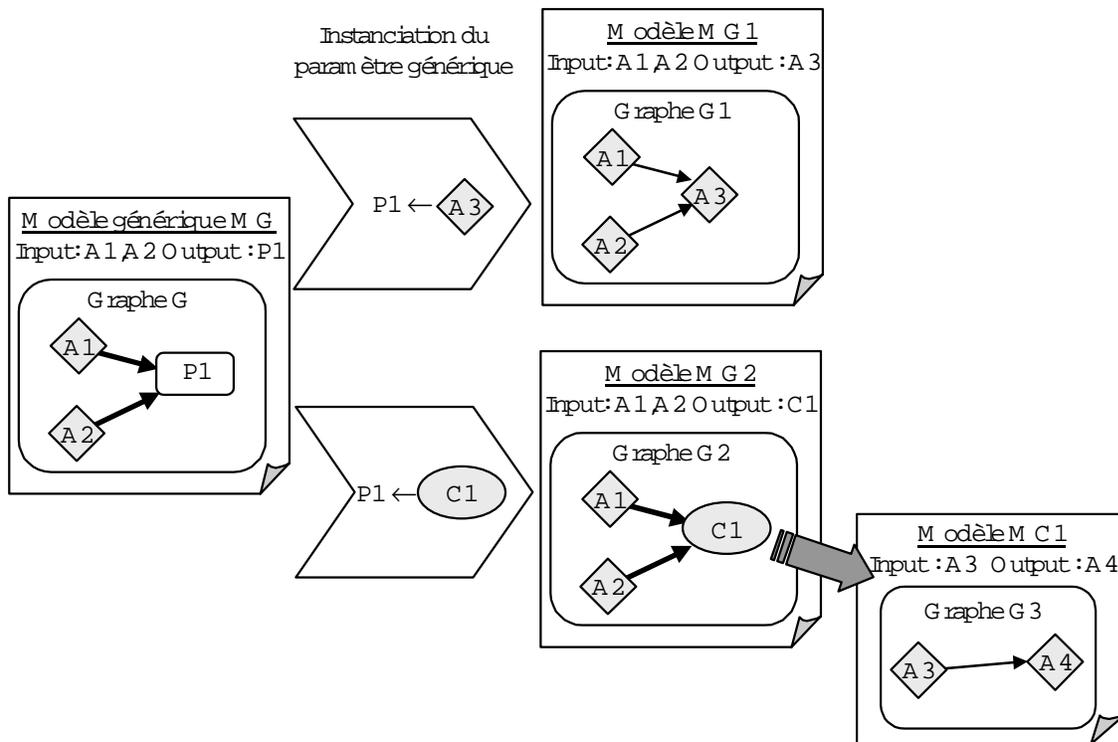


Figure 21 : Instantiation des paramètres d'un modèle générique.

Pour que le modèle puisse être utilisé en tant que modèle de composant, tous les paramètres génériques doivent être remplacés par des valeurs correctement spécifiées. Par exemple (Figure 21), un « nœud générique » doit être remplacé par un atome ou un composant correctement défini. La transformation d'un modèle générique en un modèle de composant (i.e. la prise de valeur des paramètres) est appelée *instanciation* du modèle générique. Le modèle de composant obtenu peut alors être utilisé normalement.

Pour être plus précis, à chaque fois qu'un des paramètres d'un modèle générique est remplacé par une valeur correcte, un nouveau modèle est bâti. L'opération de prise d'une valeur V_i d'un paramètre P_i peut être traduite par :

$$(G, I \{P_1, \dots, P_n\}) \xrightarrow{P_i=V_i} (G, I \{P_1, \dots, P_n\} - \{P_i\})$$

Le nouveau modèle est considéré comme *moins* générique que le précédent. S'il reste des paramètres génériques, le nouveau modèle est encore un modèle générique. Par contre, s'il ne possède plus de paramètres génériques, il devient un modèle de composant à part entière. En définitive, la prise de valeur de tous les paramètres crée un arbre de modèles génériques, chaque valeur différente créant une branche différente. La racine de cet arbre est le modèle possédant le plus de paramètres. Les feuilles sont tous les modèles de composants que l'on peut obtenir grâce aux différentes valeurs que peuvent prendre les paramètres génériques.

2.4. Graphe global

En utilisant des modèles de composant génériques, on peut désormais définir des modèles très complexes avec un moindre coût de spécification. Tant que la définition est valide, c'est-à-dire tant qu'elle ne contient pas de définitions récursives, on doit être capable de calculer les valeurs des atomes du modèle en fonction des connexions causales dont ils sont les effets. Cependant, comme on l'a vu, seules les influences nous permettent directement d'obtenir la valeur à utiliser. Les autres connexions (i.e. celles impliquant des composants) demandent un travail supplémentaire. Pour nous aider à calculer les valeurs des atomes nous définissons la notion de graphe global.

Définition 18 : Graphe global

Le graphe global d'une définition valide de modèle ne contient que des atomes et des influences. Il est obtenu à partir du graphe du modèle dans lequel on aura récursivement remplacé tous les composants par les graphes de leur propres modèles. On aura, de plus, redirigé les connexions impliquant ces composants vers les nœuds des graphes qui les ont remplacés.

La construction de ce graphe est assez simple. Après avoir remplacé tous les composants du graphe par les graphes de leurs propres modèles, il suffit de « rediriger » les connexions, suivant qu'elles utilisent un point ou un schéma de connexion :

Cas 1. Une connexion impliquant un composant et faisant intervenir un point de connexion se ramène, par définition, à une connexion impliquant le nœud désigné par ce point. Ce nœud, qui faisait partie du modèle du composant, a été ajouté au graphe global lors du remplacement du composant par son graphe.

Cas 2. Une connexion portant le nom d'un schéma de connexion se ramène à l'ensemble des connexions définissant le schéma. Ces connexions font intervenir des nœuds ont été, eux aussi, ajoutés au graphe global lors du remplacement du composant par son graphe.

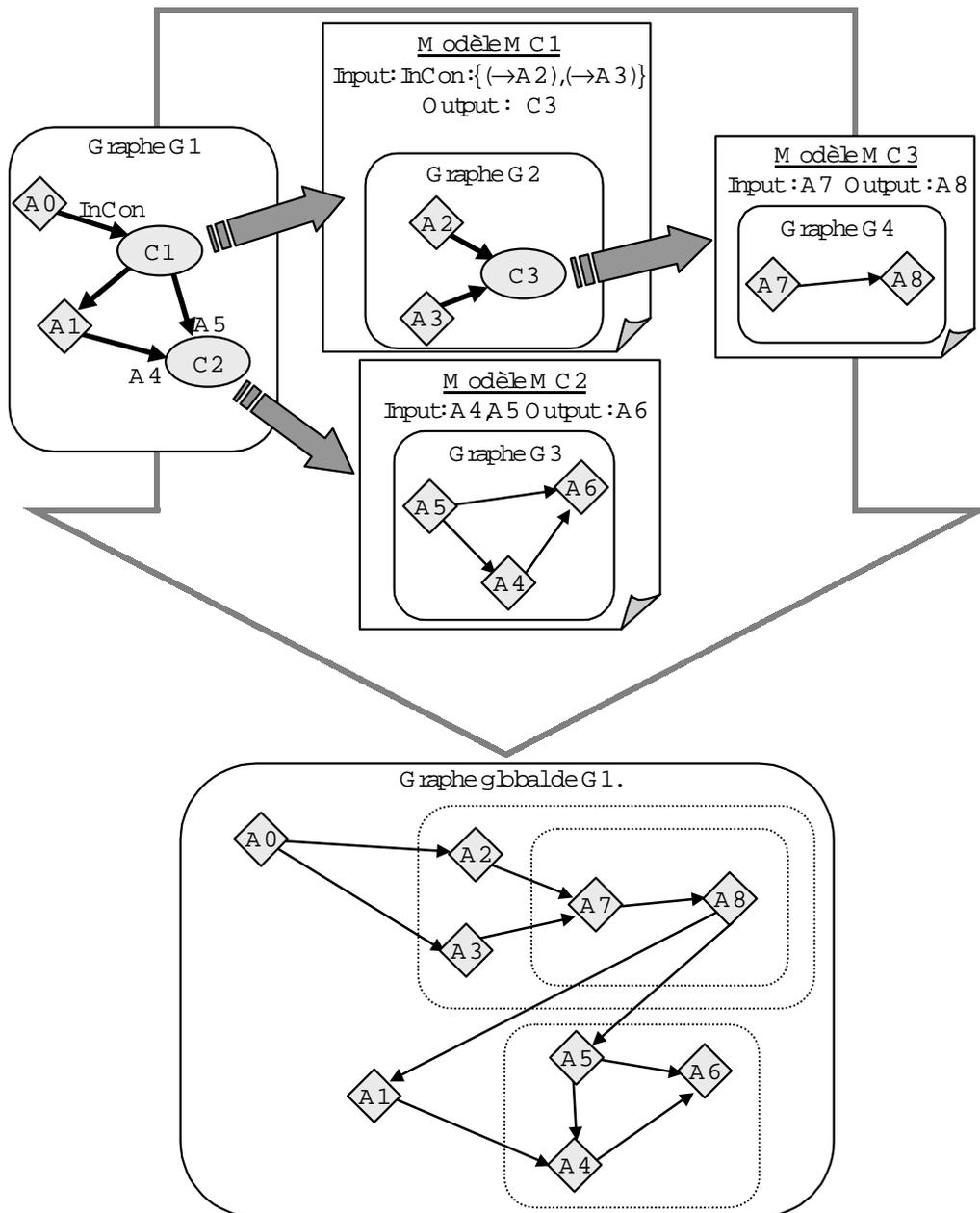


Figure 22: C onstruction du graphe global.

Si, parmi les nœuds que l'on vient d'ajouter, il y a des composants, il faut alors recommencer l'opération, et ce tant qu'il y aura des composants dans le graphe. La validité de la définition nous garantit l'arrêt de la procédure. En définitive, toute connexion est équivalente à une ou plusieurs influences. En reprenant l'exemple de la Figure 19, on peut illustrer l'obtention du graphe global associé à G_1 par la figure ci-dessus (Figure 22). La définition hiérarchique utilise 3 modèles de

composants. On remarquera que, dans le graphe G1, la connexion dirigée vers C1 utilise le schéma de connexion en entrée défini pour celui-ci car c'est seul moyen pour influencer ce composant. Au contraire, les deux connexions dirigées vers le composant C2 utilisent chacune une connexion différente en utilisant l'un des deux points d'entrée du composant. Dans le graphe global, la connexion dirigée vers C1 a été remplacée par deux influences ($A0 \rightarrow A2$ et $A0 \rightarrow A3$) et les connexions dirigées vers C2 ont été redirigées vers les atomes correspondants (A4 et A5).

2.5. Conclusion

Nous avons défini une architecture homogène pour une représentation hiérarchique de modèles de grande complexité. Notre architecture ne différencie formellement pas les niveaux structurels des niveaux fonctionnels, ce qui nous accorde une plus grande liberté dans l'organisation et l'interprétation de nos modèles. Tout graphe peut être considéré comme un modèle à part entière. Cela nous permettra, par exemple, de tester des portions de modèles avant de les assembler dans des structures de complexité croissante.

Nous savons maintenant que toute définition de graphe utilisant des composants peut se ramener à un graphe n'utilisant que des atomes et des influences. Nous nous intéresserons donc dans la suite de ce chapitre uniquement à ces deux types d'entités.

3. REPRESENTATION DE L'INFORMATION

Comme nous l'avons dit dans la première partie de ce chapitre, nous avons opté pour une représentation hybride de l'information, sous la forme d'un couple possédant une partie numérique, l'*Amplitude* et une partie symbolique, le *Type*. Ces deux parties sont complémentaires, c'est pourquoi tous les processeurs élémentaires (i.e. les atomes) seront associés à des valeurs constituées d'une amplitude et d'un type. Nous allons définir ici la forme précise de ces entités, ainsi que toutes les opérations et relations qui sont nécessaires pour les manipuler.

3.1. Amplitude

L'amplitude représente une donnée numérique imprécise. Elle est représentée par un intervalle qui délimite les bornes entre lesquelles la donnée numérique mal connue est comprise. On définit

généralement les valeurs des bornes de l'intervalle dans l'ensemble des réels (\mathbb{R}). On définit un intervalle fermé comme nous l'avons vu au Chapitre 3 (page 73, Travé-Massuyès *et al.* 1997, Sunaga 1958).

Pour pouvoir représenter des intervalles fermés de longueur infinie, on doit les définir en s'appuyant, non pas sur l'ensemble des réels, mais sur l'ensemble des réels étendus, noté \mathbb{R}^∞ . Cet ensemble est construit comme une augmentation de \mathbb{R} , auquel on ajoute deux éléments, les symboles $+\infty$ et $-\infty$. Par définition, $+\infty$ est tel que $\forall x \in \mathbb{R}, x < +\infty$. Respectivement, $-\infty$ est défini tel que $\forall x \in \mathbb{R}, x > -\infty$ (Tupper 1996, Struss 1990).

On augmente les relations et les opérations sur \mathbb{R} pour qu'elles puissent manipuler les nombres infinis. Ainsi, par exemple, pour l'addition et la soustraction, $\forall x \in \mathbb{R}, x + +\infty = +\infty, x - +\infty = -\infty, x + -\infty = -\infty$ et $x - -\infty = +\infty$, mais cependant le résultat de $+\infty + -\infty$ reste indéfini.

3.1.1. Les intervalles étendus

On définit les intervalles étendus sur \mathbb{R}^∞ de la même manière que les intervalles réels.

Définition 19 : Intervalle étendu

Un intervalle fermé étendu est un ensemble de réels étendus noté $[a, b]$, où a et $b \in \mathbb{R}^\infty$ et $a \leq b$, tel que $\forall x \in \mathbb{R}^\infty, x \in [a, b]$ si et seulement si $a \leq x \leq b$. On notera \mathcal{I}^∞ l'ensemble des intervalles étendus.

Ceci nous permet donc de représenter simplement des intervalles de longueur infinie. On peut, par exemple, représenter l'ensemble des réels étendus positifs ou nuls par l'intervalle $[0, +\infty]$ et l'ensemble \mathbb{R}^∞ par $[-\infty, +\infty]$.

Notation : Si I est un intervalle, alors on notera \bar{I} sa borne supérieure et \underline{I} sa borne inférieure.

On note *Longueur* la fonction de \mathcal{I}^∞ dans $\mathbb{R}^{\infty+}$, et dont l'image est la longueur de l'intervalle. $\forall A \in \mathcal{I}^\infty$ on définit $Longueur(A) = \bar{A} - \underline{A}$. On posera par définition que $Longueur([+\infty, +\infty]) = Longueur([-\infty, -\infty]) = 0$. Soit $\alpha \in \mathbb{R}^\infty$, on notera α tout intervalle de longueur nulle et égal à $[\alpha, \alpha]$.

On définit, sur les intervalles étendus, des relations inspirées des relations et opérations ensemblistes (inclusion, union, etc.) mais aussi inspirées du calcul sur les réels (supériorité, somme, multiplication, etc.).

3.1.2. Relations entre intervalles étendus

La première relation à définir sur \mathbb{I}^∞ est une relation d'équivalence, nous l'appellerons égalité stricte.

Définition 20 : Egalité stricte pour les intervalles étendus

$\forall A, B \in \mathbb{I}^\infty$, on dit que A est strictement égal à B, noté $A = B$, si et seulement si $\bar{A} = \bar{B}$ et $\underline{A} = \underline{B}$.

On peut interpréter cette relation comme la relation telle que, si l'on considère que les intervalles A et B représentent des valeurs réelles imprécises, alors $A = B$ est équivalent à dire que *quelle que soit* la véritable valeur de A, elle est une valeur possible de B et réciproquement.

Notation : $\forall A, B \in \mathbb{I}^\infty$, on notera $A \neq B$, A est différent de B, la proposition $\neg(A = B)$.

On peut définir l'égalité qualitative au sens où elle a été définie au chapitre précédent (page 69).

Définition 21 : Egalité qualitative

$\forall A, B \in \mathbb{I}^\infty$, on dit que A est qualitativement égal à B, noté $A \approx B$, si et seulement si $\exists x \in \mathbb{R}^\infty$ tel que $x \in A$ et $x \in B$.

On peut interpréter cette relation comme la relation telle que, si l'on considère que les intervalles A et B représentent des valeurs réelles imprécises, alors $A \approx B$ est équivalent à dire qu'il est *possible* que la véritable valeur de A soit la véritable valeur de B et réciproquement. Cette seconde relation est beaucoup moins restrictive que la relation d'égalité stricte.

L'égalité qualitative est une relation réflexive et symétrique, mais pas transitive. On remarque de plus les propriétés suivantes :

Propriétés 3 :

1. $\forall A, B \in \mathbb{I}^\infty$ si $A = B$ alors $A \approx B$.
2. $\forall A \in \mathbb{I}^\infty$ $A \approx [-\infty, +\infty]$.

On peut ensuite définir une relation d'ordre partiel inspirée de la relation d'inclusion ensembliste.

Définition 22 : Inclusion

Soit deux intervalles étendus A et B , on dit que A est inclus dans B , noté $A \subset B$, si et seulement si $\forall x \in A$ alors $x \in B$.

On remarque les propriétés suivantes,

Propriétés 4 :

1. $\forall A, B \in \mathbb{I}^\infty$ alors $A \subset B$ si et seulement si $\underline{A} \geq \underline{B}$ et $\bar{A} \leq \bar{B}$.
2. $\forall A \in \mathbb{I}^\infty$ alors $A \subset [-\infty, +\infty]$.
3. $\forall A, B \in \mathbb{I}^\infty$, $A = B$ si et seulement si $A \subset B$ et $B \subset A$.
4. $\forall A, B \in \mathbb{I}^\infty$, si $A \subset B$ alors $A \approx B$.

On peut définir une seconde relation d'ordre strict partiel qui est, elle, inspirée de la relation d'ordre des réels.

Définition 23 : Supériorité stricte

$\forall A, B \in \mathbb{I}^\infty$, on dit que A est strictement supérieur à B , noté $A \gg B$, si et seulement si $\forall x \in A$ et $\forall y \in B$ alors $x > y$.

On montre aisément que la supériorité stricte est une relation d'ordre strict partiel. On remarque de plus les propriétés suivantes :

Propriétés 5 :

1. $\forall A, B \in \mathbb{I}^\infty$, alors $A \gg B$ si et seulement si $\underline{A} > \bar{B}$.
2. $\forall A, B \in \mathbb{I}^\infty$, $A \approx B$ ou $A \gg B$ ou $B \gg A$.

Pour obtenir un ordre total sur \mathbb{I}^∞ on doit fournir une définition moins intuitive de la supériorité entre intervalles :

Définition 24 : supériorité totale

$\forall A, B \in \mathbb{I}^\infty$, on dit que A est supérieur à B , noté $A > B$, si et seulement si $\bar{A} > \bar{B}$, ou sinon $\bar{A} = \bar{B}$ et $\underline{A} < \underline{B}$.

Pour que cette relation soit complète, il faut ajouter un cas particulier lorsque les bornes supérieures sont égales. Nous avons choisi de rester cohérent avec la relation d'inclusion (i.e. A est

plus grand à B lorsqu'il le contient). Cette relation illustre la difficulté d'obtenir une relation d'ordre totale sur \mathbb{I}^∞ . On montre sans aucune difficulté que cette relation est une relation d'ordre stricte et totale sur \mathbb{I}^∞ ainsi que les propriétés suivantes :

Propriétés 6 :

1. $\forall A, B \in \mathbb{I}^\infty$, si $A \subset B$ alors $B > A$ ou $A = B$.
2. $\forall A, B \in \mathbb{I}^\infty$, si $A \gg B$ alors $A > B$.
3. $\forall A, B \in \mathbb{I}^\infty$, si $A > B$ alors $A \approx B$ ou $A \gg B$.

3.1.3. Opérations sur les intervalles étendus

L'idée générale pour la définition d'opérateurs sur les intervalles est de considérer le résultat de l'opération comme le plus petit intervalle (au sens de l'inclusion) contenant l'ensemble de tous les résultats de l'opération appliquée à chacun des éléments des intervalles initiaux. Cette idée prend tout son sens si elle est appliquée aux opérateurs ensemblistes, l'intersection (\cap) et l'union (\cup).

Définition 25 : Union d'intervalles

Soit deux intervalles étendus A et B. On appelle union de A et B, l'intervalle noté $A \cup B$ tel que $A \cup B = [\min(\underline{A}, \underline{B}), \max(\bar{A}, \bar{B})]$.

On montre aisément que $A \cup B$ contient tous les éléments qui étaient contenus dans A et dans B. Mais on peut aussi remarquer que, ainsi que nous le disions ci-dessus, $A \cup B$ peut contenir des réels étendus qui n'étaient contenus ni dans A, ni dans B. C'est, par exemple, le cas de $[1, 2] \cup [5, 6] = [1, 6]$. C'est cette particularité qui fait, cependant, l'utilité de cet opérateur (Sunaga 1958). Il faut tout d'abord démontrer la propriété suivante :

Propriété 7 :

$$\forall A \in \mathbb{I}^\infty, \text{ alors } A = \bar{A} \cup \underline{A} = \underline{A} \cup \bar{A}.$$

Cette propriété nous permet de représenter tout intervalle $[a, b]$ par l'union de ses bornes $a \cup b$.

Définition 26 : intersection d'intervalles

$\forall A, B \in \mathbb{I}^\infty$, tels que $A \approx B$, on appelle intersection de A et B, l'intervalle noté $A \cap B$ et tel que $\forall x \in A \cap B, x \in A$ et $x \in B$.

L'intersection n'existe que si les intervalles possèdent au moins une valeur commune, c'est à dire s'il sont qualitativement égaux. Toutes les propriétés de l'union et de l'intersection ensemblistes sont vérifiées pour l'union et l'intersection d'intervalles. $[-\infty, +\infty]$ est l'élément absorbant de l'union et l'élément neutre de l'intersection.

Les opérations arithmétiques, enfin, sont définies sur la base des opérations qui leur correspondent sur \mathbb{R}^∞ . On définit d'une manière générale les quatre opérations de base comme suit (Tupper 1996, Sunaga 1958).

Définition 27 : Opérations arithmétiques

$\forall A, B \in \mathbb{I}^\infty$ et \diamond un opérateur choisi parmi $\{+, \times, -, /\}$, le résultat de l'opération $A \diamond B$ est défini comme le plus petit intervalle étendu au sens de l'inclusion, contenant l'ensemble $\{x \in \mathbb{R}^\infty / \exists a \in A \text{ et } \exists b \in B, x = a \diamond b\}$.

On construira l'intervalle résultat à partir de l'union des opérations menées sur les bornes des opérands. On peut montrer que le résultat de l'opération sur les intervalles produit alors bien le plus petit intervalle contenant l'ensemble des résultats de l'opération sur les éléments des intervalles. De plus la présence de $[-\infty, +\infty]$, dans \mathbb{I}^∞ , va nous permettre de fermer les opérations sur \mathbb{I}^∞ , en associant l'intervalle $[-\infty, +\infty]$ aux valeurs qui étaient indéterminées dans \mathbb{R}^∞ . De cette manière on définit l'opposé d'un intervalle par :

$$-A = (\underline{-A}) \cup (\overline{-A}) = [\underline{-A}, \overline{-A}].$$

En posant $-\infty + +\infty = [-\infty, +\infty]$, on peut définir $A + B$ et $A - B$ comme :

$$A + B = (\underline{A} + \underline{B}) \cup (\overline{A} + \overline{B}) = [\underline{A} + \underline{B}, \overline{A} + \overline{B}],$$

$$A - B = A + (-B).$$

De même on peut définir l'inverse d'un intervalle, $1/A$, par :

$$1/A = \begin{cases} (1/\underline{A} \cup 1/\overline{A}) = [1/\underline{A}, 1/\overline{A}] & \text{si } 0 \notin A \\ [-\infty, +\infty] & \text{sinon} \end{cases}$$

Et en posant $0 \times +\infty = 0 \times -\infty = [-\infty, +\infty]$ on peut définir $A \times B$ et A/B comme :

$$A \times B = (\underline{A} \times \underline{B}) \cup (\overline{A} \times \overline{B}) \cup (\overline{A} \times \underline{B}) \cup (\underline{A} \times \overline{B}),$$

$$A/B = A \times (1/B).$$

On peut de cette manière définir d'autres fonctions sur \mathbb{I}^∞ , comme par exemple les fonctions trigonométriques ou bien les fonctions exponentielle et logarithme.

3.1.4. Relations multivaluées

Contrairement à un ensemble flou, un intervalle ne représente pas l'incertitude associée à l'information qu'il représente. Cela allège sa représentation et facilite la spécification d'une information mal connue. Cependant, l'interprétation de la validité d'une relation manipulant un intervalle peut ne pas être certaine. On pourrait, par exemple, dans le cas des relations d'égalité définies ci-dessus, admettre que le degré de certitude d'égalité de deux valeurs imprécises A et B est supérieur lorsque $A = B$ que lorsque $A \approx B$.

C'est pour prendre en compte l'incertitude générée par des données imprécises représentées par des intervalles, que nous proposons des relations multivaluées. Ces relations sont interprétées dans une logique multivaluée comme les logiques de Lukasiewicz que nous avons présentées au chapitre précédent (voir page 61). Nous allons proposer une relation d'égalité et une relation d'ordre. Dans un premier temps, nous allons définir ces relations dans L_3 , ce qui nous permettra d'en avoir une représentation simple, avant de passer à L_{N1} .

Quelle que soit la relation \bullet et les intervalles étendus A et B, on notera $V_2(A \bullet B)$, $V_3(A \bullet B)$ et $V_{N1}(A \bullet B)$ les valeurs de vérité de la relation $A \bullet B$, respectivement dans la logique booléenne, dans L_3 et dans L_{N1} . Ainsi, par exemple, dans L_2 , $V_2(A \bullet B) = 1$ si et seulement si $A \bullet B$ est vérifiée, $V_2(A \bullet B) = 0$ si la relation n'est pas vérifiée. Par extension nous définissons, $\forall A, B \in \mathbb{I}^\infty$ et $\forall \bullet \in \{=, \approx, \subset, \gg, >\}$, $V_{N1}(A \bullet B) = V_3(A \bullet B) = V_2(A \bullet B)$.

Définition 28 : Egalité dans L_3

$\forall A, B \in \mathbb{I}^\infty$, on notera $A =_3 B$, la relation égalité dans L_3 , telle que :

$$V_3(A =_3 B) = 1 \text{ si et seulement si } A = B,$$

$$V_3(A =_3 B) = \frac{1}{2} \text{ si et seulement si } A \neq B \text{ et } A \approx B,$$

$$V_3(A =_3 B) = 0 \text{ sinon.}$$

On peut montrer, par exemple sur des tables de vérité, que cette relation est réflexive ($A =_3 A$), symétrique ($(A =_3 B) \rightarrow (B =_3 A)$), transitive ($((A =_3 B) \wedge (B =_3 C)) \rightarrow (A =_3 C)$) dans L_3 . On peut aussi montrer que :

Propriété 8 :

$$\forall A, B \in \mathbb{I}^\infty \text{ alors } V_3(A =_3 B) = V_3(A = B) \vee (V_3(A \approx B) \wedge 1/2).$$

Sur la base de cette définition on peut définir la relation d'ordre suivante :

Définition 29 : Supériorité dans L_3

$\forall A, B \in \mathbb{I}^\infty$, on notera $A >_3 B$, la relation supérieur dans L_3 , telle que :

$$V_3(A >_3 B) = 1 \text{ si et seulement si } A \gg B,$$

$$V_3(A >_3 B) = 1/2 \text{ si et seulement si } A > B \text{ et } A \approx B,$$

$$V_3(A >_3 B) = 0 \text{ sinon.}$$

On montre que $>_3$ est irreflexive ($(A >_3 B) \rightarrow \neg(A =_3 B)$), antisymétrique ($((A >_3 B) \wedge (B >_3 A)) \rightarrow (A =_3 B)$) et transitive ($((A >_3 B) \wedge (B >_3 C)) \rightarrow (A >_3 C)$) dans L_3 . On montre de plus que $>_3$ est totale ($\neg(A =_3 B) \rightarrow ((A >_3 B) \vee (B >_3 A))$) dans L_3 .

Ce système de relations permet de gérer un minimum d'incertitude engendrée par les intervalles. Il est cependant possible d'obtenir une graduation de l'inégalité d'intervalles qualitativement égaux. Par exemple, on conviendra que $[1, 2]$ est plus intuitivement plus « égal » à $[1, 3]$ qu'à $[2, 5]$ pourtant $[1, 2] \approx [1, 3] \approx [2, 5]$. On peut fonder la mesure de la ressemblance des intervalles sur la longueur de leur intersection (i.e. de leur partie commune). On a $[1, 2] \cap [1, 3] = [1, 2]$ et $[1, 2] \cap [2, 5] = 2$, or $Longueur([1, 2]) > Longueur(2)$. Toutefois, cette mesure soulève deux problèmes aux cas limites, lorsque la longueur de la partie commune est infinie et quand elle est nulle. En prenant en compte les cas aux limites, nous allons proposer une relation d'égalité dans L_{N1} . Cette relation ne s'appuie que sur la relation d'égalité stricte.

Définition 30 : Egalité dans L_{N1}

$\forall A, B \in \mathbb{I}^\infty$, on notera $A =_{N1} B$, l'égalité dans L_{N1} , telle que :

$$V_{N1}(A =_{N1} B) = 1 \text{ si et seulement si } A = B,$$

$$V_{N1}(A =_{N1} B) = \frac{Longueur(A \cap B)}{Longueur(A \cup B)} \text{ si } A \neq B, A \approx B, Longueur(A \cap B) < +\infty.$$

$$V_{N1}(A =_{N1} B) = 1/2 \text{ si } A \neq B, A \approx B \text{ et } Longueur(A \cap B) = +\infty.$$

$$V_{N1}(A =_{N1} B) = 0 \text{ sinon.}$$

Le troisième cas élimine l'indétermination du quotient ∞/∞ dont la valeur est indéterminée. On pose par définition que, dans ce cas là, la valeur de vérité de la relation sera égale à $1/2$, ce qui représente une valeur de vérité indéterminée, à mi-chemin entre le Vrai (1) et le Faux (0). On remarque que, si la longueur de l'intersection des deux intervalles est nulle (e.g. $[2, 2] \cap [1, 3]$), alors la valeur de vérité de $=_{N1}$ vaut 0 (Faux) alors que la relation d'égalité qualitative est vérifiée (e.g. $[2, 2] \approx [1, 3]$ car $2 \in [1, 3]$). Cette relation est donc à considérer avec prudence, car elle n'est pas entièrement compatible avec l'égalité qualitative.

Sur la base de cette relation on peut alors définir une relation d'ordre strict total dans L_{N1} , que l'on notera $>_{N1}$ et qui est définie par :

Définition 31 : Ordre strict total dans L_{N1}

$$V_{N1}(A >_{N1} B) = V_{N1}(\neg(A =_{N1} B) \wedge (A > B)).$$

3.2. Type

Le type est la partie symbolique de l'information. Pour pouvoir réaliser simplement la réunion de plusieurs types, nous avons choisi de représenter le type par un ensemble flou de symboles. On peut, de cette manière, construire incrémentalement un type en associant des types élémentaires. De plus, pour représenter l'*atténuation* des éléments constituant un type (voir page 90), à chaque élément sera associée une valeur d'appartenance, comprise entre 0 et 1, et qui spécifiera l'atténuation de cet élément vis à vis des autres éléments du type.

Les symboles contenus dans un type sont dépendants de la catégorie d'information représentée, catégorie qui n'est pas modifiée par un changement d'atténuation. On peut construire les types en fonction des stimuli du modèle concerné et ils peuvent prendre des formes très diverses. Par exemple, pour une aire primaire auditive, les types sont constitués de symboles représentant des fréquences sonores. On notera \mathbb{S} , l'ensemble de tous les symboles d'éléments de types possibles. On définit formellement un type comme il suit.

Définition 32 : Type

On appelle type, toute application de \mathbb{S} vers $[0, 1]$. On désigne par \mathbb{T} l'espace des types. Le poids de l'atténuation d'un symbole $s \in \mathbb{S}$, dans un type $T \in \mathbb{T}$, est noté $T(s)$.

Un type associe une atténuation à tous les symboles de \mathcal{S} . L'atténuation $T(s) = 0$ indique que le symbole est totalement atténué ; on dit alors qu'il n'*appartient* pas au type T . $T(s) = 1$ revient à dire que le symbole s n'est pas atténué, son rang est le plus élevé.

Définition 33 : Types remarquables

On appelle *type composé* le type défini comme l'application associant 1 à un ensemble de symboles $\{s_1, \dots, s_n\}$ et 0 à tout autre symbole de \mathcal{S} . Si $n = 1$, on parlera de *type pur*.

On appelle *type vide* le type, noté \emptyset , tel que $\forall s \in \mathcal{S}, \emptyset(s) = 0$.

On appelle *type total* le type, noté Γ , tel que $\forall s \in \mathcal{S}, \Gamma(s) = 1$.

3.2.1. Relations sur les types

Nous allons tout d'abord définir des relations à valeurs dans l'algèbre de Boole, puis nous définirons, comme pour les intervalles, des versions de ces relations dans L_3 et L_{N1} . On définit les relations sur les types en se basant sur des comparaisons portant sur les poids d'atténuation des symboles. Ainsi, on peut définir :

Définition 34 : Egalité stricte de deux types

$\forall T_1, T_2 \in \mathcal{T}$, on dit que T_1 est strictement égal à T_2 , noté $T_1 = T_2$, si et seulement si $\forall s \in \mathcal{S}, T_1(s) = T_2(s)$.

Définition 35 : Supériorité de deux types

$\forall T_1, T_2 \in \mathcal{T}$, on dit que T_1 est supérieur à T_2 , noté $T_1 > T_2$ si et seulement si $\forall s \in \mathcal{S}, T_1(s) > T_2(s)$.

La relation d'égalité est une relation d'équivalence sur \mathcal{T} . La relation de supériorité est une relation d'ordre strict partielle sur \mathcal{T} . L'égalité stricte considère deux types égaux si toutes les atténuations sont égales. Si l'on veut être moins strict on peut admettre que deux types sont « presque » égaux si la différence entre les atténuations est faible. Cette idée nous a conduit à définir une relation d'égalité large, noté \cong . Cette relation est une relation multivaluée, on définit sa valeur de vérité comme suit.

Définition 36 : Egalité large de deux types

$\forall T_1, T_2 \in \mathcal{T}$, on définit la relation d'égalité large, $T_1 \cong T_2$ telle que $V_{N1}(T_1 \cong T_2) = 1 - \max_{s \in \mathcal{S}} (|T_1(s) - T_2(s)|)$.

On remarquera les propriétés suivantes (en notant $\circ P$ la modalité « P est nécessaire ») :

Propriétés 9 :

1. L'égalité large est une relation d'équivalence.
2. $V_{N1}(\circ(T_1 \cong T_2)) = V_{N1}(T_1 = T_2)$, soit en d'autres termes : si $V_{N1}(T_1 \cong T_2) = 1$ alors $T_1 = T_2$.

3.2.2. Opérations sur les types

Les opérations sur les types sont des opérations ensemblistes qui se ramènent, elles aussi, à des opérations sur les atténuations des symboles de \mathbb{S} .

Définition 37 : Complémentaire d'un type

On appelle complémentaire d'un type T , le type noté $-T$ tel que $\forall s \in \mathbb{S}. -T(s) = 1 - T(s)$.

Définition 38 : Opérations de composition

$\forall T_1, T_2 \in \mathbb{T}$,

On appelle union de T_1 et T_2 , le type noté $T_1 \cup T_2$, tel que $\forall s \in \mathbb{S}, T_1 \cup T_2(s) = \max(T_1(s), T_2(s))$.

On appelle intersection de T_1 et T_2 , le type noté $T_1 \cap T_2$, tel que $\forall s \in \mathbb{S}, T_1 \cap T_2(s) = \min(T_1(s), T_2(s))$.

On appelle produit de T_1 et T_2 , le type noté $T_1 \otimes T_2$, tel que $\forall s \in \mathbb{S}, T_1 \otimes T_2(s) = T_1(s) \times T_2(s)$.

On appelle somme de T_1 et T_2 , le type noté $T_1 \oplus T_2$, tel que $\forall s \in \mathbb{S}, T_1 \oplus T_2(s) = \min(T_1(s) + T_2(s), 1)$.

Propriétés 10 :

1. \cup, \cap, \oplus et \otimes sont des opérations commutatives.
2. $\forall T_1 \in \mathbb{T}, \forall T_2 \in \mathbb{T}, T_1 \otimes T_2 < T_1 \cap T_2 < T_1 \cup T_2 < T_1 \oplus T_2$.
3. $\forall T \in \mathbb{T}, \emptyset \cup T = T, \emptyset \oplus T = T, \emptyset \cap T = \emptyset, \emptyset \otimes T = \emptyset$.
4. $\forall T \in \mathbb{T}, \Gamma \cup T = \Gamma, \Gamma \oplus T = \Gamma, \Gamma \cap T = T, \Gamma \otimes T = T$.

Pour en finir avec les opérations sur les types, on définit deux opérations scalaires et deux fonctions qui nous aideront à faire évoluer un type. Les deux opérations sont la multiplication par un scalaire, qui modifie les atténuations de tous les symboles, et le filtrage, qui permet de réduire un type en éliminant les atténuations les plus faibles. Les fonctions, appelées type maximal et poids supérieurs, nous permettent d'accéder à l'ensemble des symboles appartenant au type (i.e. de poids supérieur non nul).

$\forall T \in \mathcal{T}$ et $\forall \alpha \in [0, 1]$:

On appelle multiplication par un scalaire du type T , le type noté $\alpha.T$, tel que $\forall s \in \mathcal{S}$,

$$\alpha.T(s) = \alpha \times T(s).$$

On appelle filtre- α du type T , le type noté $\alpha|T$, tel que $\forall s \in \mathcal{S}$, $\alpha|T(s) = T(s)$ si $T(s) \geq \alpha$ ou

$$\alpha|T(s) = 0 \text{ sinon.}$$

On appelle type maximal d'un type T , le type noté \hat{T} , tel que $\forall s \in \mathcal{S}$, $\hat{T}(s) = 1$ si et seulement si $T(s) > 0$.

On appelle poids supérieur d'un type T , le scalaire, noté $\text{sup}(T)$, égal à $\text{sup}(T) = \max_{s \in \mathcal{S}} (T(s))$

3.3. Conclusion

Les atomes d'une hiérarchie de composants prennent leurs valeurs dans $\mathbb{I}^\infty \times \mathcal{T}$. De cette façon, l'espace de variation des variables est identique pour tous les atomes du modèle. Pour un atome donné, α , nous représenterons par α_A la partie amplitude de la valeur de l'atome, et par α_T la partie type de cette même valeur. En utilisant les relations et les opérations que nous venons de définir, nous allons pouvoir décrire les mécanismes de propagation de l'information au travers du modèle.

4. SIMULATION DYNAMIQUE DU SYSTEME

Nous allons maintenant décrire les mécanismes de simulation qui nous permettront de représenter les variations de l'état du système. Pour ce faire, nous reprendrons les mécanismes de base du simulateur causal Ca~En (Travé-Massuyès *et al.* 1997, Bousson & Travé-Massuyès 1994, Bousson 1993), que nous avons adaptés à nos contraintes lorsqu'ils n'y satisfaisaient pas.

On a vu que l'évaluation de l'état du système consiste à calculer la valeur de tous les atomes de la hiérarchie de composants. La simulation représente l'évolution de cet état au cours du temps. Nous devons donc calculer la valeur de chaque atome en tenant compte de l'écoulement du temps. Pour réaliser ce calcul, nous ne possédons que la connaissance représentée par les liens qui unissent les atomes (i.e. les influences). Le système étant considéré comme purement réactif vis à vis de l'environnement (hypothèse fondamentale I, page 82), on peut donc représenter son activité uniquement au travers de transformations qui sont appliquées à l'information entrante (i.e. les

stimuli). Ces transformations, contraintes par les influences, peuvent donc s'exprimer par des applications de fonctions (i.e. les processeurs) sur les valeurs de l'information. Les bases du calcul des valeurs (amplitude et type) ayant été établies précédemment, nous pouvons maintenant nous intéresser à leur propagation au travers du graphe d'influences.

4.1. Horloge

Nos modèles représentent une activité variant au cours du temps. On cherche à obtenir une reproduction de la propagation de l'information cérébrale durant un certain laps de temps, au cours duquel on calcule les variations de l'état du système. On cherche aussi à exprimer des relations représentant un délai dans la propagation entre deux zones cérébrales. En d'autres termes, nous devons manipuler des données variant au cours du temps et dont les relations ne sont pas instantanées. Ceci nous oblige à exprimer explicitement le temps dans nos modèles. Pour ce faire, nous définissons la notion d'*horloge du système* modélisé qui permet de représenter et surtout d'utiliser explicitement le temps dans les relations liant les atomes.

L'horloge permet de manipuler expressément le temps dans les modèles. On effectue les calculs de l'état du système à modéliser, suivant cette horloge, à des intervalles de temps régulier. Cette unité est appelée *pas d'échantillonnage*. On considère que les variations de l'état du système entre deux pas de calcul sont négligeables. Notre système est donc un système à temps discret. Bien que l'activité du cerveau soit une activité continue dans le temps, la granularité la plus fine des outils de mesure de cette activité est supérieure ou égale à la milliseconde (e.g. EEG et MEG, cf. chapitre 1, page 20). En conséquence, l'adoption d'une horloge discrète est une approximation justifiée. De plus, ce principe de simulation est très simple à mettre en œuvre sur un ordinateur conventionnel, fonctionnant lui-même en mode discret. Le choix d'une horloge virtuelle, propre au système modélisé, nous dégage de toute considération technique liée à l'implémentation de notre formalisme.

L'horloge est relative, car elle représente le temps qui s'est écoulé depuis le début de la simulation. En effet, on a admis que le système était purement réactif (hypothèse I) et on peut donc le considérer comme à l'état de repos avant le début de la simulation.

Notation : L'horloge du système est représentée par une variable que nous notons t . La valeur du pas d'échantillonnage est noté Δt . Les valeurs de t qui correspondent aux *taps d'horloge* (i.e. les

instants où l'on réalise les calculs) sont représentées par $t_i = i \times \Delta t$. Ainsi t_0 représente le temps du début de la simulation.

Nous devons nous assurer que l'on mesure au cours du temps toutes les valeurs nécessaires à la compréhension des variations de l'état du système. Pour ce faire, il faut que le pas d'échantillonnage soit suffisamment petit pour que l'on puisse considérer que les variations entre deux tops d'horloge sont négligeables. Pour que cette hypothèse soit réalisable il faut que Δt soit suffisamment petit face à la dynamique propre du système modélisé. Comme il l'a été précisé plus haut, la meilleure résolution temporelle des techniques de imagerie neuro-fonctionnelle étant de l'ordre de la milliseconde. Par la suite, toutes les durées seront exprimées en millisecondes. *On considérera, par défaut, que Δt vaut 1 milliseconde.*

La manipulation des valeurs au cours du temps peut poser certains problèmes. Nous en soulevons deux : le problème d'accessibilité des valeurs antérieures au début de la simulation et le problème de recalage du temps sur les tops d'horloge.

Le premier problème est surtout rencontré en début de simulation lorsque, par exemple, un délai de propagation induit l'utilisation de la valeur d'un atome à une date antérieure au début de la simulation. Or, on ne peut accéder à cette valeur, puisque le calcul débute à l'instant t_0 . Dans ce cas, on peut considérer que la valeur n'existe pas et le calcul ne peut être réalisé. Ou bien, on peut considérer que le système était dans un état stable au début de la simulation, et, de plus, qu'il était ainsi depuis suffisamment longtemps. Dans ce cas, la valeur recherchée est égale à celle du même atome au début de la simulation. Ici, le calcul peut être réalisé. Cependant, il faut être très prudent avec cette considération, car une mauvaise initialisation peut avoir des effets dangereux sur le comportement du système durant toute la simulation. Le choix de l'une ou l'autre de ces méthodes est contextuel dans notre implémentation.

Le second problème vient de ce que les diverses durées utilisées dans le modèle peuvent être exprimées par des valeurs ne correspondant pas à un top d'horloge. Il faudra donc remplacer ces durées par des multiples du pas d'échantillonnage. Par exemple pour $\Delta t = 1$ milliseconde, la valeur d'un atome à 3.5 millisecondes est inconnue, on connaît la valeur de l'atome à 3 millisecondes et à 4 millisecondes, mais pas entre. Cependant, si l'on considère que Δt est compatible avec la dynamique du système, on peut convenablement approcher la valeur souhaitée par la valeur calculée à l'instant le plus proche.

4.2. Fonctions de mise à jour

Les influences décrivent des causalités liant un seul atome cause à un seul atome effet. Elles traduisent les contraintes exercées sur les transformations de l'information propagée dans le graphe. Pour pouvoir calculer la valeur de l'effet en fonction de la cause, on associe à l'influence une fonction, appelée *fonction de mise à jour*. Pour pouvoir représenter des transformations complexes, nous nous sommes inspirés de la définition des fonctions de transfert associées aux influences causales du simulateur Ca~En. Nous avons adapté et étendu cette définition pour répondre aux contraintes d'une modélisation « physiologiquement plausible ».

Définition 40 : Fonction de mise à jour

La fonction de mise à jour d'une influence ($\alpha \rightarrow \beta$) est une fonction f possédant un ensemble de sept paramètres et telle que, à tout instant $t_i = i \times \Delta t$, on obtienne la valeur de l'atome β . Cette fonction possède deux composantes. La première permet d'obtenir l'amplitude de la valeur de β , on la note f_A . La seconde permet d'obtenir le type de la valeur de β , on la note f_T . Les paramètres de la fonction de mise à jour sont :

- Une condition d'activation, notée C : cette condition est une expression logique représentant une relation faisant intervenir les valeurs des atomes du graphe. La valeur de vérité de cette expression est interprétée dans la logique multivaluée L_{N1} . Cette expression représente l'état de l'influence ; état activé lorsque la condition est vérifiée (Vrai), inactivé lorsqu'elle ne l'est pas (Faux) et par extension, « progressivement activé » lorsque la valeur de vérité varie entre Vrai et Faux.
- Deux gains dynamiques, notés G_A et G_T : ces gains sont des fonctions de $I^\infty \times T$ vers I^∞ et T . On utilise deux gains différents : le *gain d'amplitude*, pour la composante amplitude, et le *gain de type*, pour la composante type de la fonction de mise à jour.
- Un délai, noté T_D : il est exprimé par un intervalle de longueur finie. Il représente le décalage temporel plus ou moins précis de l'influence de α sur β .
- Un temps de réponse, noté T_R : il représente le temps que β met à se stabiliser à la suite des variations de α . Il est lui aussi exprimé par un intervalle de longueur finie.

L'expression de mise à jour de la valeur de β , notée $(\beta_A(t_i), \beta_T(t_i))$, est le couple d'applications suivant :

$$\begin{cases} \beta_A(t_i) = f_A[\alpha, \beta, C, G_A, T_D, T_R](t_i) \\ \beta_T(t_i) = f_T[\alpha, \beta, C, G_T, T_D, T_R](t_i) \end{cases}$$

Les expressions des deux composantes sont, respectivement :

$$(18) \quad \underline{f}_A[\alpha, \beta, C, G_A, T_D, T_R](t_1) = (a \times \beta_A(t_1 - \Delta t)) + (V_{N1}(C) \times (1 - a) \times G_A(\alpha(t_1 - t_d)))$$

$$(19) \quad \underline{f}_T[\alpha, \beta, C, G_T, T_D, T_R](t_1) = (a \cdot \beta_T(t_1 - \Delta t)) \oplus ((V_{N1}(C) \times (1 - a)) \cdot G_T(\alpha(t_1 - t_d)))$$

où : les opérateurs (\times et $+$ pour les intervalles, \cdot et \oplus pour les types) sont ceux définis dans la partie précédente de ce chapitre (pages 107 et 113),

Δt est le pas d'échantillonnage du système,

$V_{N1}(C)$ est la valeur de vérité de la condition C dans la logique multivaluée L_{N1} .

t_d est le délai T_D recalé sur les tops de l'horloge discrète et

a est égal à $e^{-3\Delta t/T_R}$. Bien que nous soyons dans un contexte différent, nous avons choisi de garder pour a la valeur déterminée dans Ca~En (Bousson et Travé-Massuyès 1994), obtenue par la discrétisation d'une équation différentielle.

Les caractéristiques de ce genre d'influence sont dépendantes des paramètres qui la définissent, soit le délai T_D , la condition d'activation C, les gains dynamiques G_A et G_T et le temps de réponse T_R . Nous allons maintenant commenter chacun d'entre eux dans le détail.

4.2.1. Condition d'activation

Les conditions d'activation sont des expressions logiques qui manipulent les valeurs des atomes du graphe. Ces expressions utilisent les relations sur les amplitudes et les types que nous avons définis dans la partie précédente de ce chapitre (pages 105 et 112). Ces conditions permettent de représenter des modifications du fonctionnement du système lorsqu'un *événement* ce produit (e.g. en activant une influence lorsque l'amplitude d'un atome dépasse une certaine valeur). De tels mécanismes sont courants dans les processus de traitement de l'information cérébrale, le plus exemplaire étant celui de la décharge d'un neurone.

Les valeurs de vérité de ces conditions sont exprimées dans la logique multivaluée L_{N1} , ce qui nous permet de prendre en compte l'incertitude générée par la manipulation de valeurs imprécises. Nous dégageons deux états remarquables de l'influence, l'état actif et l'état inactif. On dit d'une influence qu'elle est active lorsque sa condition d'activation est Vrai (i.e. une valeur de vérité égale à 1 dans L_{N1}). On dit qu'une influence est inactive lorsque sa condition est Faux (i.e. sa valeur de vérité est 0 dans L_{N1}). Entre ces deux états, nous avons défini un continuum en faisant directement intervenir la valeur de vérité de la condition dans les équations (18) et (19).

Ainsi l'état actif d'une influence ramène les équations aux équations suivantes :

$$(20) \quad f_A[\alpha, \beta, \text{Vrai}, G_A, T_D, T_R](t_1) = (a \times \beta_A(t_1 - \Delta t)) + ((1 - a) \times G_A(\alpha(t_1 - t_d)))$$

$$(21) \quad f_T[\alpha, \beta, \text{Vrai}, G_T, T_D, T_R](t_1) = (a \cdot \beta_T(t_1 - \Delta t)) \oplus ((1 - a) \cdot G_T(\alpha(t_1 - t_d)))$$

Alors que l'état inactif d'une influence ramène ces équations à :

$$(22) \quad f_A[\alpha, \beta, \text{Faux}, G_A, T_D, T_R](t_1) = a \times \beta_A(t_1 - \Delta t)$$

$$(23) \quad f_T[\alpha, \beta, \text{Faux}, G_T, T_D, T_R](t_1) = a \cdot \beta_T(t_1 - \Delta t)$$

Donc, lorsque l'influence est inactive, la valeur de β décroît progressivement jusqu'à devenir nulle. Comme les conditions d'activation manipulent les valeurs d'autres atomes que ceux directement impliqués dans l'influence (i.e. les atomes α et β), elles peuvent introduire des dépendances supplémentaires à celles explicitement exprimées dans le graphe par des influences. On appelle ces dépendances des *dépendances indirectes*. Ainsi, tout atome, χ , dont la valeur est utilisée dans une influence dirigée vers β , crée une dépendance indirecte de β par rapport à χ . Par opposition, une influence ($\alpha \rightarrow \beta$) définit une *dépendance directe* de β par rapport à α (Figure 23).

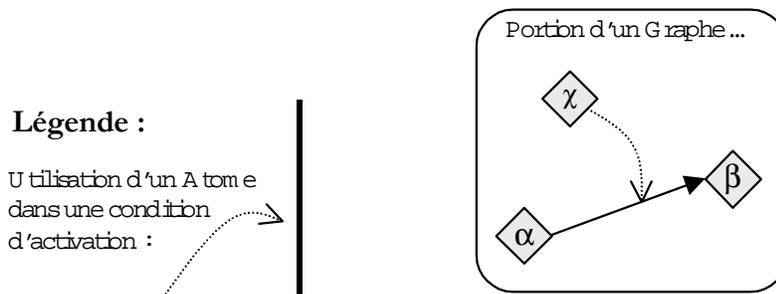


Figure 23 : Condition d'activation d'une influence.

L'existence de plusieurs états de fonctionnement d'une influence peut introduire des problèmes d'indécision lorsque celle-ci change d'état, surtout lorsque les valeurs déterminant l'état sont des valeurs imprécises. Le continuum d'états entre l'activation et l'inactivation permet de réduire les sauts de valeurs lors du changement de la valeur de vérité de la condition d'activation. Il faut, pour cela, utiliser l'une des relations multivaluées définies dans la partie précédente de ce chapitre. Ces relations, et notamment l'égalité $=_{N1}$ sur des intervalles, sont fonction de l'étendue de l'imprécision de leurs opérands (i.e. de la longueur des intervalles pour $=_{N1}$). Dans ce cas, la progression des variations de la valeur mise à jour est fonction de l'étendue de l'imprécision des valeurs manipulées

dans la condition d'activation. Plus ces valeurs sont imprécises et plus le passage entre un état actif et un état inactif est progressif.

4.2.2. Gains dynamiques

Les gains dynamiques sont utilisés pour représenter des mécanismes d'amplification ou de filtrage de l'information propagée. On trouve, encore une fois, des mécanismes équivalents au niveau neuronal. Il est donc fortement probable que les mêmes opérations soient à envisager sur une information plus abstraite.

Le mécanisme d'amplification se traduit par des gains linéaires lorsque l'amplification est constante au cours du temps (e.g. $G_A(X) = k \times X_A$, où k est une constante, ou bien $G_T(X) = k | X_T$ avec $k \in [0, 1]$). Lorsque l'amplification doit varier en fonction du temps, les gains peuvent prendre des formes plus complexes. Cette variation de l'amplification peut traduire toutes sortes de phénomènes. Elle peut, par exemple, être représentative de la sensibilité d'un processeur à un type particulier d'information. La valeur de l'information est davantage amplifiée lorsque le type souhaité se présente ou, au contraire, elle est atténuée lorsque son type ne correspond pas.

Le mécanisme de filtrage est plus directement lié à la partie type de l'information. Il correspond aussi à une réalité neuronale. On sait, par exemple, que les populations des couches corticales sont plus ou moins réactives à certaines configurations de décharges, réalisant ainsi un filtrage de l'information qui leur est transmise (Guigon *et al.* 1994). Ce genre de mécanisme peut être directement représenté par des intersections de types ou des opérations de filtres- α .

4.2.3. Le délai

Le délai de l'influence, T_D , est une expression directe du décalage temporel des variations de la valeur de β par rapport à celles de la valeur de α . Il est la manière la plus simple pour exprimer, par exemple, un délai lors d'une transmission distale entre deux zones cérébrales. Pour prendre en compte l'imprécision relative à ce genre de paramètres, nous avons choisi de représenter le délai sous la forme d'un intervalle de longueur finie. On peut, par exemple, de cette manière exprimer la variabilité inter-individuelle par des délais de transmission imprécis.

On remplace T_D par t_d dans l'expression de mise à jour des valeurs des atomes. L'intervalle t_d représente le recalage de T_D sur les tops d'horloge de telle manière que les bornes de cet intervalle soient des multiples de Δt . Ainsi, si $T_D = [a, b]$ alors $t_d = [a', b']$ tels que a' , respectivement b' , sont les

multiples de Δt les plus proches de a et respectivement de b . Par exemple, si $T_D = [1.2, 4.8]$ et si $\Delta t = 1$, le délai recalé t_d associé à T_D sera l'intervalle $[1, 5]$.

Si t_d à l'instant t_i représente un instant $t_i - t_d$ antérieur au début de la simulation, c'est-à-dire $t_i - t_d < 0$, alors on considère à cet instant que l'influence de α sur β est inexistante et on ne l'exprime pas. Durant toute la période où t_i est inférieur ou contenu dans t_d on dira que la variable β est en phase d'initialisation.

Dès que la phase d'initialisation est terminée, c'est à dire au premier top d'horloge t_i supérieur ou égal à la borne maximale de t_d , on applique la fonction de mise à jour à l'union de toutes les valeurs de α calculées entre $t - b'$ et $t - a'$. On considère de cette manière que l'imprécision du décalage se reporte sur l'imprécision des valeurs. Ainsi la fonction de mise à jour est appliquée au couple amplitude α_A et type α_T tel que :

$$\alpha_A(t - t_d) = \bigcup_{t=t_i - \overline{t_d}}^{t_i - \underline{t_d}} \alpha_A(t) \quad \text{et} \quad \alpha_T(t - t_d) = \bigcup_{t=t_i - \overline{t_d}}^{t_i - \underline{t_d}} \alpha_T(t)!$$

où, les opérateurs d'union sont les opérateurs respectivement définis sur les intervalles (page 107) et les types (page 113).

4.2.4. Temps de réponse.

Le temps de réponse d'une influence $\alpha \rightarrow \beta$ indique le temps nécessaire à β pour se stabiliser après avoir réagi à une variation de α . Il permet de paramétrer la stabilité d'un processeur ou, en d'autres termes, de jouer sur sa réactivité à l'information qui lui est transmise. Une influence avec un temps de réponse élevé engendrera des variations faibles lorsque la valeur de α variera, alors qu'un temps de réponse faible engendrera de fortes variations.

La réactivité des processeurs est un facteur important dans le traitement de l'information cérébrale. On peut, par exemple, considérer différentes échelles de réactivité entre les aires primaires et les aires associatives, les premières étant plus réactives que les secondes. Au niveau des processus automatiques, dirigés par les stimuli, le temps de réponse est un moyen explicite pour exprimer la latence de la réponse neuronale à un stimulus.

La notion de temps de réponse a été reprise de celle des influences du système Ca~En (Travé-Massuyès *et al.* 1997, Bousson & Travé-Massuyès 1994). Dans ce système, les influences représentent des discrétisations d'équations différentielles linéaires du premier ordre, de la forme :

$$(24) \quad y(t) + \tau \frac{dy(t)}{dt} = x(t)$$

La discrétisation de ce type d'équation se ramène à une équation récurrente (Bousson & Travé-Massuyès 1994) de la forme :

$$(25) \quad y(t) = a \times y(t-\Delta t) + b \times x(t), \quad \text{où } b = (1-a) \quad \text{et } a = e^{\frac{-3\Delta t}{\tau}}.$$

Nous avons donc repris l'expression du paramètre a dans notre propre fonction de mise à jour car nous fonctionnons aussi en mode discret. Comme nous ne considérons que des processus automatiques, pilotés par les stimuli, nous avons considéré que le transfert de l'information se faisait par des mécanismes semblables à ceux utilisés pour modéliser des systèmes physiques.

Cependant, compte tenu de la forme des valeurs de nos données (i.e. amplitudes et types), ainsi que de l'existence de paramètres dont les valeurs varient au cours du temps (i.e. condition d'activation et gains dynamiques), on ne peut pas considérer que notre fonction de mise à jour soit l'équation récurrente obtenue après discrétisation d'une équation différentielle linéaire.

4.2.5. Conclusion

Grâce aux diverses caractéristiques que nous venons de présenter, notre représentation de l'influence d'un atome sur un autre peut décrire des transformations très complexes, surtout si l'on combine les différents effets. La complexité des transformations est à la hauteur de l'abstraction de l'information à laquelle elles sont appliquées.

Ces transformations, aussi intégrées soient-elles, ne sont qu'une étape dans le traitement de l'information. En effet un processeur ne reçoit rarement qu'une seule influence à un même instant. Les processus de traitement plus complexes prennent en compte de multiples influences convergentes vers un même processeur.

4.3. Composition et propagation

Lorsque plusieurs influences convergent vers un même processeur, les transformations qu'elles appliquent à l'information qu'elles propagent doivent être combinées de manière à obtenir une valeur unique (Figure 24).

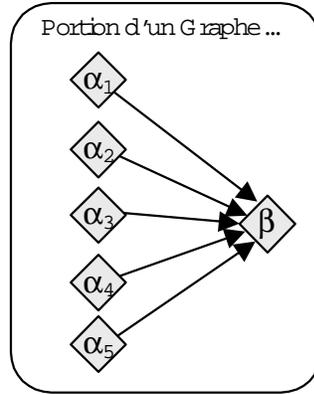


Figure 24 : De multiples influences convergentes.

Pour parvenir à obtenir la valeur de β , nous allons reprendre le concept de variation marginale décrit au chapitre précédent (page 77) et développé pour le simulateur Ca~En (Bousson 1993). Nous l'adaptions seulement à notre calcul.

Définition 41 : Valeur marginale

Si un atome, β , est influencé par plusieurs atomes α_1 à α_n , on définit les valeurs marginales de β comme les n valeurs de β^1 à β^n telles que β^i serait la valeur de β si $(\alpha_i \rightarrow \beta)$ était la seule influence dirigée vers β .

On définit ainsi l'indépendance relative des influences convergentes. Chaque valeur marginale, β^i , est obtenue indépendamment des autres en appliquant directement la fonction de mise à jour correspondant à l'influence, $(\alpha_i \rightarrow \beta)$, considérée. Pour obtenir la véritable valeur de β on doit combiner ses valeurs marginales. Cette opération est réalisée par la *loi de combinaison* de β . Une loi de combinaison est une fonction pouvant admettre un nombre variable de paramètres. Ici, elle est définie de $(\mathbb{I}^\infty \times \mathbb{T})^n$ vers $\mathbb{I}^\infty \times \mathbb{T}$, pour $n \in \mathbb{N}^+$. Son application aux valeurs marginales de β doit retourner, à chaque top d'horloge, la valeur de β . Soit, si ξ_β est la loi de combinaison de β , alors :

$$\beta(t) = \xi_\beta(\beta^1(t), \dots, \beta^n(t)).$$

Pour constituer une loi de combinaison valide, une fonction ξ doit satisfaire aux deux propriétés suivantes (Bousson 1993) appliquées aux valeurs marginales (Lafon *et al.* 1999, Lafon 1998) :

Propriété 11 : Consistance

$$\forall X \in \mathbb{I}^\infty \times \mathbb{T}, \xi(X) = X.$$

Propriété 12 : Multi-Symétrie

$$\forall X^1 \text{ à } X^n \in \mathbb{I}^\infty \times \mathbb{T}, \forall i \in \{1, \dots, n\}, \xi(X^1, \dots, X^i, X^{i+1}, \dots, X^n) = \xi(X^1, \dots, X^{i+1}, X^i, \dots, X^n).$$

La première propriété nous garantit que cette fonction est cohérente avec la définition des valeurs marginales. A savoir que, s'il n'y a qu'une seule valeur marginale à « combiner », l'atome prend cette valeur. La seconde propriété nous garantit que la combinaison n'est pas dépendante de l'ordre dans lequel on classe les valeurs marginales. Il n'y a en effet aucun ordre de priorité dans un graphe pour des influences convergentes vers un même atome.

Comme pour les fonctions de mise à jour, on peut séparer les deux composantes de cette fonction. On définit $\xi = (\xi_A, \xi_T)$, où ξ_A est la composante amplitude de ξ et ξ_T est la composante type de ξ . ξ_A est définie de $(\mathbb{I}^\infty \times \mathbb{T})^n$ vers \mathbb{I}^∞ , ξ_T est définie de $(\mathbb{I}^\infty \times \mathbb{T})^n$ vers \mathbb{T} . Nous citerons à titre d'exemple deux lois de combinaison que nous noterons Σ et Π . Ces lois sont définies comme :

$$(26) \quad \Sigma(X^1, \dots, X^n) = \begin{cases} \Sigma_A(X^1, \dots, X^n) = \sum_{i=1}^n X_A^i \\ \Sigma_T(X^1, \dots, X^n) = \bigcup_{i=1}^n X_T^i \end{cases}, \quad \Pi(X^1, \dots, X^n) = \begin{cases} \Pi_A(X^1, \dots, X^n) = \prod_{i=1}^n X_A^i \\ \Pi_T(X^1, \dots, X^n) = \bigcap_{i=1}^n X_T^i \end{cases}$$

4.4. Calculabilité d'un graphe global

Avant de conclure ce chapitre, nous allons nous intéresser à la calculabilité des valeurs des atomes dans le graphe global. Comme on vient de le voir, toutes les valeurs des atomes sont exclusivement calculées en fonction des valeurs des atomes les influençant. Or, s'il existe un cycle dans le graphe global, il est possible que, ainsi que l'illustre la figure ci-dessous (Figure 25), pour calculer la valeur d'un atome A1, qu'on ait besoin de connaître la valeur d'un atome A2, nécessitant à son tour que de connaître, récursivement, celle de A1. On ne peut alors calculer la valeur d'aucun des atomes du graphe. On dit que le graphe est incalculable.

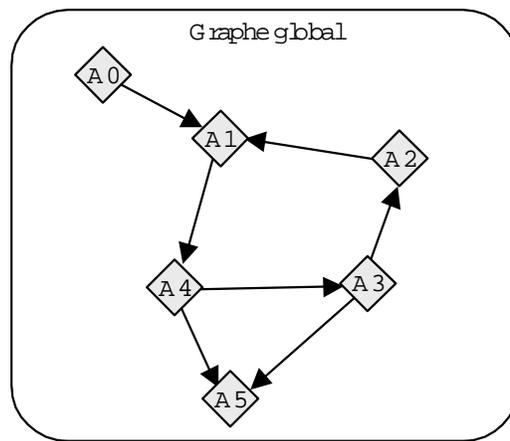


Figure 25 : Un graphe global incalculable.

On définit donc la calculabilité d'un modèle comme :

Définition 42 : Calculabilité d'un modèle

On dit qu'un modèle est calculable si et seulement si son graphe est valide et que dans son graphe global, la valeur de tout atome, à un instant donné, peut être calculée indépendamment d'elle-même.

Pour savoir si la valeur d'un atome est calculable ou pas, il faut prendre en compte les dépendances indirectes créées par les conditions d'activation. On peut cependant ignorer toutes les influences possédant un délai non nul, car la valeur nécessaire à la mise à jour a déjà été calculée à un instant précédent.

La valeur d'un atome, α , ne peut pas être calculée à un instant t , si et seulement si l'on est dans l'un des deux cas suivants :

1. Il existe, dans le graphe global, un cycle qui contient l'atome α et dont toutes les influences du cycle possèdent un délai nul (i.e. compte tenu du recalage du délai sur les tops d'horloge, ceci sera vrai dès que $T_D < \frac{\Delta t}{2}$).
2. Il existe une influence ($\kappa \rightarrow \lambda$), qui a une condition d'activation référençant la valeur de α à l'instant t , et il existe un chemin de λ à α , dont toutes les influences possèdent un délai nul.

5. CONCLUSION

Pour conclure ce chapitre, nous pouvons rappeler les principales caractéristiques du formalisme que nous venons de décrire.

Notre formalisme est fondé sur une architecture hiérarchique. Outre les qualités habituelles de ce genre de représentation (e.g. robustesse, réutilisabilité et modularité), elle nous permet une représentation descendante de la complexité du système modélisé.

L'utilisation de graphes nous dégage de toute contrainte topologique en ne considérant que les liens fonctionnels qui unissent les zones cérébrales. Nous ne sommes, de ce fait, pas soumis au problème de la localisation des processeurs dans nos modèles. Ceux-ci sont explicitement représentés.

La représentation de l'information est une représentation hybride, numérico-symbolique, qui nous permet de nous appuyer sur les mesures de l'imagerie fonctionnelle sans négliger pour autant la catégorie de l'information qui n'est pas mesurable. Le traitement de cette représentation est très intégré et nous permet de représenter explicitement les transformations qui sont appliquées à l'information.

Finalement, la simulation de la propagation de l'information nous permet d'obtenir le détail des variations de l'activité modélisée en concordance avec les meilleures définitions temporelles de l'imagerie.

LE SIMULATEUR BIOCAEN

Nous allons dans ce chapitre présenter le simulateur BIOCAEN qui implémente le formalisme que nous avons défini au chapitre précédent. Nous débuterons par quelques remarques sur l'organisation générale du logiciel ainsi que sur le langage de description des modèles qui lui est associé. Nous commenterons ensuite les algorithmes mis en œuvre pour réaliser les simulations. Pour finir, nous présenterons deux modèles différents qui offrent une première validation de notre approche. Ces modèles nous permettront d'illustrer les capacités du simulateur.

1. ALGORITHMES

1.1. Le simulateur Ca~En

La construction du simulateur BIOCAEN s'est appuyée en partie sur la deuxième version du simulateur qualitatif Ca~En (Travé-Massuyès *et al.* 1997, Bousson & Travé-Massuyès 1994) qui était fondée sur un raisonnement sur les intervalles. Ca~En est un logiciel de simulation qualitative fondé sur une représentation mixte des connaissances en faisant cohabiter une représentation causale et une représentation non causale (Bousson *et al.* 1994, Chapitre 3, partie 4.4, page 75). Seul le niveau de représentation causal nous intéresse. Nous ne connaissons pas d'équations différentielles globales (non causales) régissant la propagation de l'information dans le cerveau. Et même si de telles connaissances étaient exprimables, elles ne satisferaient pas à nos contraintes de plausibilité neurophysiologique (i.e. non-supervision de la propagation) et devraient être transformées en connaissances causales (distribuées).

Nous nous sommes donc intéressés uniquement à la partie causale de Ca-En, c'est à dire à l'algorithme de propagation des variations dans le graphe causal. De plus, nous ne raisonnons pas sur la base d'équations différentielles et nous avons donc abandonné les calculs de variations au profit de calculs de valeurs. L'attention toute particulière que nous avons accordée à cet algorithme

nous a permis de mettre en place un certain nombre de modifications et d'améliorations, notamment sur la phase d'initialisation de la simulation et sur la représentation des conditions d'activation. Tout ceci nous a conduit à une redéfinition complète des algorithmes de ce simulateur, de telle façon qu'en définitive on peut résumer l'influence de Ca~EN sur BIOCAEN à trois aspects qui sont :

1. Sur le plan représentation des modèles : une définition des graphes d'influences fondée sur la notion de variations marginales (ou d'influences marginales) et de lois de combinaisons.
2. Sur le plan de la simulation : une simulation réalisée par des calculs calés sur une échelle temporelle discrète et constante.
3. Sur le plan du logiciel lui même : Un mécanisme de compilation qui permet de réduire la complexité de l'algorithme de simulation.

Les deux premiers points ont été incorporés et adaptés à notre formalisme (voir chapitre 4, page 123). Le troisième point sera détaillé dans ce chapitre. Tout comme Ca~En, BIOCAEN est implémenté en langage LISP⁵.

1.2. Organisation générale du simulateur

La compilation des modèles permet de réaliser la simulation en deux temps, en effectuant un certain nombre d'optimisations et de vérifications sur la consistance des modèles. Ce mécanisme fait décroître la complexité de l'algorithme de simulation. L'organisation générale d'une simulation est la suivante (cf. Figure 26) :

Les modèles sont représentés dans un langage de description qui identifie tous les éléments des modèles (nœuds, connexions et paramètres). On différencie les modèles génériques, ceux-ci n'étant pas directement utilisables par le simulateur. Ils seront utilisés lorsqu'ils seront instanciés par le modèle à compiler.

Cette description est transmise au compilateur de BIOCAEN. Celui-ci construit le graphe global, à l'aide des modèles génériques nécessaires et en vérifiant la validité du graphe. C'est ce modèle « compilé » qui est transmis au simulateur.

⁵ Sous le dialecte Le-Lisp, © ILOG, version 15.26.

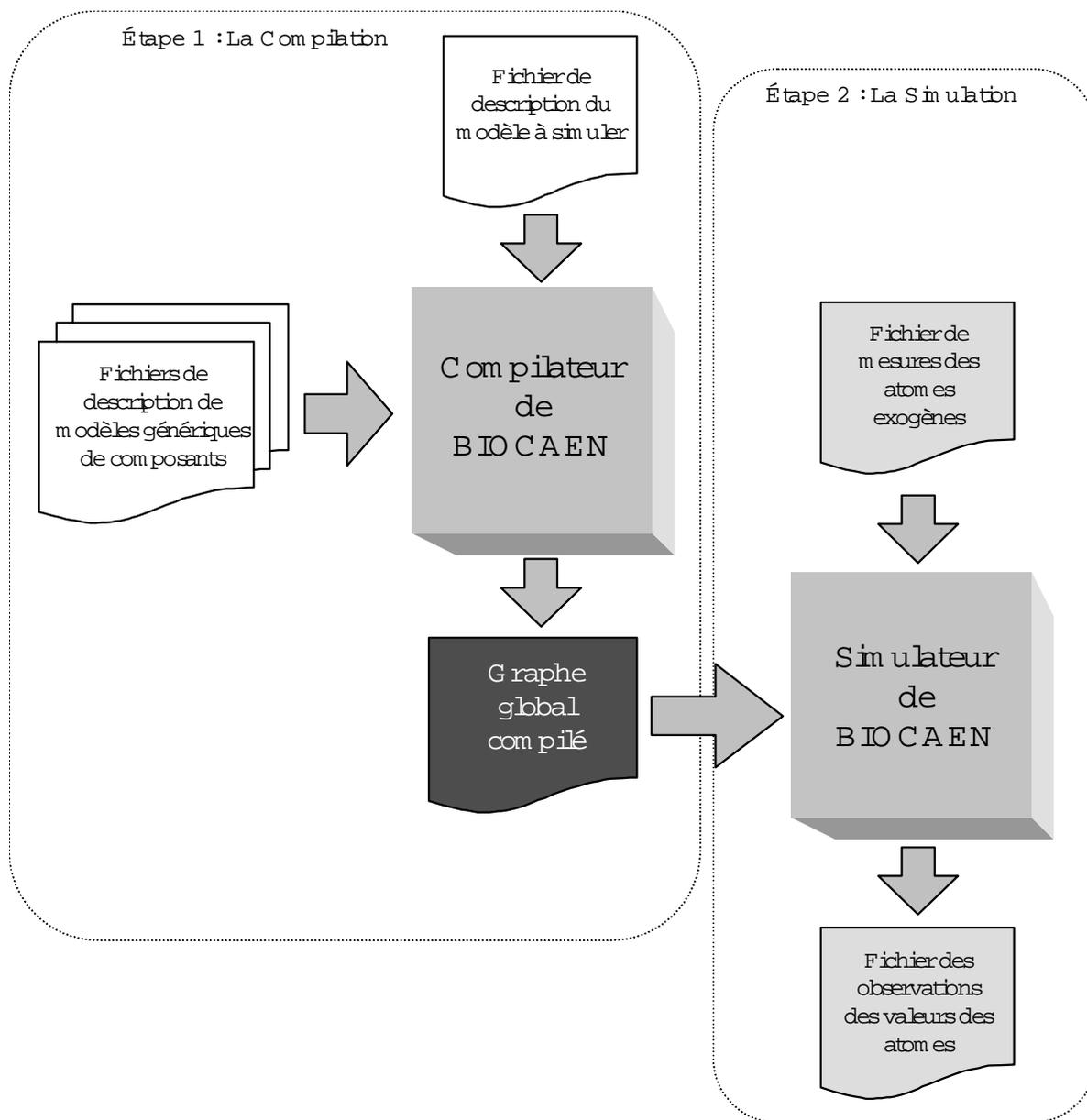


Figure 26 : Organisation d'une simulation sous BIOCAEN.

La simulation est réalisée par BIOCAEN à partir du modèle compilé et d'un fichier de valeurs fournissant, à chaque pas de calcul, la valeur de tous les atomes exogènes du modèle. Le résultat de la simulation est aussi un fichier de valeurs. Il fournit une observation des valeurs des atomes du modèle au cours du temps.

1.3. Description des modèles

La langage de description des modèles permet de représenter les éléments d'un graphe tel que nous les avons définis au chapitre précédent. Lorsqu'un élément est décrit, on précise la valeur des

paramètres nécessaires à sa définition. Sans entrer dans les détails, nous allons nous intéresser rapidement à la description des atomes, composants et influences qui constituent les éléments principaux de nos modèles. Nous allons aussi présenter le principe de la définition qualitative d'une valeur de type. De plus, BIOCAEN permet d'utiliser toutes les relations et opérateurs définis au chapitre précédent, Il propose d'autres opérateurs obtenus par composition des précédents.

Nous ne mettrons en avant que les concepts essentiels qui ont été adoptés par BIOCAEN pour décrire un modèle défini selon le formalisme du chapitre précédent. Pour une description exhaustive du langage de description de BIOCAEN, le lecteur pourra se reporter au manuel du simulateur (Lafon 1999). La description d'un modèle est constituée d'une succession de sections regroupant les définitions. Il existe ainsi une section appelée **Atoms** où sont décrits les atomes du graphe, une section **Components** et **Connexions**. Il existe d'autres sections comme la section **Interface**, définissant l'interface de connexion des modèles de composants. Reprenant la syntaxe générale des langages LISP, les sections sont délimitées par des parenthèses, chaque section pouvant contenir un nombre variable de définitions spécifiques à la section considérée.

1.3.1. Description des valeurs constantes

Les valeurs constantes apparaissant dans les modèles à décrire sont de trois sortes : les valeurs numériques (i.e. les intervalles), les valeurs symboliques (i.e. les valeurs de types) et les couples numérico-symboliques (i.e. intervalles et types). Un intervalle A est représenté par la liste $(\underline{A} \ \bar{A})$. La représentation des valeurs de types est plus complexe, compte tenu de la forme qu'ils peuvent prendre.

Le langage de description ne permet de décrire directement que des types composés (i.e. des types associant un poids de 1 à un ensemble de symboles et 0 aux autres, voir page 112). Les types plus complexes sont obtenus par composition. Si A , B et C sont des symboles, alors la fonction **Type** retourne le type composé, disons T , tel que $T(A) = T(B) = T(C) = 1$ et $T(s) = 0 \ \forall s \in S$ et $s \notin \{A, B, C\} : T = (\mathbf{Type} \ A \ B \ C)$.

BIOCAEN propose de plus un mécanisme de description qualitatif de types complexes. Ce mécanisme utilise les relations de $O(M)$ (voir chapitre 3, page 72, Mavrovouniotis & Stephanopoulos 1988, 1987). La description qualitative d'un type consiste en un ensemble de relations d'ordre entre des symboles. Ces relations définissent les valeurs des poids du type.

Si s_1 et s_2 sont deux symboles on notera $s_1 == s_2$ la proposition s_1 et s_2 ont la même atténuation. De même on notera $s_1 >\sim s_2$ la proposition l'atténuation de s_1 est légèrement supérieure à celle de s_2 , et ainsi de suite pour toute les relations de $O(M)$. On peut alors définir un type en indiquant simplement les relations d'ordre entre les symboles. Tous les symboles qui ne sont pas cités sont considérés comme ayant un poids nul. Par exemple, l'ensemble de relations $\{B >- C, A >\sim C\}$ peut définir un type T tel que $T(B) >- T(C)$ et $T(A) >\sim T(C)$.

L'algorithme de définition d'un type consiste à calculer les poids des symboles par ordre décroissant des relations de $O(M)$. Le poids du symbole le plus élevé d'après les relations d' $O(M)$ est fixé à 1, les autres poids sont calculés en conséquence. Par exemple, si A, B, C et D sont 4 symboles, l'application de la fonction `(OMType (B >- C) (A >\sim C))` permet d'obtenir le type T désiré. Selon cette définition, le poids du symbole B sera égal à 1 (i.e. le plus élevé), celui de C sera modérément plus petit que 1, et le poids de A sera légèrement supérieur au poids de C . Le poids du symbole D sera nul, car le symbole n'est pas cité dans la définition. Les poids effectif des symboles de la définition sont obtenus par *intersection* des intervalles compris entre 0 et 1 et correspondant aux relations de $O(M)$ (voir chapitre 3, page 72). Si la définition conduit à une intersection vide, elle est invalide. Par exemple, la définition $\{A >> B, B >> C, C >> A\}$ sera invalide car aucun réel, *a fortiori* compris entre 0 et 1, ne peut satisfaire ces relations. Si plusieurs types satisfont à une définition, le type retenu est, par défaut, celui qui possède les poids les plus élevés, soit le type maximal d'après la relation de supériorité des types (définie page 112).

1.3.2. Description d'un atome

Les descriptions des atomes sont regroupées dans une section appelée **Atoms**. Chaque définition doit spécifier le nom, qui identifie l'atome, sera utilisé pour le référencer, par exemple, dans la description des influences le concernant. La définition spécifie également un ensemble de paramètres, comme par exemple le domaine de variation de l'atome. Le paramètre le plus spécifique aux atomes est la loi de combinaison.

Chaque atome possède une loi qui lui est propre. Pour ce faire, on peut utiliser l'une des lois prédéfinies par BIOCAEN ou en définir une soi-même, en utilisant des opérateurs définis par le langage de description.

Les lois de combinaisons sont des lambda abstractions que le simulateur conserve dans une bibliothèque en vue de leur utilisation lors de la simulation. Les opérateurs de définition de lois de combinaison sont actuellement au nombre de deux. Ils nous garantissent que toute fonction définie

en les utilisant vérifiera les propriétés de consistance et de multi-symétrie (page 124). Par contre, ils ne permettent pas de définir toutes les lois de combinaison imaginables. On définit une loi de combinaison en utilisant la commande Define et en respectant la syntaxe suivante :

```
(Define Combination-law [Nom de la loi définie]
  [opération de combinaison des amplitudes]
  [opération de combinaison des types]
)
```

On définit indépendamment les deux parties, amplitude et type, de la loi en utilisant leurs opérateurs propres. L'opérateur de base de la définition d'une loi de combinaison est nommé **ForAll**. Il possède un seul paramètre, C, qui représente l'opération qui sera affectée aux valeurs marginales.

```
(ForAll C)
```

C est choisi dans une liste d'opérations commutatives limitée. On en distingue deux familles, les opérations portant sur des amplitudes et celles portant sur des types. Ces opérations sont celles que nous avons définies au chapitre précédent. Elles sont désignées par les symboles suivants :

Opérations sur des amplitudes :

@	unions d'intervalles,	&	intersection,	+	somme,
*	multiplication,	min	minimum	et	max maximum.

Opérations sur des types :

@T	unions de types,	&T	intersection,
+T	somme	et	*T produit.

On peut définir toute loi de combinaison ϑ du style :

$$(27) \quad \vartheta(X^1, \dots, X^n) = \begin{cases} \vartheta_A(X^1, \dots, X^n) = C_A^n X_A^i \\ \vartheta_T(X^1, \dots, X^n) = C_T^n X_T^i \end{cases}$$

où C_A , respectivement C_T , est une opération d'intervalles, respectivement de types, choisie parmi les opérations précédentes. Par exemple, voici les définitions de deux lois de combinaisons pré-définies dans BIOCAEN. Elles représentent les définitions des lois de combinaison Σ et Π mentionnées au chapitre précédent (voir page 124).

```

(Define Combination-law Sigma-Law
  (ForAll +)
  (ForAll +T)
)

(Define Combination-law Pi-Law
  (ForAll *)
  (ForAll *T)
)

```

L'algorithme, appliqué pour combiner les valeurs marginales d'un atome, et applicable à toute loi de combinaison créée par **ForAll**, est le suivant :

```

Soit X, ( la valeur de l'atome considéré,  $X_A$  et  $X_T$  sont des composantes )
      VM, ( l'ensemble des valeurs marginales de X )
       $C_A$ , ( l'opération à appliquer aux amplitudes )
       $C_T$ ; ( l'opération à appliquer aux types )

Retirer un élément V de VM;
X ← V;
Tant que VM n'est pas vide faire
  Retirer un élément V de VM; (Les composantes de V sont notées  $V_A$  et  $V_T$  )
   $X_A$  ←  $X_A$   $C_A$   $V_A$  ;
   $X_T$  ←  $X_T$   $C_T$   $V_T$  ;
fin tant que;

```

Figure 27: Algorithme de combinaison des valeurs marginales pour l'opérateur ForAll.

Cet algorithme nous garantit que lorsque X n'a qu'une seule valeur marginale, alors la propriété de consistance est respectée. Les valeurs permises pour les opérateurs C_A et C_T étant des opérations commutatives, les lois de combinaison construites par **ForAll** respectent la propriété de multi-symétrie.

Le second opérateur de définition de loi de combinaison est appelé **ForEach** et suit le même principe. Il permet d'appliquer un facteur au résultat de la combinaison. Il permet, par exemple, de calculer la moyenne des valeurs marginales.

1.3.3. Description d'une influence

Les influences sont décrites par des prédicats qui désignent les atomes concernés par l'influence, ainsi que les valeurs des paramètres de cette influence :

(**influence** X Y C Ga Gt Td Tr)

où X est le nom de l'atome source de l'influence, Y le nom de l'atome destination, C est la condition d'activation, Ga et Gt sont des fonctions de gain dynamiques, Td est le délai et Tr est le temps de réponse de l'influence. Rappelons que les valeurs de ces deux derniers peuvent être des intervalles. Ce prédicat contient tous les paramètres nécessaires à la construction de la fonction de mise à jour associée à l'influence décrite.

La syntaxe de la définition d'un gain dynamique peut être illustrée par l'exemple suivant. La commande **Define** définit ici un gain dynamique d'amplitude appelé *GASquare*, qui élève au carré l'amplitude de X, l'atome source de l'influence :

(**Define Dynamic-Gain** GASquare X (X * X))

Les gains dynamiques les plus utilisés sont des fonctions linéaires. Devoir définir de nouvelles fonctions de gain dynamique, pour chaque fonction linéaire, alourdirait la description des modèles. On considère donc toute valeur constante, passée en guise de gain dynamique d'un prédicat d'influence, comme une fonction linéaire. Ainsi la commande :

(**Influence** A B **True** 13 0.5 0 0)

définit une influence de l'atome A vers B, dont la condition d'activation est toujours vérifiée (True), le délai est nul tout comme le temps de réponse. Seuls les gains sont des gains linéaires (i.e. de coefficient égal à 13 pour l'amplitude et 0.5 pour le type). La fonction de mise à jour correspondante à cette influence se réduit donc à :

$$(28) \quad \begin{cases} B_A(t) = 13 \times A_A(t) \\ B_T(t) = 0.5 \cdot A_T(t) \end{cases}$$

En plus des influences d'atome à atome, BIOCAEN permet de décrire des *influences constantes*. De telles influences ne désignent pas d'atome source mais, à la place, une valeur constante. Ces influences constantes peuvent être utilisées pour alléger la représentation du modèle. Le compilateur de BIOCAEN les utilise pour simplifier le graphe global, en remplaçant les atomes non influencés par des constantes. Une influence constante est décrite par le prédicat :

(**Constant** Y C Val_A Val_T Tr)

où Y est l'atome destination de l'influence, C est une condition d'activation, Tr le temps de réponse de l'influence et Val_A et Val_T sont les valeurs constantes, respectivement l'amplitude et le type. Cette définition est équivalente à :

(**Influence** CTE Y C 1 1 0 Tr)

Où CTE serait un atome dont la valeur resterait constamment égale au couple (Val_A, Val_T).

1.3.4. Description d'un composant

Les descriptions de composants sont regroupées dans une section appelée **Components**. Lorsque l'on décrit un composant, on doit aussi indiquer le nom du fichier contenant le modèle générique dont il est l'instance. Les modèles génériques sont stockés dans des fichiers séparés pour faciliter leur réutilisation.

La description du composant doit fournir une valeur à tous les paramètres génériques du modèle désigné. C'est le compilateur qui remplacera les paramètres par leurs valeurs lors de la construction du graphe global.

Les descriptions des modèles de composants doivent débiter par une section appelée **Interface** contenant la description de leur interface de connexion. Cette interface décrit les schémas de connexion et désigne les points de connexion du composant. On y spécifie aussi les noms des paramètres génériques du modèle. Un exemple complet de description d'une modèle de composant est donné en annexe (voir Annexe A, page 177).

1.4. Algorithme de compilation

La compilation a pour but, outre les vérifications de syntaxe de la description des modèles, de construire le graphe global du modèle. De plus, l'algorithme de compilation réalise un certain nombre de simplifications du modèle et de pré-calculs qui allégeront d'autant l'algorithme de simulation. La compilation peut donc être scindée en deux parties distinctes et successives : la génération du graphe global, puis la simplification de celui-ci.

En entrée du compilateur, lui sont transmis le fichier contenant la description du modèle à compiler ainsi que tous les fichiers contenant les descriptions de tous les modèles génériques utilisés. Le

compilateur ne fournit en sortie qu'un seul fichier qui pourra être directement utilisé par le programme de simulation de BIOCAEN.

```

Soit Atoms,      ( l'ensemble des atomes du graphe global )
      Infls;      ( l'ensemble des influences du graphe global )
      Courant,    ( le fichier contenant la description de modèle qui est en train d'être compilé )
      Ouverts;    ( la liste des fichiers actuellement ouverts )

Atoms ← {}; Infls ← {}; Ouverts ← {};      ( Initialisation des variables )
Courant ← le Fichier à compiler;

Procédure Génération du graphe global;
Soit ConSchem, ConPoint; ( l'ensemble des schémas et points de connexions définis par les composants de Courant )
ConSchem ← {};
Conpoint ← {};

Ouverts ← {Courant} ∪ Ouverts; ( ajout de Courant à l'ensemble des fichiers en cours de traitement )

Pour tout composant décrit dans le fichier Courant faire
  CX ← Courant;
  Courant ← Le fichier du composant;
  Si Courant ∈ Ouverts Alors Arrêter l'algorithme; ( La définition n'est pas valide )
  Remplacer tous les paramètres génériques par leurs valeurs;
  Appeler la procédure Génération du graphe global; ( Appel récursif de la procédure )
  ConSchem ← ConSchem ∪ {Schémas de connexions de l'interface de Courant};
  ConPoint ← ConPoint ∪ {Points de connexions de l'interface de Courant};
  Courant ← CX;
Fin Pour tout composant;

Ouverts ← Ouverts - {Courant};

Pour toute Connexion C décrite dans Courant faire
  Si C est une influence Alors
    Si C utilise un nœud P de ConPoint Alors
      C ← L'influence C redirigée vers l'atome désigné par P;
      Infls ← Infls ∪ {C};
  Sinon ( C n'est pas une simple influence mais est une connexion )
    SC ← L'ensemble des influences associées à C dans ConSchem;
    Infls ← Infls ∪ SC;
  Fin Pour toute Connexion C;

Fin de la procédure;

```

Figure 28 : Algorithme de compilation de BIOCAEN, première partie, obtention du graphe global.

1.4.1. Génération du graphe global

La première étape de la compilation consiste donc à parcourir le graphe causal du modèle et à remplacer récursivement les composants par les graphes de leur modèle (voir chapitre 4, partie 2.4, page 101). L'algorithme ci-dessus (Figure 28, page 136) parcourt la hiérarchie de modèles en *profondeur d'abord* en s'assurant de la validité du graphe global, c'est-à-dire qu'aucun descendant d'un modèle n'utilise ce même modèle dans son propre graphe, ce qui créerait une définition récursive (voir la définition d'un graphe valide, en page 99).

Cet algorithme est divisé en deux parties. On commence par remplacer les composants par les graphes de leurs modèles associés (partie descendante du parcours de la hiérarchie). Puis, on reconnecte les atomes entre eux en re-dirigeant les connexions (partie ascendante du parcours). L'arrêt de l'algorithme est garanti par le test de validité qui vérifie si le fichier Courant n'est pas contenu dans l'ensemble des fichiers en cours de traitement (i.e. la variable Ouverts). Ce test interdit les définitions récursives de modèles et empêche de ce fait la procédure de s'appeler elle-même indéfiniment. Si l'algorithme se termine correctement, *Atoms* et *Infls* contiennent tous les atomes et influences du graphe global. Seuls ces éléments sont conservés pour la suite du traitement.

1.4.2. Simplifications du graphe global

La seconde partie de la compilation construit un fichier compact ne contenant que les éléments indispensables à la simulation du modèle. Avant d'appliquer toute une série de simplifications, le compilateur vérifie la calculabilité des valeurs des atomes du graphe (voir page 125).

La valeur d'un atome ne peut être calculée que si les valeurs des atomes dont il dépend ont déjà été calculées. L'ordre de calcul des valeurs des atomes n'est pas innocent. Pour ordonner correctement les atomes, le compilateur doit construire le *graphe de dépendance directe*. Les nœuds de ce graphe orienté sont les atomes du graphe global. Il contient un arc entre deux atomes de X vers Y si et seulement si il existe une influence de délai nul de X vers Y, ou si X est utilisé, sans décalage temporel, dans la condition d'activation d'une influence dirigée vers Y. Tout en ordonnant la liste des atomes, on vérifie qu'il n'existe pas de circuit dans le graphe de dépendance directe. L'algorithme est classique (Figure 29), il construit le dictionnaire des nœuds du graphe et l'utilise pour déterminer le rang des atomes. Le rang d'un nœud est le nombre maximal de prédécesseurs qu'il possède dans le graphe (Xuong 1992), tout nœud d'un cycle possède un rang infini.

On s'assure que l'algorithme se termine bien en supprimant les éléments de *Atoms* au fur et à mesure que l'on les classe dans *ListAtoms*. *Rg* étant un sous-ensemble de *Atoms*, lorsque *Atoms* est vide, *Rg* l'est aussi.

```

Étant donnés
  Atoms,      (L'ensemble des atomes du graphe global)
  DD[X];     (Pour tout atome x, l'ensemble des atomes dont il dépend directement, le dictionnaire de l'atome x)

Soit ListAtoms,  (la liste des atomes correctement ordonnés)
      Rg;          (l'ensemble des atomes de même rang)

ListAtoms ← {};
Rg ← {};
Pour tout élément X de Atoms faire
  Si DD[X] = {} Alors Rg ← Rg ∪ {X};
Fin Pour tout élément X;
Tant que Rg n'est pas vide faire
  Ajouter en tête de ListAtoms les éléments de Rg;
  Atoms ← Atoms - Rg;

  Rg ← {};
  Pour tout élément X de Atoms faire
    DD[X] ← DD[X] ∩ Atoms;
    Si DD[X] = {} Alors Rg ← Rg ∪ {X};
  Fin Pour tout élément X;
Fin tant que;

```

Figure 29: Algorithme de compilation de BIOCAEN, seconde partie, test de validité du graphe global.

A la fin de l'algorithme, si *Atoms* n'est pas vide, alors il existe des circuits dans le graphe de dépendance directe. Le modèle n'est alors pas calculable et on doit arrêter la compilation. Les atomes de même rang dans le graphe de dépendance directe peuvent être calculés indépendamment les uns des autres. On pourrait très bien calculer leurs valeurs en parallèle de manière à gagner du temps de calcul.

En plus d'avoir vérifié la calculabilité des atomes du graphe, on obtient la liste ordonnée des atomes qui permettra de calculer les valeurs des atomes dans le bon ordre (i.e. les valeurs de tous les atomes directement dépendants peuvent être séquentiellement mises à jour). Après avoir ordonné les atomes, l'algorithme de compilation exprime tous les délais en fonction de l'horloge discrète du système (i.e. en les recalant sur les pas de calculs). Le compilateur calcule pour chaque atome le décalage temporel maximal avec lequel l'atome est référencé. Ce décalage maximal sera utilisé par le simulateur. Le compilateur construit les fonctions de mise à jour des influences du graphe. Ces fonctions seront directement transmises au simulateur dans le fichier compilé.

Pour finir, l'algorithme réalise un certain nombre de simplifications du graphe. Les simplifications consistent à calculer les expressions à valeurs constantes et à les remplacer par leurs valeurs. Le compilateur applique ce principe aux paramètres numériques et symboliques mais aussi aux atomes du graphe. Si un atome est une racine du graphe global et qu'il n'est pas exogène, c'est à dire que sa valeur n'est pas dépendante de l'extérieur du modèle, alors sa valeur ne sera pas modifiée au cours de la simulation. Dans ce cas, l'atome est supprimé du graphe et toutes les influences dont il était la source sont remplacées par des influences constantes. De même, si un atome n'est influencé que par des influences constantes dont la condition d'activation est égale à la constante True⁶, alors l'atome est lui aussi supprimé. Le compilateur calcule sa valeur à partir des influences constantes qui étaient dirigées vers lui et remplace toutes les influences dont il était la source par des influences constantes.

1.5. Algorithme de simulation

Du fait des traitements effectués lors de la compilation, l'algorithme du simulateur est relativement simple. Il s'agit d'une boucle qui répète la mise à jour des valeurs des atomes, en incrémentant à chaque fois l'horloge, ceci un nombre fini de fois, déterminé par la durée de la simulation, spécifiée par l'utilisateur. Le compilateur transmet au simulateur uniquement les informations nécessaires à la mise à jour des valeurs des atomes du modèle. Ces informations sont :

1. La liste des atomes ordonnée de manière à ce que la mise à jour des valeurs puisse être effectuée séquentiellement.
2. Pour chaque atome :
 - Le décalage temporel maximal, T_m , avec lequel l'atome est référencé. Le décalage détermine le nombre de valeurs de l'atome à conserver en mémoire ; ces valeurs sont appelées *historique* de l'atome. Le simulateur ne conserve que les T_m dernières valeurs, pour réduire l'occupation de la mémoire.
 - La liste des influences dont l'atome est la destination, et, pour chacune d'elles, sa fonction de mise à jour et la valeur de son délai.
 - La loi de combinaison de l'atome.

⁶ La valeur de vérité de cette constante est, vrai, égale à 1 dans L_2, L_3 et L_{N1} .

En plus de cette description compacte du modèle, on transmet au simulateur un fichier de mesures contenant les valeurs destinées aux atomes exogènes du modèle, ainsi que la liste des atomes dont les valeurs seront enregistrées dans le fichier des observations (voir page 129, Figure 26) et la durée de la simulation. La durée de la simulation est recalée sur l'horloge discrète du système et transformée en nombre de pas d'horloge.

Étant donnés :

TS, (Le nombre de pas d'horloge constituant la durée de la simulation)
 Obs, (La liste des atomes dont les valeurs doivent être relevées)
 Atoms, (La liste ordonnée des atomes)
 TM[x], (Le décalage temporel maximal de l'atome x)
 Inf[x], (Pour tout atome x, fournit l'ensemble des influences dont il est la destination. Pour chaque influence I de cet ensemble, est donné son délai I.d et sa fonction de mise à jour I.Fn(t))
 LC[x]; (La liste de combinaison de l'atome x)

Soit t, (Le top courant de l'horloge discrète)
 VM, (L'ensemble des valeurs marginales de l'atome courant)
 H[x]; (Une liste de valeur, l'historique des valeurs de l'atome x)

t ← 0;

Tant que t < TS **faire**

Lire les valeurs des atomes exogènes dans le fichier de valeurs;

Pour tout élément x de Atoms **faire**

VM ← {};

Pour toute influence I de Inf[x] **faire**

Si t > I.d **alors** (si t n'est pas supérieur au délai de I alors elle est ignorée, car son délai réclame une valeur antérieure à la date de début de la simulation)

VM ← VM ∪ { I.Fn(t) }; (Applique la fonction de mise à jour de I à l'instant t)

Fin pour toute influence I;

Si VM ≠ {} **Alors** (Si il y a des valeurs marginales à combiner alors..)

Ajouter en tête de H[x] le résultat de l'application de LC[x] à VM;

Sinon (L'atome X n'est pas influencé; sa valeur reste constante)

Ajouter en tête de H[x] la première valeur de H[x];

Si t > TM[x] **Alors** (Suppression de la valeur la plus ancienne de l'historique de X)

Supprimer la dernière valeur de H[x];

Si x ∈ Obs **Alors** (Enregistrement des valeurs observées)

Écrire la première valeur de H[x] dans le fichier d'observations;

Fin de pour tout élément x;

t ← t + 1;

Fin de Tant que;

Figure 30: Algorithme de simulation de BIOCAEN.

La complexité de l'algorithme de simulation (Figure 30, page 140) est grandement réduite par l'accès à la liste des atomes du graphe, spécialement ordonnée pour réaliser séquentiellement leur mise à jour. De cette manière, la simulation peut se résumer au parcours de cette liste, en appliquant à chaque fois la loi de combinaison de l'atome considéré aux fonctions de mises à jour des influences actives. Ceci revient, pour un graphe global de n atomes et m influences, à une complexité en temps de $O(n + m)$, ce qui représente une complexité équivalente à un algorithme classique de parcours de graphe (Xuong 1992), pour chaque pas de calcul.

L'instruction primordiale de cet algorithme est donc l'application de la fonction de mise à jour qui permet de calculer les valeurs marginales des atomes (voir page 123). Cette opération nécessite en fait un certain nombre de calculs, comme le calcul de la valeur de la condition d'activation. Lorsque le temps de réponse n'est pas nul, elle peut nécessiter la mémorisation de la valeur marginale, pour sa réutilisation au pas de calcul suivant (voir page 121). Lorsque le délai n'est pas nul, l'opération de mise à jours utilise aussi l'historique de la valeur de l'atome influençant (voir page 120). L'utilisation d'historiques de valeurs permet de réduire les besoins en espace mémoire d'une simulation qui nécessite par ailleurs de conserver un nombre important de données, comme les m fonctions de mises à jours des influences, les n lois de combinaisons des atomes, ainsi que la mémorisation de $m \times n$ valeurs marginales.

1.6. Implémentation

BIOCAEN est implémenté en LISP sous système UNIX. L'utilisation d'un langage applicatif interprété offre, entre autres, deux avantages remarquables. Le premier, au niveau de la compilation, est de permettre une évaluation directe de la syntaxe des descriptions des modèles en utilisant l'interpréteur LISP pour évaluer les expressions décrivant les éléments du modèle. On peut, de plus, tirer ici profit du caractère récursif de l'algorithme de construction du graphe global en utilisant des expressions LISP récurrentes. Le second avantage tient dans l'uniformité de représentation entre les données et les fonctions en LISP (i.e. les listes). On peut ainsi stocker directement dans le fichier compilé les fonctions de mises à jour et les lois de combinaisons sous la forme de lambda-expressions qu'il ne restera plus qu'à appliquer lors de la simulation.

Compte tenu du nombre relativement important de calculs à réaliser à chaque pas de l'horloge, la simulation réclame un grand nombre de ressources système. La durée réelle de la simulation est, de plus, allongée par l'accès, à chaque pas, aux fichiers de valeurs des atomes exogènes et à l'écriture

des valeurs dans les fichiers résultats. Ainsi, sur une station de travail SUNSparc10, en interprétation non compilée sous Le-Lisp, la simulation d'un modèle tel que ceux présentés ci-dessous nécessite plus de 3 minutes de calculs pour réaliser une simulation de 40 secondes, avec un pas de calcul de 1 milliseconde (soit un rapport approximatif de 4.5).

2. APPLICATION ET VALIDATION

Nous allons présenter, dans cette seconde partie, deux applications de BIOCAEN. Les modèles que nous allons décrire ici nous ont servi à tester et valider le simulateur. C'est dans cet esprit qu'il sont incorporés dans ce document. C'est pourquoi nous les présenterons en mettant en priorité l'accent sur l'expression des hypothèses dans notre formalisme, au détriment de l'élaboration même de ces hypothèses, ce dernier point ne relevant en aucune manière des capacités de l'auteur.

La première application de BIOCAEN consiste en la proposition d'un modèle générique regroupant un ensemble assez important d'hypothèses fonctionnelles. Ce modèle a été bâti dans le but avoué de tester les capacités du formalisme à exprimer explicitement diverses connaissances, ou hypothèses, dans un même modèle. Il est présenté comme une tentative de modélisation de certaines fonctionnalités des aires associatives. Il doit illustrer, ici, l'éventail des possibilités offertes par notre représentation. La pertinence des hypothèses qu'il exprime ne sera pas particulièrement discutée.

La seconde application représente une méthodologie plus proche de l'utilisation finale de BIOCAEN. Les modèles proposés reposent sur des hypothèses élaborées à la suite d'études d'activation, tirées de la littérature en neuropsychologie, ou sur des résultats d'expériences sur l'animal. L'objectif de la modélisation consiste à confronter diverses hypothèses pouvant expliquer les phénomènes observés.

2.1. Modèle générique de zones associatives

Le modèle que nous allons présenter ici est originellement inspiré du modèle présenté dans (Lacotte 1996) et (Pastor *et al.* 1997). La première version de ce modèle pour BIOCAEN est présentée dans (Lafon *et al.* 1997). Nous en présentons ici une version non publiée. Le but de ce travail est de démontrer les capacités du formalisme à exprimer diverses hypothèses fonctionnelles,

comme des mécanismes de filtrage et d'intégration de l'information, ou bien encore des processus d'activation, d'inhibition et de mémorisation. Comme certaines de ces fonctionnalités apparaissent dans le comportement des aires associatives (e.g. intégration d'informations convergentes, inhibition ou mémoire de travail), on peut proposer ce modèle comme une première tentative de modélisation de ce type de fonctionnement.

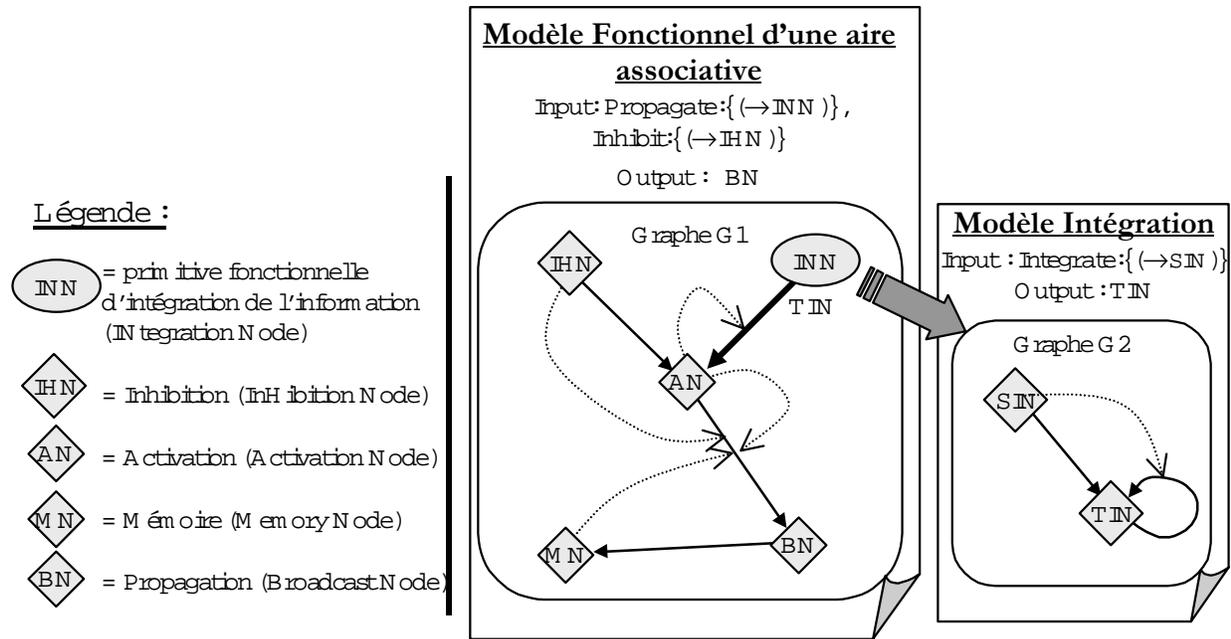


Figure 31 : Modèle fonctionnel générique d'une aire associative.

Comme tout modèle générique, ce modèle se place dans le contexte de reproductibilité des fonctionnalités cérébrales, mais on ne peut présumer de l'étendue de cette hypothèse. Ce qui est important pour notre première modélisation, c'est de souligner la potentielle réutilisation des mécanismes décrits par ce modèle. Les modèles qui suivront en reprendront d'ailleurs certaines.

Pour décrire ce modèle nous allons suivre la propagation et les transformations qui seront appliquées à l'information (Figure 31), de l'entrée de l'information dans le zone jusqu'à sa propagation aux zones suivantes. Avant d'entrer dans le détail du modèle, on peut observer que chaque primitive fonctionnelle implémentée dans ce modèle est identifiée par un nœud du graphe du modèle.

Le fichier source de description, destiné à BIOCAEN, est donné en annexe (Annexe A) où il est largement commenté.

2.1.1. Interface d'entrée du modèle

En entrée du modèle nous allons exprimer l'*hypothèse* qu'il n'existe que deux manières de transmettre une information à une aire fonctionnelle associative. La première manière consiste en une propagation d'information à intégrer (un stimulus). La seconde manière consiste en la propagation d'un signal inhibiteur qui devra empêcher l'activation de la zone.

Notre modèle décrit deux types de connexion en entrée, associées respectivement à la propagation et à l'inhibition. Chacune de ces connexions pourra être utilisée zéro, une ou plusieurs fois dans le modèle structurel qui utilisera ce modèle générique. Les liens représentés ne pourront pas changer de fonction, ce qui présuppose une identification fonctionnelle des connexions au niveau structurel du modèle.

La principale différence entre les deux schémas de connexion est leur nœud de destination. Les connexions de propagation seront dirigées vers la primitive fonctionnelle d'intégration, identifiée par le nœud INN, alors que celle correspondant à une connexion inhibitrice est dirigée vers un autre nœud, IHN, regroupant l'inhibition totale de la zone.

2.1.2. Nœud d'intégration

Les primitives d'intégration de l'information jouent, dans notre modèle, un rôle important. Nous les avons modélisées en utilisant un sous-modèle fonctionnel, appelé « modèle d'intégration ». Après avoir filtré l'information en entrée, elles réalisent deux intégrations successives, une intégration spatiale et une intégration temporelle.

a) Filtrage en entrée de l'information.

Le filtrage des informations est la première opération à être appliquée à toutes les informations convergeantes vers la zone. La zone écrase les informations pour lesquelles elle n'est pas réceptive, ce qui revient à effectuer une sélection des informations en entrée.

Ce mécanisme est bien connu pour les aires fonctionnelles primaires qui sont plus ou moins réactives aux stimuli qu'elles reçoivent, ceci revient à considérer que la zone primaire possède une liste, *fixée*, des stimuli ordonnés en fonction de leur catégorie, du stimuli auquel elle est le plus réceptive à celui auquel elle l'est le moins. Cependant, il est vraisemblable que pour des zones plus profondes (correspondant à des aires associatives), cette liste soit plus dynamique, probablement contrôlée. Pour l'instant nous la considérerons comme un paramètre générique de notre modèle.

Cette liste est appelée Table de préférence de la zone, elle est représenté par une valeur de type que l'on pourra définir, par exemple, en utilisant les opérateurs d'O(M). Cette table est utilisée dans le gain de l'influence dirigée vers le nœud SIN, de manière à « écraser » l'information en fonction du poids associé à sa catégorie dans la table. L'expression du gain d'amplitude peut être résumé par la fonction G_A suivante, si TPT est la table de préférence :

$$(29) \quad G_A(X(t)) = \uparrow (X_T(t) \cap TPT) \times X_A(t)$$

b) Intégration spatiale.

L'intégration spatiale représente la faculté qu'a la zone de combiner de multiples informations convergentes. On peut exprimer diverses hypothèses sur la manière dont les informations sont combinées entre elles, pourvu que l'opération respecte les propriétés des lois de combinaison. C'est l'atome SIN qui réalise cette intégration, dont la plus simple expression pourrait être une simple sommation des informations. Dans le modèle générique, la loi de combinaison de ce nœud restera non spécifié (i.e. elle est un des paramètres génériques du modèle).

c) Intégration temporelle

La seconde opération consiste en une véritable intégration, temporelle, de l'information, identifiée par l'atome TIN. L'intégration est simulée par une boucle de TIN sur lui-même avec un délai non nul, égal à 1 milliseconde, de manière à combiner l'information entrante à celle qui était en cours d'intégration. Nous supposons que l'intégration est conditionnée par la cohérence temporelle de l'information présentée. Cette hypothèse est modélisée par une condition d'activation de l'influence d'intégration (i.e. la boucle, TIN→TIN). De plus, cette condition d'activation est évaluée dans une logique multivaluée de telle manière que :

1. Si la catégorie du nouveau stimulus au temps t est identique à celle du stimulus au temps t-1, alors l'intégration est totalement activée. Elle est réalisée pleinement.
2. Si la catégorie du nouveau stimulus au temps t est totalement différente de celle du stimulus au temps t-1, alors l'intégration est totalement supprimée. Elle est réinitialisée.
3. Si la catégorie du nouveau stimulus au temps t n'est ni totalement identique à celle du stimulus au temps t-1, ni totalement différente, alors l'intégration est maintenue dans un degré proportionnel au degré d'égalité des deux types (i.e. $V_{N1}(SIN_T(t) \equiv TIN_T(t-1))$).

2.1.3. Nœud d'inhibition

Le nœud d'inhibition, IHN, représente la sommation des informations inhibitrices perçues par la zone à l'instant courant. L'amplitude de IHN est représentative de la *force* de l'inhibition. Cette valeur (i.e. $IHN_A(t)$) est utilisée pour interdire l'activation de la zone.

2.1.4. Nœud d'activation

L'activation de la zone est représentative de l'avancement de l'intégration des stimuli par la zone. Le nœud AN est influencé par les nœuds d'entrée INN et IHN. L'influence de INN est une influence positive et l'influence de IHN est négative (i.e. un gain linéaire avec un coefficient négatif).

Sauf en cas d'inhibition, la valeur de AN est égale à celle de TIN. Dans ce cas, l'amplitude de la valeur de cet atome est représentative du niveau d'activation de la zone.

L'influence $AN \rightarrow BN$ est l'influence qui déclenche la propagation de l'information. C'est donc la condition d'activation de cette influence qui va déterminer quand la zone va propager l'information qu'elle a intégrée. Cette condition prend en compte divers aspects comme la non inhibition de la zone (i.e. $IHN_A(t) = 0$). Nous supposons ici que la décharge se produit en général lorsque l'amplitude de l'atome AN dépasse un certain seuil, appelé *seuil d'émission*. Cette valeur numérique est un paramètre générique de la zone.

2.1.5. Nœud de propagation

Le nœud de propagation, noté BN (i.e. Broadcast Node), est l'unique point de connexion de sortie de notre modèle. Il matérialise l'hypothèse selon laquelle la zone propage la même information à toutes les zones auxquelles elle est connectée.

2.1.6. Nœud de mémoire de travail

Une primitive fonctionnelle de mémorisation est implémentée par l'atome MN. Il correspond à une mémoire de travail propre à la zone, et donc structurellement localisée dans celle-ci. L'hypothèse, formulée ici, est que l'information mémorisée est l'information propagée (i.e. influence $BN \rightarrow MN$). L'influence liant BN à MN est une influence de transfert dont le temps de réponse détermine la vitesse de mémorisation de la zone. Plus le temps de réponse de l'influence est court et plus la zone mémorise, et oublie, vite les informations qu'elle a traitées. Ce paramètre fait partie des paramètres génériques du modèle.

La valeur de ce processeur est utilisée pour déclencher l'activation de la zone (i.e. l'influence AN→BN) avant que l'amplitude de AN n'atteigne le seuil d'activation si et seulement si le type de l'information actuellement traitée par la zone fait partie de l'information mémorisée par celle-ci. Ceci modélise l'hypothèse stipulant que des présentations successives d'un même stimulus conduisent à une facilitation de l'activation de la zone.

2.1.7. Surcharge de la zone

En plus des primitives fonctionnelles, incarnées par les atomes, que nous venons de décrire, on peut aussi représenter d'autres mécanismes en agissant sur les paramètres des influences du modèle (i.e. comme on le fait, par exemple, avec l'inhibition). Nous procédons ainsi, pour décrire un mécanisme de prévention contre la « surcharge » de la zone.

L'état de surcharge peut se produire lorsque la zone reçoit un flot d'information important et qu'elle augmente son activité sans parvenir à émettre, par exemple, parce que l'information est très bruitée). Nous faisons l'hypothèse que lorsque cela se produit, la zone coupe ses entrées afin d'éviter une saturation non production d'information. Nous appelons ce phénomène, état de surcharge. Nous représentons cette hypothèse avec une condition d'activation telle que, si l'amplitude de l'activation dépasse un certain seuil (i.e. le seuil de surcharge, ST), alors les connexions de propagation deviennent inactives.

2.1.8. Conclusion

Etant donné le nombre d'hypothèses représentées dans ce modèle, on peut difficilement en imaginer une application directe. Cependant, elles peuvent être indépendamment réutilisées dans des modèles plus simples, pour une première validation. Par exemple, le modèle générique de l'intégration de l'information peut être directement réutilisé.

Enfin, pour conclure, on peut remarquer que malgré le nombre d'hypothèses assez important, le modèle reste tout de même assez peu complexe : Il ne contient que six processeurs et sept influences. On peut aussi souligner la variété des hypothèses, qui peuvent porter sur des aspects très différents de la propagation. Ces hypothèses peuvent être modifiées assez simplement. Prenons l'exemple de la primitive fonctionnelle de mémorisation. Nous avons formulé l'hypothèse que l'information mémorisée était celle que la zone propageait en décrivant l'influence BN→MN. Mais on pourrait, par exemple, supposer que l'information mémorisée est l'information qui est intégrée, en remplaçant cette influence par une influence TIN→MN. De cette manière, on représenterait

une hypothèse totalement différente selon laquelle une zone conserve des traces mnésiques des informations qui ne l'ont pas forcément conduite à émettre.

Cette dernière observation souligne deux aspects importants de notre approche. Le premier est le caractère *expressif* de notre formalisation. L'utilisation de graphes favorise une représentation visuelle et simplifiée de nos modèles (Figure 31, page 143). On obtient ainsi une vision globale du modèle et une idée claire des dépendances des unités du système. Le second aspect concerne plus directement la notion de composant. La modification des hypothèses exprimées dans le modèle fonctionnel ne modifie pas les modèles structurels qui l'utilisent. Dans une modélisation complexe faisant, par exemple, intervenir un grand nombre de zones, la modularité de la modélisation est un grand avantage.

2.2. Modèles de la réponse de l'aire visuelle primaire à un stimulus simple

Nous allons maintenant présenter une application plus poussée de notre formalisme. Cette application représente, à une petite échelle, la démarche d'une véritable modélisation utilisant BIOCAEN. Nous venons de tester l'expressivité de notre approche, nous allons maintenant passer à la simulation proprement dite de nos modèles.

Nous présentons deux modèles dont l'objectif est de reproduire les résultats des expériences TEP de Peter T. Fox et Marcus E. Raichle (1985, 1984). Les modèles présentés ici ont été originellement publiés dans (Pastor *et al.* 2000).

2.2.1. Présentation de l'étude d'activation

L'expérience de Fox et Raichle est une étude d'activation qui a le mérite d'être assez simple à comprendre. Elle ne fait intervenir que des processus automatiques, où le sujet n'a aucun contrôle à exercer sur le déroulement de l'expérience, aucune réponse à formuler. Il reste passif durant toute la durée de l'expérience. Pour une description détaillée des études de Fox et Raichle, le lecteur devra consulter les articles originaux (Fox & Raichle 1985, 1984) ou bien des études plus récentes sur le sujet (Mentis *et al.* 1997).

Le protocole des expérimentations a été appliqué à 9 sujets pour la première version de l'expérience (Fox & Raichle 1984), puis à 4 sujets pour la seconde (Fox & Raichle 1985). Les sujets reçoivent une injection de marqueur radioactif avant d'être placés sous une caméra PET (voir chapitre 1,

partie 2.2, page 22). La caméra mesure l'accumulation du marqueur durant environ 40 secondes (i.e. durée minimale d'un *run* durant lequel la stimulation doit être maintenue) pour obtenir 1 image.

On place sur les yeux du patient des lunettes supportant de petits écrans composés d'une matrice de 36 diodes. Les stimuli sont visuels. Dans la première version de l'expérience, les stimuli sont des flash (toutes les diodes s'allument puis s'éteignent en même temps). Dans la seconde version de l'expérience, une diode sur deux est toujours allumée, comme sur un damier. Les stimuli sont des inversions des diodes (i.e. inversion du damier). Les stimuli ont une durée (5 millisecondes) et une intensité lumineuse (6400 \AA pour le flash) constantes. La seconde expérience permet de conserver une luminosité constante (égale à 6240 \AA) durant toute l'expérience.

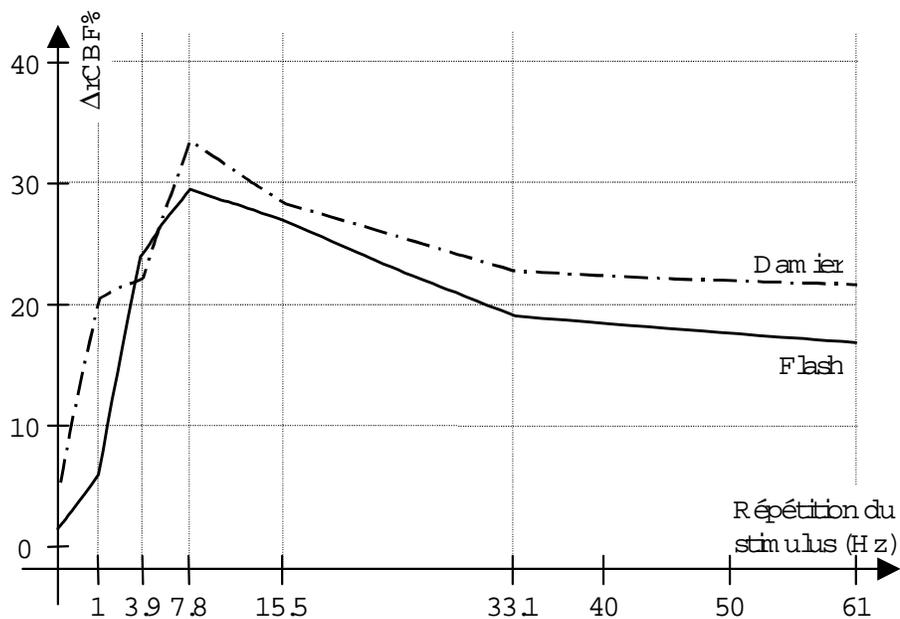


Figure 32: Résultats de l'expérience de Fox et Raichle, d'après Fox et Raichle (1985).

Suivant la version de l'expérience, la stimulation débute 40 ou 60 secondes avant l'acquisition de l'image PET. Un stimulus identique (i.e. Flash ou Inversion) est soumis au sujet, pour tout le *run*. Entre chaque *run*, on fait simplement varier la fréquence de présentation des stimuli, de 1 à 61 Hz.

Fox et Raichle mesurent la variation du débit sanguin local (rCBF) dans la région du cortex strié (i.e. l'aire visuelle primaire). La figure ci-dessus (Figure 32) résume, pour l'ensemble des sujets, les valeurs des variations relatives moyennes du rCBF en fonction de la fréquence de stimulation. Ces valeurs sont obtenues par comparaison aux mesures effectuées au repos (i.e. lorsque le sujet ne percevait pas de stimulation) et elles sont exprimées sous forme de pourcentage ($\Delta rCBF\%$).

Les résultats montrent que le débit sanguin et, *a priori*, la réponse neuronale sous-jacente, ne varient pas linéairement en fonction de la fréquence d'apparition d'un stimulus. Compte tenu des variations inter-individuelles, les auteurs observent que le $\Delta rCBF\%$ augmente jusqu'à 7.8 Hz, pour atteindre un maximum entre 7.8 et 15 Hz et décroît ensuite lentement pour se stabiliser largement en dessous de la valeur maximale (une différence d'environ -10%). Les auteurs soulignent donc que la fréquence de présentation du stimulus peut influencer sur tout système neuronal (Fox & Raichle 1985) et recommande son contrôle pour toute étude, et pas seulement celles impliquant une stimulation visuelle. En ce qui concerne la dépendance proprement dite du rCBF par rapport à la fréquence de présentation, les auteurs imputent la décroissance à deux causes possibles, ou à leur combinaison. La première cause avancée par les auteurs serait une baisse de l'amplitude de l'activation neurale (i.e. une sorte d'habituation à la stimulation). La seconde cause pourrait être une absence de réponse à certains stimuli, lorsque ceux-ci sont trop rapprochés les uns des autres dans le temps.

Outre l'extrême simplicité du réseau d'aires cérébrales impliquées dans cette étude (1 seule zone), nous attirons l'attention sur l'aspect générique des observations des auteurs mais aussi, et surtout, par la confrontation de deux hypothèses. Deux modèles ont été construits dans notre équipe et présentés dans (Pastor *et al.* 2000). Ils fournissent une argumentation en faveur d'une action conjuguée des deux causes de baisse énoncées par Fox et Raichle, tout en formalisant et précisant les hypothèses avancées. La description BIOCAEN complète et commentée est donnée en Annexe (Annexe A, page 177).

2.2.2. Premier modèle : La capacité d'apprentissage de l'aire visuelle primaire

Le premier modèle (Figure 33) reprend certains des mécanismes que nous avons exprimés dans le modèle générique précédent (partie 2.1 de ce chapitre). Le réseau de zone modélisé ne fait intervenir que l'aire visuelle primaire (i.e. V1 dans la figure).

Le nœud IGN réalise un filtrage en entrée de l'information, en fonction d'une table de préférence de types qui détermine les catégories d'information pour lesquelles la zone est la plus réceptive. Dans cette application, la table de préférence est minimaliste, puisque qu'un seul type de stimuli est attendu (i.e. des flash ou des inversion de damier). Après avoir été filtrée, l'information est directement transférée au nœud AN, par l'influence $IGN \rightarrow AN$. La zone ne s'active et ne transmet l'information traitée, via OGN, que lorsque l'amplitude du nœud AN dépasse un seuil d'émission (FT), ou lorsque la catégorie de l'information traitée par AN correspond *suffisamment* à celle de l'atome de mémorisation, l'atome MN, qui conserve une image des informations traitées

précédemment. Cette dernière condition permet d'accélérer l'émission de la zone, en autorisant la zone à émettre avec un niveau d'activation plus faible que celui requis par FT. Lorsque la zone émet (i.e. lorsque l'amplitude OGN est non nulle), on désactive l'influence $IGN \rightarrow AN$. La zone devient alors réfractaire et ne peut traiter une nouvelle information que lorsque l'amplitude de OGN redevient nulle.

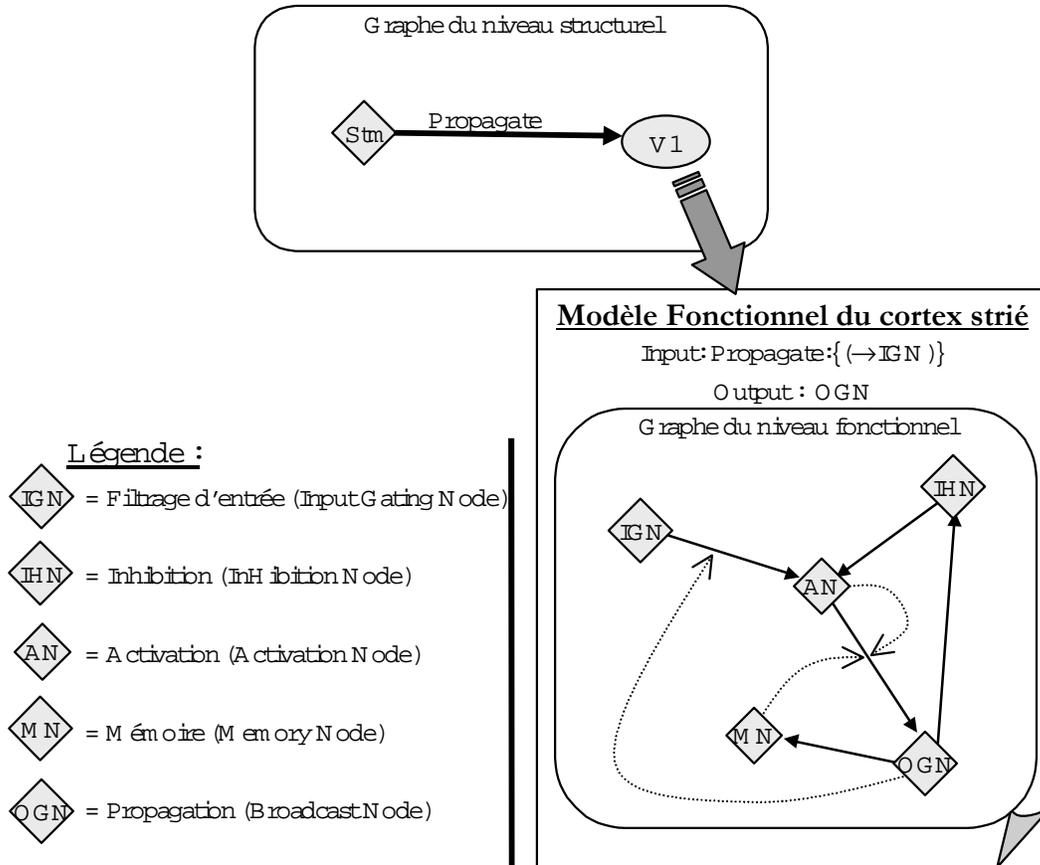


Figure 33 : Modèle fonctionnel du cortex strié, d'après Pastor et al. (2000).

L'atome IHN représente l'inhibition de la zone, il influence négativement AN ce qui empêche la zone d'émettre. Ce nœud est influencé par OGN ($OGN \rightarrow IHN$), ce qui revient à dire que la zone s'auto-inhibe après avoir émis. Ce mécanisme d'auto-inhibition allonge la période réfractaire de la zone.

2.2.3. Second modèle : La boucle cortico-thalamique

Le second modèle proposé modélise le fonctionnement de la boucle cortico-thalamique. Chaque région corticale semble en effet connectée à une partie du thalamus, partie sous-corticale du diencephale (voir page 14). Ces boucles sont supposées réguler l'excitabilité des cellules pyramidales des cortex, en réduisant leur seuil de décharge (Pastor *et al.* 2000). Ce second modèle propose donc

un réseau structurel représentant cette organisation (Figure 34). Chacune des deux zones impliquées possède son propre modèle fonctionnel.

Le modèle fonctionnel de l'aire visuelle primaires, la zone V1, est quasiment identique à celui du modèle précédent à l'exception près de la disparition de la mémoire de travail et de l'ajout d'un seuil d'émission dynamique. Ce seuil, matérialisé par l'atome FT, est connecté à la structure thalamique via le schéma de connexion *Firing-control*, qui définit un nouveau type de connexion pour ce modèle. L'information traitée par le cortex strié est propagée à la structure thalamique, via la connexion *Propagate(V1→Thalamus)*. La structure thalamique possède un modèle fonctionnel très simplifié, puisqu'il n'est constitué que d'un atome d'entrée et un atome de sortie, liés par une influence de transfert.

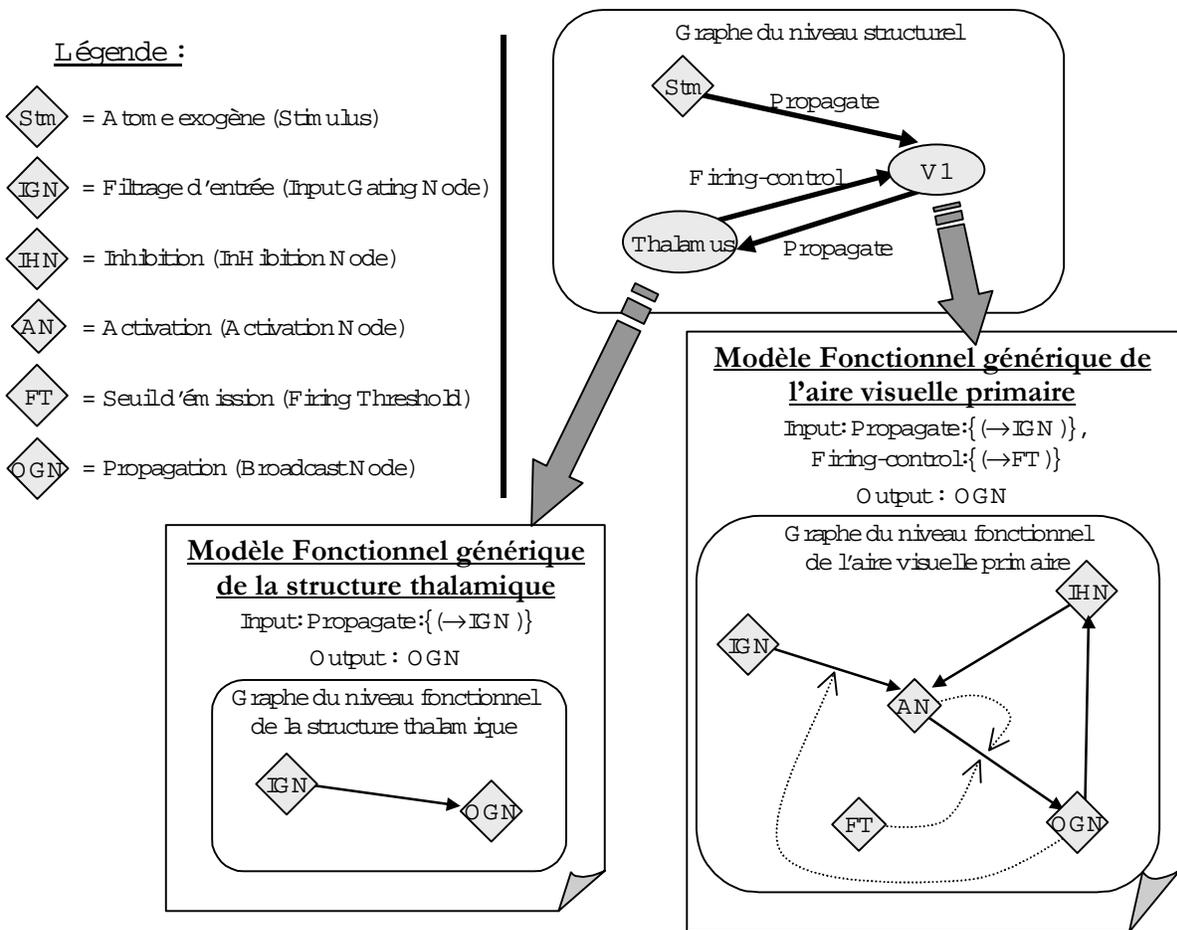


Figure 34: Modélisation de la boucle cortico-thalamique, d'après Pastor et al. (2000).

Ce second modèle peut être vu comme une implémentation plus « biologique » du premier modèle. La modélisation de la mémoire est ici faite par la boucle structurelle, alors que dans le modèle précédent elle était faite au niveau fonctionnel.

2.2.4. Simulation

Après avoir compilé la description de chacun des deux modèles, nous avons réalisé plusieurs simulations en utilisant des fichiers de mesures présentant un stimulus, de type Flash, d'une durée de 5 millisecondes. La fréquence d'apparition du stimulus varie d'un fichier à l'autre. On obtient, dans le fichier de mesure, les valeurs des différents atomes et en particulier celle de AN (Figure 35). La durée de la simulation est de 40 secondes, ce qui avec un pas d'échantillonnage de 1 milliseconde, représente 40.000 valeurs mesurées pour chaque atome.

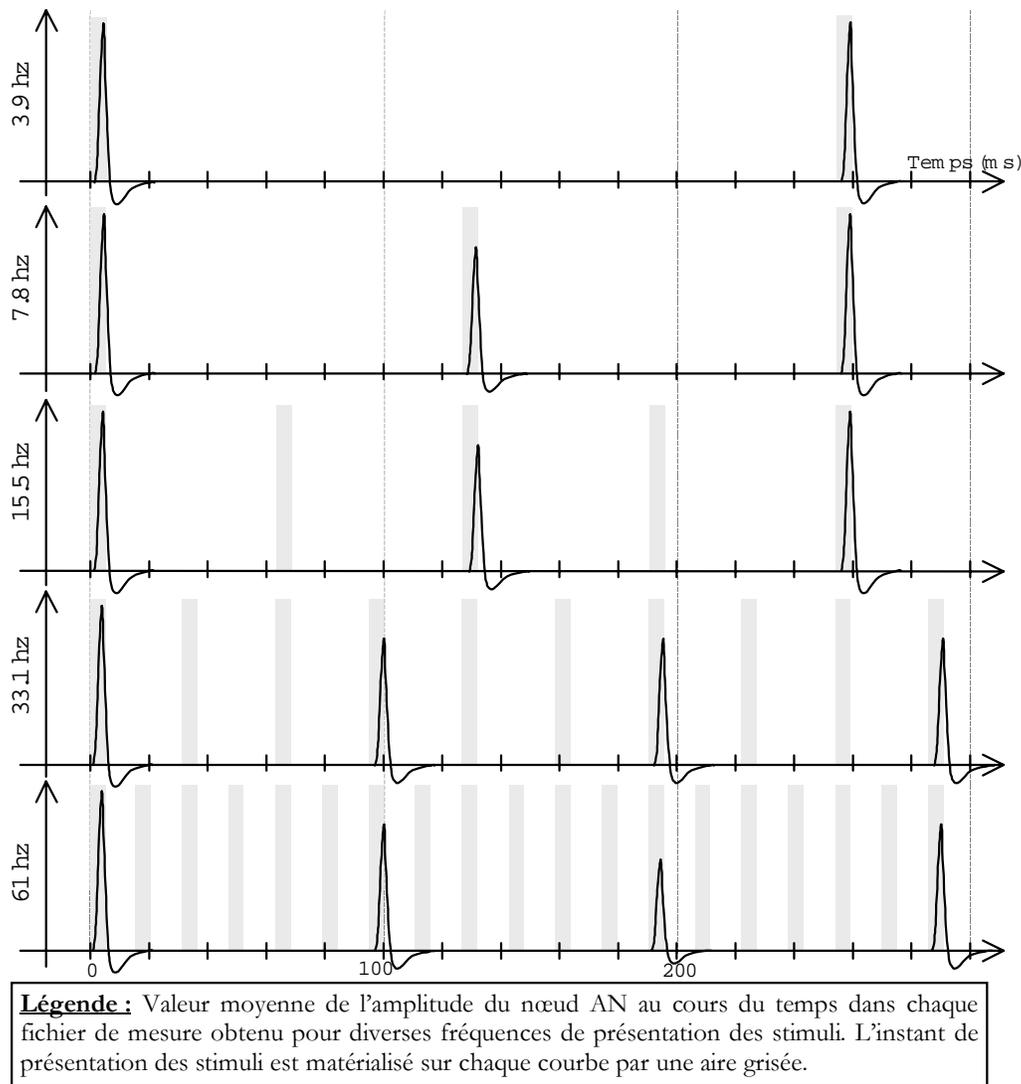


Figure 35: Extraits des fichiers de mesures (Mesures de l'amplitude de l'atome V1 AN).

Les valeurs mesurées sont alors comparables à des mesures de l'activité de populations de neurones, dans la localisation et dans le temps. A partir de ces valeurs (Figure 35), on souhaite obtenir des valeurs comparables à celles obtenues par Fox et Raichle, des mesures du rCBF obtenues au bout de 40 secondes. Pour y parvenir, il faut intégrer l'amplitude des valeurs mesurées de l'activation

pendant le laps de temps correspondant à un *run* (i.e. la durée d'acquisition d'une image TEP). Ceci correspond à 40 secondes. Nous sommes les valeurs activations de la zone correspondant à l'amplitude du nœud AN. Cette somme est notée $\Delta rCBF^{BC}$. Il faut ensuite ramener ces valeurs à la même échelle que les valeurs de l'expérience (Figure 32), c'est à dire à un rapport activité sur repos. Pour y parvenir nous utilisons un rapport direct entre la somme des mesures de l'expérience, notées $\Delta rCBF^{0\%exp}$, et la somme de nos mesures, $\Delta rCBF^{BC}$. De cette manière nous prenons en compte la même mesure du repos qui avait permis d'obtenir le $\Delta rCBF^{0\%exp}$. On obtient donc une valeur normalisée, $\Delta rCBF^{0\%BC}$, à partir de l'équation suivante (Pastor *et al.* 2000) :

$$(30) \quad \Delta rCBF\%^{BC} = 100 \times \Delta rCBF^{BC} \times \frac{\sum \Delta rCBF\%^{exp}}{\sum \Delta rCBF^{BC}}$$

Nous obtenons ainsi une mesure qualitative comparable à celle de l'expérience initiale (Figure 36, Pastor *et al.* 2000). Le comportement des deux modèles est sensiblement identique à celui observé par Fox et Raichle, à savoir une phase d'augmentation jusqu'à environ 7.8 Hz, puis une lente décroissance et enfin une stabilisation au delà de 33.1 Hz.

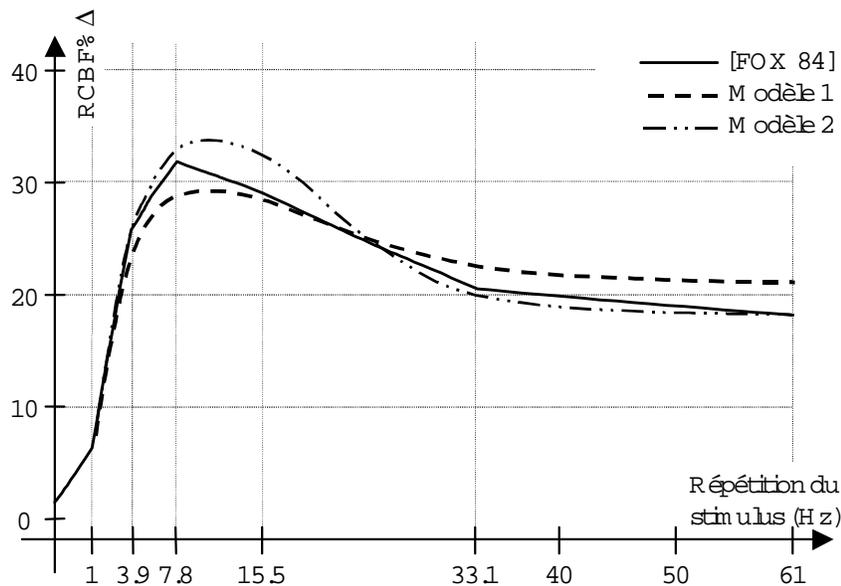


Figure 36 : Résultats globaux de la simulation, d'après Pastor *et al.* (2000).

Nos modèles permettent d'interpréter, en termes de réponses de populations neuronale, à une échelle temporelle de l'ordre de la milliseconde, la réponse qui n'a pu être observée qu'en termes de données d'activation très intégrées, à une échelle de temps de 40 secondes. En observant le détail de nos fichiers de mesures et notamment les valeurs du nœud AN (Figure 35, page 153), on

observe que l'activation de la zone correspond bien aux hypothèses de Fox et Raichle. Il y a effectivement une atténuation de l'activation pour les fréquences de présentation élevées, mais l'on observe aussi que l'activation n'est pas systématique pour les hautes fréquences. C'est la *conjonction* de ces deux phénomènes qui nous permet de reproduire les valeurs intégrées que Fox et Raichle ont pu observer. On voit alors que les modèles permettent de trancher parmi les hypothèses formulées par Fox et Raichle (1984).

En plus de cette corroboration globale des hypothèses des neuropsychologues, la formalisation de nos modèles permet d'en expliquer en détail leur fonctionnement. L'absence de réponse à certains stimuli est dû à la période réfractaire de la zone, conséquence d'une auto-inhibition. Ce phénomène se déclare, dans nos modèles, pour des fréquences assez faibles, à partir de 15.5 Hz. D'une manière générale, il permet de ne pas établir une relation linéaire entre fréquence de présentation et valeur de l'activation. Cependant, l'observation de la courbe montre que ce phénomène seul permettrait au mieux de ralentir l'augmentation de l'activation (puisque'il y a une réponse sur deux à 15.5 Hz, une sur trois à 31 Hz...). La diminution de l'activation, entre 15.5 et 33.1 Hz, ne peut être obtenue que par l'ajout d'une atténuation de la réponse à chaque stimulus. L'atténuation est ici expliquée par l'habituation aux stimuli dû aux mécanismes de mémorisation dont les modèles sont dotés (i.e. modélisation fonctionnelle ou boucle thalamo-corticale). De cette manière, même si le nombre de réponses augmente, leur amplitude respectives étant plus faible, l'activation totale au bout de 40 secondes aura été moins élevée. Ces observations nous permettent donc de confirmer l'hypothèse d'une action conjugué des deux mécanismes, mais elles nous permettent surtout de fournir une explication détaillée et neuropsychologiquement plausible de leur mise en œuvre.

Pour conclure, nous soulignerons encore le caractère explicite des hypothèses modélisées, l'auto-inhibition de la zone est, par exemple, clairement identifiable sur les graphes des modèles (par le cycle $AN \rightarrow OGN \rightarrow INN \rightarrow AN$, Figure 33 et Figure 34). Il en est de même pour la mémorisation fonctionnelle ou la boucle thalamo-corticale. Même si les modèles construits pour BIOCAEN représentent une modélisation simpliste du fonctionnement cérébral, ils permettent de reproduire son comportement. Ils offrent surtout un complément de validité aux hypothèses avancées, et permettent d'en dégager de nouvelles qui seront mieux définies, car elles seront exprimées dans une représentation rigoureusement formalisée.

3. CONCLUSION

Nous venons de décrire les deux premières applications de BIOCAEN. Elles illustrent certaines des possibilités de notre formalisme à exprimer de manière directement interprétable des hypothèses de haute abstraction sur l'organisation fonctionnelle du cerveau. La seconde application illustre plus particulièrement les capacités de BIOCAEN à servir de terrain formel pour la confrontation d'hypothèses. On doit tout spécialement souligner la relative simplicité des modèles que nous avons présentés. Une dizaine de processeurs nous suffisent pour simuler une propagation qui, dans la réalité, implique un très grand nombre d'unités. Cette relative simplicité est en partie obtenue grâce au haut niveau d'abstraction de la représentation de l'information. L'architecture fondée sur des graphes causaux contribue elle aussi à améliorer l'interprétation de nos modèles. L'organisation hiérarchique des graphes empêche que l'augmentation du nombre de nœuds ne réduise cette lisibilité.

En ce qui concerne les futurs développements techniques de BIOCAEN lui-même, les perspectives *immédiates* concernent essentiellement une augmentation des capacités de description du langage. En effet, plusieurs possibilités de notre formalisme ne sont pas aujourd'hui exploitées par le simulateur. La principale illustration concerne les lois de combinaison existantes, dont les opérateurs de définition restent assez limités, alors que certaines études ont déterminé les classes de polynômes qui respectent les propriétés des lois de combinaison (Prats & Travé-Massuyes 1999). Les autres possibilités d'augmentation concernent la définition de schémas de connexion plus évolués et l'extension de l'utilisation de paramètres génériques. Les connexions actuelles permettent de définir des schémas remplaçant des influences qui seront dupliqués à chaque nouvelle utilisation du schéma. Toutes ces influences seront dirigées vers le même nœud, mais on peut, par exemple, imaginer des connexions qui dupliqueront le nœud qu'elles désignent. On représenterait de cette manière, avec un même schéma, des processus de connexion indépendants les uns des autres (e.g. ce type de mécanisme était proposé dans le modèle générique présenté par Bertrand Lacotte (1996)). Concernant les paramètres génériques, ils peuvent, dans la version actuelle de BIOCAEN, prendre la place de tout paramètre numérique ou symbolique, mais ils ne peuvent prendre la place d'un atome, d'un composant ou d'une portion du graphe.

Nous avons montré qu'il était possible de construire des modèles de l'organisation fonctionnelle globale du cerveau, tout en conservant une représentation explicite de la propagation de l'information. Nous ne nous sommes pas concentrés sur une modélisation unique, limitée à une expérience spécifique. Au contraire, nous avons choisi de construire un formalisme original, entièrement dédié à cette représentation, offrant ainsi un cadre ouvert pour ce type de modélisation, et dont BIOCAEN représente la réalisation directe. L'objectif ultime de notre démarche reste, à terme, de proposer un outil d'aide à l'interprétation des résultats des études d'activation telles que celles qui sont menées au sein de l'unité 455 de l'INSERM.

1. FORMALISME GENERAL POUR LA MODELISATION COGNITIVE

Notre ambition a donc été de proposer un formalisme dédié à la modélisation de la propagation de l'information au niveau des aires cérébrales. Dans cette optique, nous avons toujours cherché à conserver une représentation très générale qui pourra être utilisée dans le cadre de modélisation le plus vaste possible. Ce type d'approche est en définitive assez peu courant dans le domaine des approches structurales (Chapitre 2, partie 3.3). Dans le même ordre de contraintes, il faut aussi souligner le lien étroit que nous avons voulu conserver entre nos modèles et les études d'activation (ainsi que les techniques d'imagerie qui leur sont associées).

1.1. Représentation de l'information

Le choix d'une représentation en couple de l'information illustre parfaitement notre souci de généralité. Notre représentation abstraite, à deux dimensions, de l'information propagée par des populations de neurones, nous assure un bon compromis entre plausibilité neuronale et transparence sémantique (i.e. au sens défini au chapitre 2, page 31). L'amplitude, qui reflète l'énergie de l'information peut être rapprochée, en premier lieu, d'un point de vue métabolique des mesures d'activations obtenues par les diverses techniques d'imagerie tomographique (IRMf, TEP) ou de l'amplitude du signal des techniques de surface (EGG, MEG). En second lieu, elle peut être

rapprochée, d'un point de vue physiologique cette fois, du taux de décharge ainsi que du nombre de neurones localement impliqués dans le traitement. Même si le lien direct entre les mesures d'activation et l'information neuronale n'est pas précisément connu, ces mesures en restent les seules évidences.

Le type, qui se rapporte à la catégorie de l'information, peut être facilement justifié au niveau des stimuli, dont la catégorisation est aisée. On peut ainsi catégoriser des tons purs suivant leur fréquence. Des stimuli plus « cognitifs » peuvent aussi être catégorisés. Des phonèmes peuvent, par exemple, être différenciés selon leurs traits phonétiques, en voyelles et consonnes. Ces dernières peuvent être, à leur tour, être catégorisés en sourdes ou sonores, constrictives ou occlusives (e.g ; « s » constrictive sourde, « z » constrictive sonore, « p » occlusive sourde, « b » occlusive sonore). La notion de catégorie est aussi sous-jacente aux mécanismes cérébraux. On peut ainsi citer en exemple l'organisation interne des aires primaires (auditives, visuelles ou sensitives) où l'on trouve des sous-populations de neurones spécialisées dans le traitement de catégories particulières de stimuli (Arbib *et al.* 1998, Ardila 1993, Alexander *et al.* 1992). Mais cette organisation peut être observée au delà des aires primaires (Alexander *et al.* 1992), ce qui tendrait à prouver l'existence de catégories dans l'ensemble du traitement cérébral. Le type, tout comme l'amplitude, peut donc être considéré comme une caractéristique générale de l'information cérébrale.

1.2. Traitement automatique de l'information

La caractéristique fondamentale du traitement de l'information dans nos modèle est le traitement parallèle et distribué, dû à l'utilisation de graphes d'influences causales. L'usage de graphes orientés est totalement cohérent avec le traitement cérébral, autant physiologiquement que fonctionnellement. Le fonctionnement neuronal est un fonctionnement typiquement causal. En effet, l'influx créé par l'activation de neurones, va se propager unidirectionnellement le long de leurs axones et va finir par causer l'activation (ou l'inhibition) d'autres neurones.

Malgré cette adéquation très forte avec le fonctionnement cérébral, notre formalisme ne permet de représenter que les traitements automatiques du cerveau, c'est à dire uniquement les traitements dirigés par les stimuli. Cette limitation nous empêche donc de représenter les traitements cérébraux hautement cognitifs, comme les mécanismes de planification. Cependant, notre formalisme est assez souple pour tester différents types de propagation entre zones, comme, par exemple, le filtrage en entrée ou en sortie.

La principale limitation actuelle de notre formalisme reste l'absence de mécanismes d'apprentissage. Notre approche n'est pas moins adaptée à la modélisation des processus cognitifs mis en jeu lors des expériences d'activation. En effet, dans les protocoles des études d'activation actuelles, l'apprentissage de la tâche à accomplir est effectué avant le début des mesures. En effet pour obtenir une mesure de l'activité cérébrale, les techniques d'imagerie nécessitent une répétition de la tâche au cours d'un certain temps, durant lequel la mesure est réalisée. Le niveau d'expertise doit être constant durant l'acquisition de l'image, ce qui impose donc un apprentissage avant le début de la mesure. Ceci permet que la configuration du réseau d'aires mis en œuvre, pour résoudre la tâche, soit stable au cours de l'acquisition des images. Pour connaître l'effet de l'apprentissage il faut réaliser la mesure en trois temps : (a) mesure avant apprentissage, (b) apprentissage, (c) mesure après apprentissage.

Notre formalisme permet donc de représenter les réseaux mis en place par l'apprentissage, mais pas la mise en place de ces réseaux, elle-même. Compte tenu des contraintes décrites ci-dessus, notre formalisme reste adéquat à la modélisation de réseaux mis en jeu lors des études d'activation. Toutefois, un mécanisme d'apprentissage reste l'élément à acquérir pour augmenter la généralité du traitement de l'information dans nos modèles.

2. FORMALISME EXPLICITE

En définissant un formalisme original, nous nous sommes assurés que celui-ci répondrait point par point aux hypothèses fondamentales de notre domaine d'application. En effet, chacune des caractéristiques de notre travail peut se justifier vis-à-vis des caractéristiques du fonctionnement cérébral au niveau où on le représente (i.e. réseaux d'aires cérébrales). Il est important de souligner le nombre limité d'hypothèses fondamentales que nous avons retenues pour bâtir notre formalisme (elles sont exprimées ici en 6 points, dans le chapitre 4, partie 1). Ce nombre réduit assure une grande souplesse au formalisme. Moins il y a d'hypothèses fortes sous-entendues par le formalisme, plus on pourra exprimer d'hypothèses exploratoires différentes grâce à celui-ci.

Nous avons donc cherché à obtenir un formalisme neuro-fonctionnellement plausible. Mais nous recherchons aussi à construire une représentation explicite des hypothèses que l'on peut formuler à partir des diverses sources d'observation du fonctionnement cérébral. Une modélisation explicite permet une interprétation directe des hypothèses représentées, car elles sont clairement identifiées dans les modèles. Pour obtenir une représentation explicite, l'information manipulée doit être de

haut niveau sémantique. On se doit, de plus, de proposer une représentation très structurée, de manière à limiter la perte de transparence en fonction de l'augmentation de la complexité des modèles

2.1. Haut degré d'abstraction de l'information

Notre représentation de l'information est considérée comme une abstraction qualitative de l'information manipulée et transmise par les zones fonctionnelles. Nous la représentons par une variable à deux dimensions, numérique et symbolique, qui modélise l'amplitude et le type. Cette représentation, volontairement très éloignée d'une représentation neuronale de l'information, peut y trouver diverses interprétations, puisqu'elle peut être considérée comme l'abstraction du codage de l'information effectué par une population neuronale. Si elle a une réelle justification neurophysiologique (page 157), cette abstraction qualitative vise avant tout à fonder notre représentation sur une grande transparence sémantique.

L'amplitude reste une abstraction qualitative avant tout. Elle n'est pas une expression numérique directe de la quantité mesurée par telle ou telle technique (e.g. débit sanguin ou champ électromagnétique). De son côté, le type est aussi une abstraction qualitative, car il est très indéterminé et ne peut généralement pas être observé.

2.2. Représentation hiérarchique homogène

La base de notre approche est une représentation où deux niveaux d'organisation sont clairement identifiés. Ils sont le niveau structurel, matérialisant un réseau de zones interconnectées, et le niveau fonctionnel, décrivant le fonctionnement local de chacune des zones du niveau structurel. L'identification de ces deux niveaux de représentation est en adéquation avec une décomposition hiérarchique du système modélisé en *sous-systèmes*. Ce principe de modélisation est, par ailleurs, largement utilisé pour la modélisation de systèmes complexes (Okuda & Ushio 1991, Zeigler 1990). L'application de ce principe de décomposition peut être accrue en décomposant chacun de ces deux niveaux en un ensemble de sous-niveaux. On obtient alors une véritable hiérarchie de modèles telle que nous l'avons définie au chapitre 4.

Une organisation orientée composant originale nous permet d'obtenir des modèles individuellement moins complexes que le modèle total, tout en conservant une totale *homogénéité*

entre la représentation des niveaux structurels et celle des niveaux fonctionnels. On conserve ainsi une cohérence certaine à la description. La simulation du modèle total en est aussi fortement simplifiée, car tous les éléments utilisent les mêmes représentations et les mêmes opérateurs, quel que soit le niveau considéré. Cette organisation permet aussi de diminuer la complexité de la représentation de chaque niveau, par rapport à la complexité totale du modèle.

La modélisation causale s'allie parfaitement à ce style de représentation, grâce à la notion de composant. De plus, les modèles causaux, représentés par des graphes, permettent d'exprimer clairement les relations qui unissent les processeurs. La notion d'influence marginale, telle que nous la définissons, permet de représenter simplement un large éventail de relations avec un minimum de paramètres à définir. Les notions de gain dynamique et de loi de combinaison permettent d'*identifier* les processus complémentaires d'influence et de combinaison. Cette séparation contraint la description des relations mais elle contribue aussi à l'homogénéisation du formalisme, car toutes les relations sont bâties sur le même schéma.

2.3. Modularité et réutilisabilité des modèles

La principale conséquence d'une organisation hiérarchique homogène est qu'elle permet de construire des modèles modulaires. L'importance de la modularité dans nos modèles n'est pas réellement illustrée par les modélisations de petite envergure que nous avons présentées au chapitre précédent.

La principale caractéristique de la modularité est la segmentation des éléments d'un modèle. Chaque module est défini par un ensemble de relations, dites internes, qui lui sont propres et qui sont indépendantes des relations que ce module entretient avec le reste du modèle. Dans notre organisation, les modules sont représentés par les *composants*. Les relations internes, qui définissent le comportement du module, sont décrites dans le *modèle du composant*. Les liens entre les relations internes et les relations externes du module sont identifiés et formalisés par l'*interface* du modèle de composant. Donc, plus un modèle définira de composants, plus il sera modulaire. Cette propriété nous permet une réutilisation maximale des divers éléments d'un modèle. Tout modèle de composant peut être réutilisé dans différents modèles, ou même plusieurs fois dans un seul modèle, chaque instance étant considérée comme une copie indépendante du même modèle.

La modularité permet donc une grande réutilisabilité des modèles de composant, mais elle est accrue par la notion de généralité. Cette notion n'est pas, rappelons-le, réservée aux modèles

fonctionnels de zones, mais elle peut être appliquée à toute structure, ou toute primitive fonctionnelle, qui pourrait posséder un ensemble de caractéristiques non spécifiques, et donc réutilisables. A terme, cette grande réutilisabilité fait diminuer la durée de conception des modèles et facilite leur définition car, grâce à elle, nous sommes en mesure de proposer différentes classes de modèles génériques. Nous pouvons donc construire une bibliothèque de modèles facilement interfaçables et offrant ainsi une base évolutive à la définition de modèles plus complexes.

3. POSITION DE NOTRE APPROCHE

Comme nous l'avons vu au chapitre 2, il est possible de classer les modélisations cognitives computationnelles en deux grandes familles selon que les modélisations prennent en compte ou pas des contraintes biologiquement plausibles. Dans ce cadre, notre approche est une approche typiquement structurelle (Chapitre 2, partie 3) et située au niveau organisationnel global (page 50, Burnod 1991). Elle est toutefois innovante par le niveau d'abstraction où elle place sa représentation de l'information. Cette abstraction est rendue nécessaire par notre volonté d'obtenir une modélisation explicite.

3.1. Approche structurelle hybride

Une modélisation explicite peut conceptuellement s'opposer à une *modélisation émergente*, telle qu'une modélisation connexionniste fondée sur des réseaux de neurones formels. Contrairement à de tels modèles, notre approche laisse peu de place à l'auto-organisation et donc à l'émergence de comportement non représentés dans les modèles. Cependant, les interactions des hypothèses exprimées par les processeurs de nos modèles font quand même émerger un comportement global. Ce comportement n'est pas le produit d'une auto-organisation, mais il est le résultat direct des éléments fonctionnels modélisés. Cette approche nous permet de contrôler *expressément* le comportement du modèle sans passer par l'intermédiaire d'un apprentissage supervisé.

Malgré la différence du niveau d'abstraction, notre approche respecte, comme les réseaux de neurones formels, un certain nombre de contraintes propres au fonctionnement neuronal. Ainsi, l'aspect parallèle, distribué et non supervisé du traitement reste le grand point commun de ces deux approches. Ces contraintes, exprimées dans nos hypothèses fondamentales, limitent nécessairement le caractère explicite de notre représentation.

En considérant nos unités comme des processeurs fonctionnels de haut niveau, organisés en une représentation physiologiquement plausible, on peut considérer notre approche comme un compromis, à mi-chemin entre une modélisation fonctionnelle, de haute transparence sémantique, et une approche structurelle connexionniste. Notre modélisation peut être qualifiée d'approche *hybride*, mais uniquement dans le sens où notre formalisme n'est ni totalement connexionniste ni totalement *symbolique*. Notre approche est cependant très différente de modélisations comme celle de Wallace *et al.* (1993) où cohabitent des réseaux de neurones et des modules symboliques.

3.2. Simulation qualitative

Le respect de la plausibilité neuro-fonctionnelle dans notre formalisme a conduit à un accroissement non négligeable de la complexité du traitement de l'information qualitative. La manifestation directe de cet accroissement est la représentation de l'information, par un couple, qui conduit à un dédoublement de tous les calculs de propagation. Nous avons, de plus, transporté un raisonnement qualitatif dans un type de modélisation assez peu exploré par cette discipline, la modélisation biologiquement plausible du fonctionnement du cerveau.

Dans le cas plus particuliers des modèles causaux qualitatifs, très répandus, nous proposons une représentation originale. Pour schématiser, les modèles causaux décrivent généralement des relations entre des variables d'états. Ces relations sont d'ailleurs souvent obtenues à partir d'équations numériques, enrichies d'une relation orientée, décrivant le sens de la causalité. Dans notre approche, nous décrivons des relations entre processeurs qui traitent tous la même grandeur. Et les relations causales liant ces processeurs ne peuvent être obtenues à partir de modèles numériques de leur comportement. Nous nous situons donc dans une approche purement causale et nous nous contentons de relations exprimant directement une causalité pour décrire l'évolution du système. Pour y parvenir, nous avons donc porté un intérêt tout particulier aux mécanismes de la propagation qualitative et défini un type d'influence généralisée.

Outre la notion de gain dynamique, nos influences sont caractérisées par des conditions d'activation qualitatives. Ces dernières permettent de définir les différents états de fonctionnement du modèle (i.e. pour chaque influence, un état actif et un état inactif). Elles utilisent un système d'évaluation multivalué qui permet de limiter le problème d'indécision produit par le changement d'état d'une l'influence. En effet, l'utilisation de valeurs qualitatives peut, dans certain cas, conduire à une indétermination de l'état de l'influence.

Ce problème d'indécision est courant dans les méthodes de simulations qualitatives. Par exemple, une solution, proposée par la méthode QSIM, consiste à créer un *arbre de comportements* (Travé-Massuyès *et al.* 1997). Lorsque plusieurs états sont possibles au même moment, on crée autant de branches distinctes, chacune représentant un état du système. On calcule ensuite l'état du système pour chaque branche en parallèle. On est ainsi assuré d'obtenir une prédiction de tous les états possibles à la fin de la simulation. Ce calcul exhaustif alourdit la simulation, surtout dans des systèmes où l'on réalise de nombreux calculs, ou bien, dans des systèmes où il y a de nombreux changements d'états. Il existe toutefois diverses méthodes pour limiter l'expansion de l'arbre de comportements (e.g. supprimer les états redondants).

Dans notre cas, nous ne conservons qu'un seul état à chaque pas de calcul. Ce choix tient au fait que l'on peut supposer que l'indécision, entre l'état actif et l'état inactif d'une influence, a une durée limitée. On peut alors considérer ce laps de temps comme une transition entre ces deux états (i.e. un troisième «état»). Il faut, dans ce cas, définir la représentation de cette transition. La représentation la plus simple, consiste à représenter la transition comme la réunion des deux états qu'elle sépare. Ce qui revient, dans notre cas, à effectuer l'union des résultats des calculs correspondant à l'état actif et à l'état inactif de l'influence. Ce n'est qu'ainsi que l'on est assuré que la solution effective contient toutes les solutions possibles (i.e. complétude des résultats). Cependant, ce mécanisme conduit à une augmentation excessive de l'imprécision des valeurs. C'est pourquoi nous avons opté pour un compromis entre complétude des calculs et contrôle de l'imprécision. Ce compromis est incarné par l'utilisation de relations multivaluées qui permettent de représenter l'état de transition comme une valeur intermédiaire entre les deux états d'activation.

4. PERSPECTIVES

Les perspectives d'évolution de ce travail sont multiples. Elles concernent autant l'application du formalisme que son développement propre. La voie générale de cette évolution est la prise en compte de processus cérébraux toujours plus complexes. Cette voie est caractérisée par deux perspectives parallèles mais complémentaires. D'une part, nous devons augmenter le nombre des propositions de modèles d'hypothèses (e.g. primitives fonctionnelles, modèles génériques de zones, etc.), ceci dans le but de bâtir un éventail varié de fonctionnalités plus intégrées. Ces éléments pourront être directement utilisables dans des modèles de plus grande envergure. D'autre part, ce sont les capacités de représentation du formalisme qui doivent être augmentées.

BIOCAEN et le formalisme qu'il implémente se sont limités à la représentation de processus cognitifs automatiques. Ces processus, qui forment une grande partie de l'activité du cerveau (Pastor 1999), peuvent être représentés par des mécanismes dirigés par les stimuli et dont la modélisation pouvait être inspirée de processus automatique industriels (Lacotte 1996). L'inspiration des techniques de l'ingénierie automatique nous a fourni un support solide pour l'élaboration de notre formalisme, puis du simulateur BIOCAEN.

Nous avons donc laissé de côté les mécanismes de contrôle et de planification, qui sont des processus plus cognitifs. La représentation de ce genre de processus représente un cadre plus vaste et une approche différente de celle qui nous a permis de construire BIOCAEN. Cette évolution s'accompagnera certainement de l'ajout de nouvelles hypothèses fondamentales, plus spécifiques des processus cognitifs de haut niveau que celles que nous avons formulées et qui, de plus, doivent être préalablement validées. Le changement des fondements du formalisme devront entraîner de grandes modifications dans sa structure, ce qui nécessitera une refonte complète des concepts que nous avons définis. Cependant, dans sa version actuelle, notre travail peut fournir une base pour l'expression des futures hypothèses fondamentales. Prenons, par exemple, le cas de l'hypothèse du filtrage de l'information en entrée des zones (voir page 144, Pastor *et al.* 2000, Pastor 1999). Cette hypothèse est exprimée dans les modèles que nous avons présentés ici. Nous la considérons, dans BIOCAEN, comme une hypothèse exploratoire. Or, elle apparaît, de plus en plus, comme fondamentale. Elle devrait donc être incluse dans le formalisme, ce qui aurait des répercussions sur le mode de propagation, actuellement incarné dans BIOCAEN par la notion d'influence marginale. D'une manière générale, la validation d'hypothèses fondamentales pour un formalisme demande un véritable travail de fond sur une délimitation de la frontière entre exploratoire et fondamental.

Sur la voie de cette perspective générale, plusieurs développements théoriques sont d'ores et déjà envisageables. Nous les mettons en exergue dans ce qui suit.

4.1. Contrôle de l'accroissement de l'imprécision

Le contrôle de l'accroissement de l'imprécision est un problème récurrent dans les formalismes causaux qui représentent des informations incomplètes (Travé-Massuyès *et al.* 1997). D'une manière générale, la représentation de connaissances imprécises conduit à des inférences plus faibles. Dans le cas des modèles causaux, l'effet est accentué par la réinjection de l'information dans le graphe causal (i.e. l'existence de cycles causaux, ou la modélisation de systèmes dynamiques). Même si les intervalles restent un bon modèle de la représentation informatique des nombres réels

(Tupper 1996), l'algèbre classique des intervalles pose un certain nombre de problèmes dès que l'imprécision des données devient importante (SIGLA/X group 1998, Struss 1990). Les lacunes de cette algèbre se traduisent par un accroissement excessif de l'imprécision, c'est à dire de la longueur des intervalles, au cours de la propagation de l'information. L'exemple le plus représentatif peut être illustré par la résolution de l'équation suivante :

$$[0, 1] + X = [0, 1] \quad \text{d'ou} \quad X = [0, 1] - [0, 1] = [-1, 1] \neq 0$$

Bien qu'il contienne le résultat réel de l'équation (i.e. $0 \in [-1, 1]$), le résultat de la soustraction est bien plus large. Ce problème est latent dans la définition même des opérations sur les intervalles. Les résultats de ces dernières doivent contenir *toutes* valeurs possibles de la même opération appliquée à tous les réels contenus dans les opérandes.

Dans un modèle causal où l'information est réinjectée dans le système, l'imprécision a tendance à augmenter au cours de la simulation, même si l'imprécision des entrées du système n'augmente pas. Sur ce plan-là, l'algèbre des intervalles telle que nous l'utilisons, c'est-à-dire définie sur la base d'intervalles fermés, est très laxiste. L'utilisation d'intervalles ouverts (Struss 1990) ou d'hypperréels (Tupper 1996) serait déjà un premier pas vers une représentation plus contrôlée. Leur utilisation supprimerait un certain nombre de particularités aux cas limites, notamment dans le cas de l'égalité qualitative au sens où nous l'avons définie (e.g. $[0, 10] \approx [10, 100]$).

Pour contrôler l'accroissement de l'imprécision, il est nécessaire d'exercer des contraintes sur sa représentation. Or, dans un raisonnement qualitatif, la représentation de cette imprécision n'est pas directement accessible. Nous avons tenté de la prendre en compte en utilisant une logique multivaluée pour l'évaluation des conditions d'activation. Cependant, pour maîtriser au mieux cet accroissement, la solution passe inmanquablement par un enrichissement de la connaissance représentée par l'intervalle. On peut, par exemple, citer deux moyens de contrôle qui vont dans ce sens.

Le premier consiste à utiliser des intervalles flous (que nous avons présenté au chapitre 3, page 64) pour représenter les valeurs numériques. La fonction d'appartenance floue diminue cet accroissement, car elle représente *explicitement* l'incertitude engendrée par une valeur imprécise. On réalise ici un enrichissement de la représentation qui consiste à spécifier la fonction d'appartenance propre à chaque intervalle.

Le second moyen repose sur la notion d'intervalle modal (Armengol *et al.* 1998, SIGLA/X group 1998). L'application de cette notion va plus loin que l'utilisation d'intervalles flous dans le sens où elle s'accompagne d'une modification de la sémantique représentée par la notion d'intervalle. Le concept d'intervalle modal est assez simple à formaliser, il consiste à associer à l'intervalle un quantificateur (i.e. une modalité). L'intervalle modal $([1, 2], \forall)$ représente la proposition « quel que soit le réel x tel que $x \in [1, 2]$ », on le représentera par $[1, 2]^M$, alors que l'intervalle modal $([1, 2], \exists)$ représente la proposition « il existe un réel x tel que $x \in [1, 2]$ », on le représentera $[2, 1]^M$, il est appelé intervalle *dual* de $[1, 2]^M$. Pour faire court, on peut donc représenter la proposition, pour a et $b \in \mathbb{R}$, tels que $a < b$: $\forall x \in [a, b], \exists y \in [-b, -a]$ tel que $x + y = 0$, par notation :

$$[a, b]^M + [-a, -b]^M = 0.$$

Ceci fait de zéro l'élément neutre pour l'addition des intervalles modaux. Et, ainsi, la solution de l'équation :

$$[0, 1]^M + X = [0, 1]^M$$

devient :

$$X = [0, 1]^M - [0, 1]^M = 0.$$

Cette représentation change totalement la notion d'intervalle telle que nous la concevons normalement. Un changement de la sémantique associée à l'intervalle modifie radicalement le raisonnement que l'on peut réaliser à partir de cette représentation.

4.2. Inférence distribuée sur les types

L'une des caractéristiques principales de la manipulation de la catégorie de l'information dans le cerveau humain est la faculté de génération de nouvelles catégories. S'il existe actuellement des opérateurs d'union et de combinaison des types dans BIOCAEN, nous ne pouvons pas réellement générer une nouveaux types. De plus, les opérations sur les types ont tendance à faire globalement décroître les poids des symboles manipulés. Nous avons aussi tendance à appliquer, par défaut, les mêmes opérations à l'amplitude et au type de l'information, alors que ces deux notions représentent deux dimensions totalement différentes. Le mimétisme des opérations est un reliquat des premières versions de notre formalisme, et il ne peut être correctement justifié du point de vue cognitif. Toutes ces limitations représentent un gros handicap pour la représentation d'opérations cognitives plus élevées.

Pour pouvoir représenter des processus cérébraux complexes, il apparaît indispensable de se munir d'un système de génération de nouvelles catégories. L'élaboration de ce système passe par la mise en œuvre de mécanismes d'inférence distribués appliqués au type de l'information propagée. Cette évolution peut être, par exemple, guidée par divers travaux de modélisations cognitives fonctionnelles (i.e. ACT, SOAR, page 31) ou trouver une inspiration auprès de techniques de l'intelligence artificielle (i.e. logique floue, apprentissage automatique probabiliste, algorithmes génétiques).

4.3. Architecture hiérarchique généralisée

Nous avons utilisé une architecture hiérarchique pour la définition de nos modèles. Nous avons vu que ce type d'architecture était tout à fait indiqué pour représenter des systèmes complexes, mais il est possible d'en tirer un profit plus grand encore que l'utilisation que nous en faisons aujourd'hui. Une architecture modulaire, comme la nôtre, peut facilement servir d'interface entre divers formalismes. Une application directe de cette possibilité serait de pouvoir utiliser des modèles de zones exprimés dans des formalismes hétérogènes. On pourrait, par exemple, utiliser un modèle connexionniste pour telle zone et un modèle causal pour une autre à l'intérieur d'un même modèle structurel. On pourrait alors tirer directement profit de modélisations existantes. Ou bien, on pourrait appliquer, à chaque niveau et à chaque localisation, un formalisme, au choix, plus ou moins biologiquement plausible, suivant l'état d'avancement des connaissances biologiques. Cette perspective élargit les possibilités de modélisations réutilisables.

On réaliserait de cette manière de véritables modèles hybrides (i.e. au sens de Wallace, page 55). Il faudrait pour cela définir des *interfaces de connexion* entre des formalismes différents. Cette opération demande la définition de méthodes et d'opérations de conversion de la représentation de l'information d'un formalisme à l'autre, avec tous les problèmes que cela sous-entend (i.e. niveaux d'abstraction différents).

Cette optique d'évolution est tout à fait envisageable à partir de la notion actuelle de hiérarchie de composants, mais elle supprimerait notre effort de conservation d'homogénéité de la représentation. Elle demande donc une révision complète des concepts que nous avons définis, pour en élaborer une version généralisée, même si la structure de base (i.e. les graphes causaux) semble très appropriée à cette évolution. Pour avancer dans ce sens, il semble donc très intéressant de s'intéresser davantage encore à la notion de causalité pour revenir à une formalisation généralisée de celle-ci (Pearl 1999).

Avant même de s'intéresser à une modélisation hybride, la représentation de causalités plus évoluées est motivée par la représentation des processus complexes. On peut, par exemple, penser à une généralisation du concept d'influence binaire à celui d'influence n^{aires} qui permettrait peut être d'exprimer un contrôle dynamique sur la propagation de l'information.

5. CONCLUSION

Nous avons montré qu'il était possible, sur la base d'une représentation qualitative de l'information et d'une architecture causale, de proposer un formalisme apte à la représentation explicite d'hypothèse sur les processus cognitifs contraints par leur substrat cérébral. En nous limitant à la représentation de processus automatiques, nous avons pu construire un simulateur permettant de tester ces hypothèses en les confrontant aux observations obtenues par des techniques d'imagerie fonctionnelle.

La dernière perspective d'évolution de nos travaux que nous avons citée permet d'apercevoir BIOCAEN comme le premier l'élément d'un *méta-simulateur* de la propagation de l'information cérébrale au travers des réseaux de zones fonctionnelles. Un système plus vaste, qui prendra en compte les processus que BIOCAEN ne représente pas, utilisera des fondements plus spécifiques que les hypothèses qui ont été utilisées pour bâtir ce premier simulateur.

Cette perception ne doit cependant pas occulter l'esprit d'ouverture qui caractérise le formalisme que nous avons défini. Nous avons toujours, dans notre démarche, gardé un regard critique qui nous a permis de spécifier chaque élément qui compose notre système de manière à ce qu'il soit, à la fois, cohérent avec nos objectifs et le moins limitant possible pour une évolution du système. C'est ainsi que l'on peut penser à utiliser un type semblable d'architecture pour une représentation plus développée.

Tout comme les concepts que nous avons définis peuvent être réutilisés et adaptés à la définition d'un outil plus sophistiqué et destiné à la même application, on peut aussi s'intéresser à l'application de notre formalisme à d'autres domaines. Même si les applications directes de notre approche semblent limitées, la réutilisation de la notion de composant, par exemple, est un atout important pour la modélisation de tout système complexe, biologique ou même physique. Ces applications diversifiées pourront peut-être en retour apporter de nouveaux concepts au formalisme de BIOCAEN.

BIBLIOGRAPHIE

- Alary F., Doyon B., Loubinoux I., Carel C., Boulanouar K., Ranjeva J.P., Celsis P. & Chollet F., 1998, Event-Related potentials elicited by passive movements in humans: Characterization, source analysis, and comparison to fMRI, *NeuroImage*, 8:377-390.
- Alexander G.E., DeLong M.R. & Crutcher M.D., 1992, Do cortical and basal ganglionic motor areas use "motor programs" to control movement?, *Behavioral and Brain Sciences*, 15:656-665.
- Alliot J.M. & Schiex T., 1994, *Intelligence Artificielle & Informatique Théorique*, CÉpaduès Editions.
- Anderson J.R., 1983, A Spreading Activation Theory of Memory, *Journal of Verbal Learning and Verbal Behavior*, 22:261-295.
- Anderson J.R., 1989, A Theory of the Origins of Human Knowledge, *Artificial Intelligence*, 40(3-4):313-351.
- Andres V., Dubois D., Lang J. & Prade H., 1990, Logique possibiliste et logique floue, *Modèles logiques et systèmes d'intelligence artificielle*, Luisa Iturrioz & Alain Dussauchoy, Hermès, 163-181.
- Arbib M.A., 1985, Brain theory and cooperative computation, *Human Neurobiology*, 4:201-218.
- Arbib M.A., Érdi P. & Szentágothai, 1998, *Neural organization. Structure, Function and Dynamics*, A Bradford Book, MIT Press, Cambridge (Massachusetts, USA).
- Ardila A., 1993, Toward a model of phoneme perception, *International Journal of Neuroscience*, 70:1-12.
- Amengol J., Travé-Massuyès L., Vehi J. & Lluís De La Rosa J., 1998, On modal interval analysis for envelope determination within the Ca-En qualitative simulator. *Proceeding of 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'98)*, Paris, France, 110-117.
- Arnold A. & Guessarian I., 1994, *Mathématiques pour l'informatique*, 2^e édition, Masson.
- Bellazzi R., Guglielmann R. & Ironi L., 1999, A qualitative-fuzzy framework for nonlinear black-box system identification, *Proceedings of the sixteenth International Joint Conference on Artificial Intelligence*, 2:1041-1046.
- Bousson K. & Travé-Massuyès L., 1993, Fuzzy Causal Simulation in Process Engineering, *Proceedings of the 13th International Joint Conferences on Artificial Intelligence*, Chambéry (France), 1536-1541.
- Bousson K., 1993, *Raisonnement causal pour la supervision de processus basée sur des modèles*, Thèse de doctorat, Institut National des Sciences Appliquées de Toulouse.
- Bousson K. & Travé-Massuyès L., 1994, Putting more numbers in the qualitative simulator Ca-En, *Proceedings of the Second International Conference on Intelligent Systems Engineering*, Hambourg (Allemagne), 62-69.
- Bousson K., Zimmer L. & Travé-Massuyès L., 1994, Causal model-based diagnosis of dynamic systems, *Proceedings of the 5th International workshop on principles of diagnosis*, New Paltz (USA).
- Brown T.H., Kairiss E.W. & Keenan C.L., 1990, Hebbian Synapses: biophysical mechanisms and algorithms, *Annual Review of Neuroscience*, 13:475-511.
- Bumod Y., 1991, Organizational levels of the cerebral cortex: an integrated model, *Acta Biotheoretica*, 39:351-361.
- Carpenter G. & Grossberg S., 1993, Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions, *Trends in Neuroscience*, 16(4):131-137.
- Churchland P.S. & Sejnowski T.J., 1992, *The computational brain*, MIT Press, Cambridge (Massachusetts, USA).
- Clark A., 1989, *Microcognition*, MIT Press, Cambridge, MA, USA.
- Crick F., 1989, The recent excitement about neural networks, *Nature*, 337:129-132.
- Dayan P., 1994, Computational modeling, *Current Opinion in Neurobiology*, 4:212-217.
- Dehaene S. & Changeux J.-P., 1997, A hierarchical neuronal network for planning behavior, *Proceedings of the National Academy of Sciences*, 94:13293-13298.
- Dehaene S., Kerszberg M. & Changeux J.-P., 1998, A neuronal model of a global workspace in effortful cognitive tasks, *Proceedings of the National Academy of Sciences*, 95:14529-14534.

- Delacour J., 1995, *Le cerveau et l'esprit*, Presses Universitaires de France, « Que sais-je ? », n° 2938.
- Démonet J.F., Chollet F., Ramsay S., Cardebat D., Nespoulous J.L., Wise R., Rascol A. & Frackowiak R., 1992, *The Anatomy of Phonological and Semantic Processing in Normal Subjects*, *Brain*, 115:1753-1768.
- Dubois D. & Prade H., 1987, *Théorie des possibilités - Applications à la Représentation des Connaissances en Informatique*, M ASSON, 2^{ème} édition.
- Dubois D. & Prade H., 1989, *Order-of-magnitude reasoning with fuzzy relations*, *Revue d'Intelligence Artificielle*, 3(4):69-94.
- Dubois D., Prade H. & Smets P., 1994, *Partial Truth is not Uncertainty*, *Fuzzy Logic versus Possibilistic Logic*, *IEEE Expert*, 9(4):15-18.
- Evans A.C., 1996, *La neuroanatomie entre dans l'âge informatique*, *La Recherche*, 289:42-45.
- Fischbach G., 1992, *Le cerveau et la pensée*, *Pour la science*, 181:28-38.
- Foulloy L., 1993, *Qualitative control and fuzzy control: towards a writing methodology*, *AI Communications*, 6(3-4):147-154.
- Fox P.T. & Raichle M.E., 1984, *Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by Positron Emission Tomography*, *Journal of Neurophysiology*, 51:1109-1120.
- Fox P.T. & Raichle M.E., 1985, *Stimulus rate determines regional brain blood flow in striate cortex*, *Annals of Neurology*, 17(3):303-305.
- Fox P.T. & Lancaster L., 1996, *Un atlas du cerveau sur Internet*, *La Recherche*, 289:49-51.
- Frezza-Buet H. & Alexandre F., 1998, *Multimodal sequence learning with a cortically-inspired model*, 3rd International conference on computational intelligence and neuroscience, Research Triangle Park (North Carolina, USA), 2:21-23.
- Guigon E., Grandguillaume P., Otto I., Boutkhal L. & Burnod Y., 1994, *Neural network models of cortical functions based on the computational properties of the cerebral cortex*, *Journal of Physiology (Paris)*, 88:291-308.
- Hayes P.J., 1985, *Naïve physics I: Ontology for liquids*, J. R. Hobbs & R. C. Moore (eds.), *Formal Theories of the Common Sense World*, Ablex Publishing Corporation, 71-107.
- Hedaen H., 1972, *Introduction à la neuropsychologie*, Larousse Université.
- Hickok G. & Poeppel D., 2000, *Towards a functional neuroanatomy of speech perception*, *Trends in Cognitive Sciences*, 4(4):131-138.
- Hinton G.E., Plaut D. & Shallice T., 1993, *La simulation des lésions cérébrales*, *Pour la science*, 194:36-43.
- Hinton G.E., Dayan P., Frey B.J. & Neal R.M., 1995, *The «wake-sleep» algorithm for unsupervised neural networks*, *Science*, 268:1158-1161.
- Hunter L., Krichmar J.L. & Olds J.L., 1997, *Qualitative reasoning as a modeling tool for computational neuroscience*, *Qualitative Reasoning*, 11th International Workshop, Cortona (Italie), 265-276.
- Iturriz L., 1990, *Logiques multivaluées, Un outils souple et pluriel*, *Modèles logiques et systèmes d'intelligence artificielle*, Luisa Iturriz & Alain Dussachoy, Hermès, 113-138.
- Iwasaki Y., 1997, *Real-world applications of Qualitative Reasoning*, *IEEE Expert*, 12(3):16-21.
- Joliot M., Crivello F., Badier J.M., Diallo B., Tzourio N. & Mazoyer B., 1998, *Anatomical congruence of metabolic and electromagnetic activation signals during a self-paced motor task: A combined PET-MEG study*, *NeuroImage*, 7:337-351.
- De Kleer J. & Brown J.S., 1983, *Assumptions and ambiguities in mechanistic mental models*, D. Gentner & A.L. Stevens (eds.), *Mental Models*, L. Erlbaum Publishers, 159-190.
- Koch C., 1993, *Computational approaches to cognition: the bottom-up view*, *Current Opinion in Neurobiology*, 3:203-208.
- Koch P. & Leisman G., 1996, *Wave theory of large-scale organization of cortical activity*, *International Journal of Neuroscience*, 86:179-196.
- Koller F. & Pfeffer A., 1997, *Object-Oriented Bayesian Networks*, *Proceedings of the thirteenth Annual Conference on Uncertainty in Artificial Intelligence*, Providence (Rhode Island, USA), 302-313.
- Kosko B. & Isaka S., 1993, *La logique floue*, *Pour la science*, 191:62-68.
- Kuipers B.J., 1984, *Commonsense Reasoning about Causality: Deriving behavior from structure*, *Artificial Intelligence*, 24:169-204.
- Kuipers B.J., 1993, *Reasoning with Qualitative Models*, *Artificial Intelligence*, 59:125-132.
- Kuipers B.J., 1996, *A Hierarchy of qualitative representations for space*, *Qualitative Reasoning*,

- 10th International Workshop, Fallen Leaf Lake (California, USA).
- Lacotte B., 1996, Modélisation par réseaux causaux qualitatifs de la structure fonctionnelle du cerveau humain, Mémoire de DEA. R.C.F.R., Université Paul Sabatier.
- Lafon M., Travé-Massuyès L. & Pastor J., 1997, BIOCAEN: A causal model simulator for cerebral dynamics, Proceedings of Automation 2001, Vienne (Autriche), 61-70.
- Lafon M., Pastor J., Travé-Massuyès L., Doyon B., Démonet J-F. & Celsis P., 1998, Qualitative modeling of cerebral information propagation mechanisms, Proceedings of the 3rd International conference on computational intelligence and neuroscience, JCIS, Research Triangle Park (Caroline du nord, USA), 2:21-23.
- Lafon M., 1998, BIOCAEN: simulateur d'hypothèses sur la propagation de l'information cérébrale, Actes des Rencontres Jeunes Chercheurs en Intelligence Artificielle, Toulouse (France), 139-142.
- Lafon M., Travé-Massuyès L. & Pastor J., 1999, Hierarchical causal modeling of cerebral information propagation mechanisms, Proceedings of Qualitative and Model Based Reasoning for Complex Systems and their Control, IJCAI, Stockholm (Suède), 26-32.
- Lafon M., 1999, Manuel d'utilisation de BIOCAEN, Version 1.1, INSERM U455 et LAAS/CNRS Groupe AS.
- Laird J.E., Rosenbloom P. S. & Newell A., 1986, Chunking in SOAR: The anatomy of a general learning mechanism, Machine Learning, 1:11-46.
- Laplante D., 1985, La querelle des localisations cérébrales: débat scientifique ou idéologico-politique?, Science numéro un, 3:24-36.
- Laplante D., 1994, Réflexions épistémologiques sur la question de l'organisation cérébrale, Revue de Neurologie (Paris), 150(8-9):555-563.
- Leyval L., Gentil S. & Feray-Beaumont S., 1994, Model-based causal reasoning for process supervision. Automatica, 30(8):1295-1306.
- Litvan I., DeLeo J.M., Haw J.J., Daniel S.E., Jellinger K., McKee A., Dickson D., Horoupian D.S., Lantos P.L. & Tabaton M., 1996, What can artificial neural networks teach us about neurodegenerative disorders with extrapyramidal features?, Brain, 119:831-839.
- Lukasiewicz J., 1970, Selected works, Edition L. Borkowski, Studies in Logic, Amsterdam, Hollande.
- Malhotra A. & Giannakopoulos F., 1996, Population networks: a large-scale framework for modeling cortical neural networks, Biological Cybernetics, 75:441-452.
- Mavrouniotis M.L. & Stephanopoulos G., 1987, Reasoning with orders of magnitude and approximate relations, Proceedings 6th National Conference on Artificial Intelligence (AAAI 1987), Seattle, WA, 626-630.
- Mavrouniotis M.L. & Stephanopoulos G., 1988, Formal order of magnitude reasoning in process engineering, Computer Chemical Engineering, 12:867-880.
- Mazoyer B. & Belliveau J.W., 1996, Les nouveaux progrès de l'imagerie, La Recherche, Juillet-août 289:26-33.
- McClelland J.L. & Rumelhart D.E., 1981, An Interactive Activation Model of Context Effects in Letter Perception, Part I: an Account of Basic Findings, Psychological Review, 88(5):375-407.
- McClelland J.L., Rumelhart D.E. & Hinton G.E., 1986, The appeal of parallel distributed processing, Parallel distributed processing, D.E. Rumelhart, J.L. McClelland and the PDP Research Group Edition, MIT Press, 1:3-40.
- McClelland J.L. & Elman J.L., 1986, The TRACE model of speech perception, Cognitive Psychology, 18:1-86.
- McClelland J.L. & Plaut D.C., 1993, Computational approaches to cognition: Top-down approaches, Current Opinion in Neurobiology, 3:209-216.
- McClelland J.L., 1994, The organization of memory: a parallel distributed processing perspective, Revue de Neurologie (Paris), 150(8-9):570-579.
- Meibaw., 1993, Synaptic integration in an excitable dendritic tree, Journal of Neurophysiology, 70(3):1086-1101.
- Mentis M.J., Alexander G.E., Grady C.L., Horowitz B., Krasuski J., Pietrini P., Stassburger T., Hampel H., Schapiro M.B. & Rapoport S.I., 1997, Frequency variation of a pattern-flash visual stimulus during PET differentially activates brain from striate through frontal cortex, Neuroimage, 5:116-128.
- Mesulam M.M., 1990, Large-Scale Neurocognitive Networks and Distributed Processing for Attention, Language and memory, Annals of Neurology, 28(5):597-613.
- Milne R., Nicol C., Travé-Massuyès L. & Quevedo J., 1996, TIGER: knowledge based gas turbine condition monitoring, AICommunications, 9:92-108.
- Mitchell I.J., Brothie J.M., Brown G. & Crossman A., 1991, Modeling the Functional Organization

- of the Basal Ganglia. A Parallel Distributed Processing Approach, *Movement Disorders*, 6(3):189-204.
- Monchi O. & Taylor JG., 1998, Extending an anatomically based model of working memory to allow for a realistic set-up of delayed response tasks and investigate learning, 3rd International conference on computational intelligence and neuroscience, Research Triangle Park (North Caroline, USA), 2:36-39.
- MQ&D group, 1996, Qualitative Reasoning: A Survey of Techniques and Applications, Special Issue of *AICommunications*, 8(3/4):119-192.
- Newell A. & Simon H A., 1963, GPS: A Program that Simulates Human Thought, *Computer and Thought*, E.A. Feigenbaum & J. Feldman (Eds), R.O Klenbourg KG., 279-293.
- Newell A., Rosenbloom P.S. & Laird J.E., 1989, Symbolic Architectures for cognition, *Foundations of Cognitive Science*, Posner Editor, The MIT Press, 93-131.
- Ohayon M., 1990, Processus cognitifs et réseaux neuronaux, *Anales médico-psychologiques*, 148(8):669-695.
- Okuda K. & Ushio T., 1991, Hierarchical Qualitative Simulation for Large Scale Dynamical Systems, *Applications of Artificial Intelligence in Engineering VI*, Oxford (Grande Bretagne), 301-317.
- Omar R., 1994, Artificial Intelligence through Logic?, *AICommunications*, 7(3-4):161-174.
- Pal S. & Konar A., 1996, Cognitive reasoning using fuzzy neural nets, *IEEE Transactions on Systems, Man and Cybernetics*, 26(4B):616-619.
- Parks R.W., Long D.L., Levine D.S., Crockett D.J., McGeer E.G., McGeer P.L., Dalton I.E., Zec R.F., Becker R.E., Coburn K.L., Siler G., Nelson M.E. & Bower J.M., 1991, Parallel distributed processing and neural networks: origins, methodology and cognitive functions, *International Journal of Neuroscience*, 60:195-214.
- Parsons S., 1991, Interval Algebra and Order of Magnitude Reasoning, *Applications of Artificial Intelligence in Engineering VI*, Oxford (Grande Bretagne), 945-961.
- Pastor J., Lafon M., Travé-Massuyès L., Démonet J.F., Doyon B. & Celsis P., 2000, Information processing in large-scale cerebral networks: the causal connectivity approach, *Biological Cybernetics*, 82(1):49-59.
- Pastor J., 1999, Raisonement et mécanismes cérébraux sous-jacents, approche par l'intelligence artificielle, Habilitation à diriger les Recherches, Université de Rennes I.
- Pastor J., Travé-Massuyès L., Démonet J.F., Doyon B. & Celsis P., 1997, BIOCAEN: A causal qualitative network for cerebral information propagation modeling, *Proceedings of 11th International Workshop on Qualitative Reasoning*, Cortona (Italie), 305-315.
- Pearl J., 1995, Bayesian Networks, *Handbook of Brain Theory and Neural Networks*, M. Arbib (Ed.), MIT Press, 149-153.
- Pearl J., 1999, Reasoning with cause and effect, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, Stockholm (Suède), 2:1437-1449.
- Pedrycz W., Chi-Fung Lam P. & Rocha A.F., 1995, Distributed fuzzy system modeling, *IEEE Transactions on Systems, Man and Cybernetics*, 25(5):769-780.
- Pernier J., Perrin F. & Bertrand O., 1992, Identification des générateurs de l'activité électrique cérébrale, *Le courrier du CNRS*, 79.
- Philippon J., 1988, La neurochirurgie, *Presses Universitaires de France*, « Que sais-je? », n° 1369.
- Pons R., Travé-Massuyès L. & Porcheron M., 1999, Model Based Diagnosis and Maintenance of Time-varying Dynamic Systems, *Proceedings of the tenth international workshop of diagnosis (DX'99)*, Loch Awe (Ecosse), 211-219.
- Posner M.I., Petersen S.E., Fox P.T. & Raichle M.E., 1988, Localization of cognitive operations in the human brain, *Science*, 240:1627-1631.
- Prats F. & Travé-Massuyès L., 1999, Influence combination polynomial admissible operators. LEA-SICA, Rapport de recherche UPC, Barcelone, Espagne.
- Pulvermüller F., 1996, Hebb's concept of cell assemblies and the psychophysiology of word processing, *Psychophysiology*, 33:317-333.
- Pylyshyn Z.W., 1989, Computing in cognitive science, *Foundations of Cognitive Science*, Posner Editor, The MIT Press, 133-159.
- Reeke Jr G.N. & Sporns O., 1993, Behaviorally based modeling and computational approaches to neuroscience, *Annual Review of Neurosciences*, 16:597-623.
- Rosch E., 1978, Principles of categorization, *Cognition and Categorization*, E. Rosch & B.B. Lloyd (Eds.), L. Erlbaum Publishers, 27-48.
- Rosenzweig M. & Leiman A.L., 1991, *Neurophysiologie*, deuxième édition, Inter-Edition.

- Rumelhart D E., Hinton G E. & Williams R J., 1986, Learning representations by back-propagating errors, *Nature*, 323:533-536.
- Rumelhart D E., 1989, The architecture of mind : a connectionist approach, *Foundations of Cognitive Science*, Posner Editor, The MIT Press, 133-159.
- Sakai K., Sajda P., Yen S C. & Finkel L H., 1997, Coarse-grain parallel computing for very large scale neural simulations in the NEXUS simulation environment, *Computer in Biology and Medicine*, 27(4):257-266.
- Schank R C. & Abelson R P., 1977, *Scripts, Plans, Goals and Understanding*, L. Erlbaum Publishers, chapitres 1-3, 1-68.
- Schank R C. & Farrell R G., 1988, *Memory, Understanding Cognitive Science*, M F. M cTear Ed., 120-133.
- Schutter E D., 1992, A consumer guide to neuronal modeling software, *Trends in Neuroscience*, 15(11):414-421.
- Schneider W., 1993, Varieties of working memory as seen in biology and in connectionist/control architectures, *Memory & Cognition*, 21(2):184-192.
- Segev I., 1992, Single neuron models: oversimple, complex and reduced, *Trends in Neuroscience*, 15(11):414-421.
- Sergent J., 1994, Brain-imaging studies of cognitive functions, *Trends in Neuroscience*, 17(6):221-227.
- SIGLAX group, 1998, *Modal Intervals*, draft for a basic tutorial, Rapport de recherche IM A 98-07-RR, Université de Girona, Espagne.
- Struss P., 1990, Problems of interval-Based Qualitative Reasoning, *Readings in Qualitative Reasoning about physical systems*, Editeurs D S. Weld & J. de Kleer, 288-305.
- Sunaga T., 1958, *Theory of an Interval Algebra and its Application to Numerical Analysis*, *RAAG Memoirs*, vol. 2, Msc. II, 547-564.
- Thierry G., Doyon B. & Démonet J F., 1998, ERP Mapping in phonological and lexical semantic monitoring tasks: A study complementing previous PET results, *Neuroimage*, 8:391-408.
- Travé-Massuyès L., Bousson K., Evrard J M., Guenin F., Lucas B., Mossier A., Tomasena M. & Zimmer L., 1993, Non-causal versus causal qualitative modeling and simulation, *Intelligent Systems Engineering Journal*, 2(3):159-182.
- Travé-Massuyès L., Dagues P. & Guenin F., 1997, *Le raisonnement qualitatif, pour les sciences de l'ingénieur*, Hermès.
- Tupper J A., *Graphing Equations with generalized interval arithmetic*, 1996, Master of Science Degree, Graduate Department of Computer Science, University of Toronto.
- Vescovi M R., Lamengo M M. & Farquhar, 1997, Modeling and Simulation of a Complex Industrial Process, *IEEE Expert*, 12(3):42-46.
- Wallace J G., Silberstein R B., Bluff K. & Pippingas A., 1993, Semantic Transparency, *Brain Monitoring and Evaluation of Hybrid Cognitive Architectures*, *Connection Science*, 6(1):43-58.
- Xuong N H., 1992, *Mathématiques discrètes et informatique*, Masson.
- Yao Y Y., 1998, Qualitative Fuzzy Sets, Sixth international workshop on rough sets, data mining and granular computing, proceedings of JCIS'98, Research Triangle Park (Caroline du nord, USA), 2:347-350.
- Zeigler B P., 1990, *Object-Oriented Simulation with Hierarchical Modular Models*, Academy Press INC.
- Zilles K., 1996, Pour un nouveau découpage du cortex, *La Recherche*, 289:46-48.

(Nombre de références bibliographiques : 129)

- · 108
 \approx · 105, 106, 107, 108, 113, 114
 \cdot · 114
 $/$ · 108
 $|$ · 114
 $+$ · 108
 $=$ · 105, 112
 $=_3$ · 110
 $=_{N1}$ · 111
 $>$ · 107, 112
 $>_3$ · 110
 \gg · 106

A

ACT · 33; ACT* · 34
 Adaptive Control of Thought · *Voir*
 ACT
 Adaptive Resonance Theory · *Voir* ART
 addition: d'intervalles · 108
 addition qualitative · 69, 71, 74
 aire fonctionnelle · 26
 aires: auditives primaires · 16; de Broca ·
 16, 17; de Wernicke · 16, 17; motrices
 primaires · 15; visuelles primaires · 16
 aires cérébrales · 12
 aires de Brodmann · 17
 algèbre des intervalles · 73; étendus · 104
 algèbre des signes · 69
 algorithme: de combinaison des valeurs
 marginales · 133; de compilation · 137,
 138; de simulation · 141
 algorithme wake-sleep · 44
 amplification · 120
 amplitude · 89, 103
 approche causale · 77
 approche connexionniste · 36
 approche symbolique · 31
 arbre de comportements · 164
 arbre dendritique · 18
 ART · 51 *Voir* Réseau ART
 atlas de Talairach · 24
 Atlas de Talairach · 17
 atome · 93, 103
 atténuation · 111
 atténuation de types · 91
 axone · 18

B

BIOCAEN · 127
 bouton synaptique · 18
 branchement · 76, 164

C

$Ca \sim En$ · 76, 127
 calcul des probabilités · 58
 calculabilité d'un graphe global · 125
 cellule nerveuse · *Voir* neurone
 cerveau · 14
 cervelet · 14
 chaîne de Markov · 60
 coefficient de momentum · 41
 combinaison de types · 90

combinaison d'influences · 123
 compilation des modèles · 128
 complémentaire: d'un type · 113
 composant · 93, 97
 concept de possibilité · 62, 66
 condition d'activation · 117, 119
 connaissance incomplète · 57
 connexion · 92; influence · 94
 connexion de proximité · 83
 connexion distale · 83
 consistance: loi de combinaison · 124
 corps calleux · 15
 cortex · 14; moteur · 15
 cortex cérébral · 14, 17
 cybernétique · 28
 cytoarchitecture · 17

D

délat · 118, 121
 dendrite · 18
 dépendance directe · 119
 distribution de possibilité · 66
 distribution de probabilités · 59
 division: d'intervalles · 108
 domaine d'intervalle · 74

E

EEG · *Voir* électroencéphalographie
 égalité dans L_3 : d'intervalles · 110, 111
 égalité large: de types · 113
 égalité qualitative: Algèbre des signes ·
 69; Intervalles · 74; Ordres de
 grandeur · 71; sur les intervalles
 étendus · 105
 égalité stricte: de types · 112;
 d'intervalles · 105
 électroencéphalographie · 20
 émergence · 36
 encéphale · 13
 énergie · 89 *Voir* amplitude
 ensemble des réels étendus · 104
 ensemble flou · 64
 équation différentielle linéaire du
 premier ordre · 122
 espace des épreuves · 58
 espace des qualités · 68
 étude clinique · 12
 étude d'activation · 9, 11
 événement probabiliste · 58
 événement qualitatif · 74

F

filtrage · 120
 filtre- α : d'un type · 114
 fonction cognitive · 12
 fonction de combinaison · 79
 fonction de mise à jour · 117
 fonction noyau · 53
 fonction qualitative associée · 71
 fonction stochastique · 39

G

gain dynamique · 118, 120
 GPS · *Voir* système expert
 graphe causal · 77, 92
 graphe d'influence · *Voir* graphe causal
 graphe global · 101
 gyri occipitaux · 16
 gyrus · 14
 gyrus précentral · *Voir* Aires motrices
 primaires
 gyrus temporal · 16; supérieur · *Voir*
 aires auditives primaires

H

HAM · 33
 Hebb · 41
 hémisphères · 14
 horloge du système · 115, 116
 Human Associative Memory · *Voir*
 HAM
 hyperréels · 166
 hypothèse: fonctionnelle · 27;
 structurelle · 27
 hypothèse explicite · 81
 hypothèse exploratoire · 81
 hypothèse faible · *Voir* hypothèse
 exploratoire
 hypothèse fonctionnelle · 84, 142
 hypothèse forte · *Voir* hypothèse
 fondamentale
 hypothèse structurelle · 84
 hypothèses · 27
 hypothèses exploratoires · 142
 hypothèses fondamentales · 81; I · 82; II
 · 83; III · 83; IV · 83; V · 87; VI · 89

I

imagerie de surface · 20
 imagerie fonctionnelle · 20
 imagerie par résonance magnétique
 fonctionnelle · 23
 implication matérielle · 62
 inclusion: d'intervalles · 106
 influence · 78, 94
 influence marginale · 78
 influx nerveux · *Voir* potentiel d'action
 information cérébrale · 20
 instanciation · 87
 interface de connexion · 96; schéma de
 connexion · 95
 interface de connexion: point de
 connexion · 95
 intersection: de types · 113; d'intervalles
 · 108
 intervalle dual · 167
 intervalle modal · 167
 intervalles: définition · 74; étendus · 104
 inverse: d'un intervalle · 108
 IRMF · 24 *Voir* imagerie par résonance
 magnétique fonctionnelle

L

langage de description · 128, 129
lobe · 15; frontal · 15; limbique · 15;
occipital · 15; pariétal · 15
Lobe · 15
logique floue · 60
logique L_3 · 62
logique L_{N1} · 63, 119
logique multivaluée · 61
logique possibiliste · 66
logiques de Lukasiewicz · 62
loi de combinaison · 124
longueur d'un intervalle · 105

M

macro-organisation · 25
magnétoencéphalographie · 20
maps · 52
MEG · *Voir* magnétoencéphalographie
mesure de nécessité · 66
mesure de possibilité · 66
modèle causal · 77
modèle de composant · 97; générique · 100
modèle fonctionnel · 85; générique · 87
modèle générique · 87; fonctionnel · 87;
structurel · 87
modèle PDP · 37
modèle probabiliste · 59
modèle structurel · 84; générique · 87
modélisation computationnelle · 28
modélisation fonctionnelle · 29, 31
modélisation structurelle · 30, 47
modus ponens généralisé · 66
MOPs · *Voir* système expert
moteur d'inférences · 32
multiplication: d'intervalles · 108
multiplication par un scalaire: d'un type · 114
multiplication qualitative · 69, 71, 74
multi-symétrie · 124
myéline · 20

N

neurone · 17; formel · 38
neurosciences computationnelles · 28
NEXUS · *Voir* système NEXUS
niveau cellulaire · 48
niveau du tissu · *Voir* niveau régional
niveau fonctionnel · 84
niveau global · 50, 76
niveau local · 76
niveau modulaire · 49
niveau régional · 50
niveau structurel · 84
niveaux d'organisation du cortex · 48
neud d'un graphe · *Voir* atome *Voir*
composant
notion de qualité · 68
noyau: d'ensemble floue · 65

O

O(M) · 72 *Voir* ordres de grandeur
relatifs
ordres de grandeur · 70; absolus · 70;
relatifs · 72

P

paradoxe de Russell · 61, 66
parallel distributed processing · 36
paramètre générique · 100
parents markoviens · 60
pas d'échantillonnage · 115
PDP · 36 *Voir* approche connexionniste
phase d'initialisation · 121
physique naïve · 68
plasticité du cerveau · 18
poids maximal: d'un type · 114
point d'entrée · *Voir* point de connexion
point de connexion · 95
points de sortie · *Voir* point de sortie
populations de neurones · 53
potentiel d'action · 19
potentiel évoqué · *Voir* réponse évoquée
primitive fonctionnelle · 85
principe de non-contradiction · 61
principe de tiers exclu · 59, 61
probabilité · 58
processeur élémentaire · 85
processeur fonctionnel · 92
produit: de types · 113
PUPS · 33

Q

QSIM · 76, 164

R

raisonnement qualitatif · 68
rCBF · 149
recalage · 121
règles floues · 66
relation « englobe » · 97
réponse évoquée · 21
réseau ART · 43
réseau bayésien · 60
réseau de colonnes formelles · 50
réseau de Kohonen · 42
réseau de neurones formel · 36 *Voir*
RNF
rétinopathie · 16
rétro-propagation des erreurs · 41
RNF · 37, 49, 50 *Voir* réseau de
neurones formels; algorithme wake-
sleep · 44; apprentissage · 41;
apprentissage non supervisé · 42;
apprentissage auto-organisateur · 42;
apprentissage supervisé · 41; fonction
de sortie · 39; plasticité · 42, 44; règle
d'activation · 39; réseau à deux
couches · 38; réseau à multicouches ·
38; réseau ART · 43; réseau complet ·
38; réseau de Kohonen · 42; réseau de
neurones flous · 66; rétro-propagation
· 41; stabilité · 42, 44
RNF: réseau de Hopfield · 45

S

schéma de connexion · 95
scissure: centrale · 15; de Rolando · 15;
de Sylvius · 15; hémisphérique · 15;
latérale · 15
Scissure · 14

semi-qualitatif · 75
seuil d'activation · 19
simulation qualitative · 75
SNC · 13
SOAR · 36
soma · 18
somatotopie · 15
somme: de types · 113
soustraction: d'intervalles · 108
substance: blanche · 19; grise · 19
supériorité: de types · 112
supériorité dans L_3 : d'intervalles · 110
supériorité dans L_{N1} : d'intervalles · 111
supériorité stricte: d'intervalles · 106
supériorité totale: d'intervalles · 107
support: d'ensemble floue · 65
synapse · 18
système expert · 32; general problem
solver · 32; Memory Organizer
Packets · 33
système nerveux central · 13
système NEXUS · 52

T

technique de surface · *Voir* imagerie de
surface
techniques de tomographie · *Voir*
tomographie
temps de réponse · 118, 122
TEP · *Voir* tomographie par émission de
positron
théorie de la localisation · 11
théorie des ensembles flous · *Voir*
logique floue
tomographie · 22
tomographie par émission de positron ·
22
tonotopie · 16
tops d'horloge · 116
transparence sémantique · 32
tronc cérébral · 14
type · 111; composé · 112; pur · 112;
total · 112; vide · 112

U

union: de types · 113; d'intervalles · 107
unités cognitives · 34

V

valeur d'activation · 38, 39
valeur de sortie · 38, 39
valeur marginale · 124
validité d'un graphe · 99
variable: endogène · 78; exogène · 78
variation marginale · 78
volume de l'activité · 88

W

Wave theory · 54

Z

zone fonctionnelle · 83
zone fonctionnelle cérébrale · 83

ANNEXE A :
EXEMPLES COMMENTES DE DESCRIPTIONS BIOCAEN.

Nous allons commenter ici les listings des fichiers de description compilés par BIOCAEN⁷ et correspondant aux modèles présentés au chapitre 5 (pages 142 et 148). Cette présentation n'est qu'un complément d'information et nécessite donc un lecture préalable du chapitre 5, elle ne fournit pas non plus une description complète des possibilités du langage de description utilisé par BIOCAEN. Dans les listings qui suivent les mots clés du langage sont écrits en **gras**, les commentaires sont en *italique* (et précédés d'un point virgule).

Pour faciliter la compréhension de ces listings voici quelques-unes des fonctions utilisées :

(**Type** ATN) permet d'accéder au type de la valeur de l'atome ATN. Pour accéder à l'amplitude, il existe aussi une fonction (**Magn** ATN) mais, par défaut, ATN suffit y pour l'obtenir.

(**Before** 5 ATN) permet d'obtenir la valeur de l'atome ATN il y a 5 millisecondes dans le passé, soit si t représente l'instant courant, ATN(t-5).

1. DESCRIPTION DU MODELE GENERIQUE DE ZONES ASSOCIATIVES

Nous commentons ici le listing du modèle générique présenté dans la partie 2.1 du chapitre 5 (page 142 et suivantes). Cette description définit cinq paramètres génériques dont la table de préférence des types. Il définit aussi deux schémas de connexion en entrées (i.e. **In-Influences**) et un unique point de sortie (i.e. **Out-Node**). Le graphe décrit ici est composé de quatre atomes et d'un composant. La description du modèle associé à ce dernier est stocké dans un fichier appelé « Integrate.bio », dont le listing est donné plus bas.

a) Modèle de la zone :

```
; Modèle de zone générique. (fichier GenZon00.bio) (révision 01,00)
(interface                ; section de description de l'interface du modèle
Parameters              ; définition des paramètres génériques
```

⁷ Ces listings utilisent le langage de description défini par la version 1.1 de BIOCAEN qui présente quelques modifications du langage de description mais qui reste entièrement compatible avec les descriptions de la littérature (Pastor et al. 2000, Lafon et al. 1999, Lafon 1998 et dans une certaine mesure Lafon et al. 1997). Une présentation exhaustive du langage de description peut être obtenue dans le manuel de BIOCAEN (Lafon 1999).

```

Typ-Pref                ; Table de préférence des types
Spatial-Int             ; Loi de combinaison réalisant l'intégration spatiale
ST                      ; Seuil de saturation (Saturation Threshold)
BT                      ; Seuil d'émission (Broadcast Threshold)
CT                      ; Seuil de validité du type pour émission (Consistence Threshold)
NAN                    ; Valeur résiduelle de l'amplitude de l'activation (Natural Value of AN).
MR                      ; Temps réponse de la mémoire de la zone.
)                        ; Fin de la définition des paramètres génériques

(In-Influences          ; définition de schémas de connexions en entrée
((propagate X Y delay)
 (Integrate X INN Ready_state Typ-Pref delay)
)                        ; définit la connexion Propagate qui schématise une connexion dirigée vers le
                        ; composant INN (qui définit le schéma Integrate).
((inhibit X Y delay tr)
 (Influence X IHN true 1 delay tr)
)                        ; définit la connexion inhibit qui schématise une influence dirigée vers l'atome IHN.
)                        ; fin de la définition des schémas de connexion en entrée.

(Out-Node BN)          ; définition d'un unique point de sortie, l'atome BN.
)                        ; fin de l'interface du modèle.

(display " Generic Zone Model." ) ; Affiche des messages lors de la compilation.
(display " Version : Jan 2000." )
(display " Copyright INSERM Unit 455." )

(Atoms                ; Section de description des Atomes du graphe
  IHN                   ; Nœud d'inhibition de la zone.
  (AN display)         ; Nœud d'activation de la zone, le mot-clé display, stipule que la valeur du nœud
                        ; peut être affichée dans le fichier de résultats de la simulation.
  BN                   ; Nœud de propagation de la zone, point de sortie du modèle (Broadcast Node)
  MN                   ; Nœud mémoire de la zone.
)                        ; fin de la section de description des Atomes du graphe

(Components           ; Section de description des composants du graphe
(INN "Integrate.bio" Typ-Pref Spatial-Int) ; Composant d'intégration de la zone.
                        ; son modèle est décrit dans le fichier « integrate.bio », les paramètres génériques qui y sont
                        ; définis sont instanciés par de nouveaux paramètres génériques (transfert de généralité).
)                        ; fin de la section de description des composants du graphe

(Conditions           ; Section de définition de conditions d'activations.
(Inhibited (INH <> 0)) ; est vrai si la zone est inhibée.
(Firing (((before 1 AN) > BT)
          and (((type AN) c (type MN)) or ((maxp AN) > CT)))) ; la zone est activée.
(Saturated (AN > ST)) ; la zone est saturée.
(Ready_state (not (Inhibited or Firing or Saturated)))
)                        ; fin de la section Conditions.

(Connexions           ; Section de description des connexions du graphe
(Influence INN AN (Not Saturated) 1 1 0 0) ; Connexion de INN sur AN
(Influence INH AN true -1 1 0 0) ; Influence négative de IHN sur AN
(Constant AN true NAN ( ) 0) ; Activation résiduelle de la zone.
(Influence AN BN ((Not Inhibited) and Firing) 1 1 1 0) ; Influence d'émission
(transfer BN MN true 1 1 0 MR) ; Mémorisation de l'information émise.
)                        ; fin de la section de description des connexions.
)                        ; fin de la description du modèle.

```

b) Modèle d'intégration :

```

;M odèle de zone générique. (fichier Integrate.bio) (révision 01/00)
(interface                               ; section de description de l'interface du m odèle
(Parameters                               ; définition des param ètres génériques
  Typ-Pref                                 ; Table de préférence pour le filtrage en entrée
  Spatial-Int                               ; Loide combinaison du nœud SIN
)      ; Fin de la définition des param ètres génériques

(In-Influences                             ; définition de schém as de connexion en entrée
((Integrate X INN Ready_state Typ-Pref delay)
 (Influence X SIN Ready_state Gating 1 delay 0)
)      ; définit la connexion Integrate qui schém atise une connexion dirigée vers l'atom e SIN .
)      ; fin de la définition des schém as de connexion en entrée.

(Out-Node TIN)                             ; définition du point de sortie, l'atom e TIN .
)      ; fin de l'interface du m odèle.

(display "   Generic Integration Model." )      ; Affiche des m essages lors de la compilation.
(display "   Version : Jan 2000." )
(display "   Copyright INSERM Unit 455." )

(define dynamic-gain Gating X              ; définit le gain dynam ique de l'influence dirigé sur SIN ,
  ((maxp ((Type X) &T TPT) * X)           ; Filtre les inform ations; « écrase » l'amplitude de X .
)

(Atoms                                     ; Section de description des Atom es du graphe
(SIN display Spatial-Int) ; Nœud d'intégration spatiale, sa loide combinaison est un param ètre générique.
TIN                                     ; Nœud d'intégration temporelle.
)      ; fin de la section de description des Atom es du graphe

(Connexions                               ; Section de description des connexions du graphe
(Influence SIN TIN true 1 1 0 0)          ; Influence de SIN sur TIN .
(Influence SIN TIN (Not (SIN =Ln (Before 1 TIN)) 1 1 1 0) ; Boucle d'intégration
)      ; fin de la section de description des connexions.
; fin de la description du m odèle.

```

2. MODELISATION DE LA REPONSE DE L'AIRES VISUELLE PRIMAIRE A UN STIMULUS SIMPLE.

Nous présentons ici les listings des modèles publiés dans Pastor *et al.* 2000.

2.1. La capacité d'apprentissage de l'aire visuelle primaire

a) Modèle générique de l'aire visuelle primaire :

```

;M odèle de zone générique. (fichier Zone.bio) (révision 01/00)
(interface                               ; section de description de l'interface du m odèle
(Parameters                               ; définition des param ètres génériques
  TPT                                       ; Table de préférence de type
  FT                                       ; Seuil d'émission
  Again                                    ; Gain statique de l'influence IN → AN ;
  Aresp                                   ; Temps de réponse du transfert de l'inform ation à AN
  Igain                                   ; Gain statique de l'influence d'auto-inhibition.
  OGresp                                  ; Temps de réponse de l'influence d'émission (i.e. AN → OGN )
  Mgain Mresp                             ; Gain et temps de réponse de la mém orisation.
)      ; Fin de la définition des param ètres génériques

```

```

(In-Influences ; définition de schémas de connexion en entrée
((Input-Gating stim input-node delay)
(Influence stim IGN True IGgain 1 delay 0))
) ; fin de la définition des schémas de connexion en entrée.

(Out-Node OGN) ; définition d'un unique point de sortie, l'atome OGN.
) ; fin de l'interface du modèle.

(Atoms ; Section de description des Atomes du graphe
(IGN display (space (0 inf))) ; Atome de filtrage d'entrée (InputGatingNode)
(AN display) ; Atome d'Activation
(MN display (space (0 inf))) ; Mémoire de travail (MemoryNode)
(OGN display (space (0 inf))) ; Atome de sortie (OutputGatingNode)
(IN display (space (0 inf))) ; Inhibition
) ; fin de la section de description des Atomes du graphe

(Conditions ; Section de définition de conditions d'activations.
(firing (((before 1 AN) >= FT) ; Condition d'émission
or (((minp (before 1 MN)) > 0.01)
and ((minp (before 1 AN)) > 0.01)
and ((before 1 AN) > 0))))
) ; fin de la section Conditions.

(define dynamic-gain IGgain X ; définit le gain dynamique de l'influence dirigé sur IGN,
((maxp (in-var *T TPT)) * X))

(Connexions ; Section de description des connexions du graphe
(Influence IGN AN ((before 1 OGN) = 0) Again 1 0 Aresp)
(Influence IN AN True -1 1 0 Aresp)
(Influence AN OGN firing 1 1 0 OGresp)
(Influence OGN MN True Mgain Mgain 1 Mresp)
(Influence OGN IN True Igain 1 1 0)
) ; fin de la section de description des connexions.
; fin de la description du modèle.

```

b) Niveau structurel :

C'est ce fichier qui est compilé par le compilateur, il contient la description du niveau structurel du modèle, il définit la seule et unique zone , V1, comme l'instance du modèle contenu dans le fichier listé ci-dessus.

```

(Config ; Section de définition des paramètres généraux du système.
(checkinterval 1 msc) ; définition du pas d'échantillonnage (1 milliseconde).

(Atoms ; Section de description des Atomes
(Stm stimulus) ; définit un atome exogène.

(zones ; Section de description des composants
(V1 "zone.bio" (type flash) 1 1.5 (8 10) 0.8 (8 10) 5 (64 128)))

(Connexions ; Section de description des connexions
(Input-gating Stm V1 0))
; fin de la description du modèle.

```

2.2. Modèle de la boucle cortico-thalamique

a) Modèle fonctionnel du cortex strié (aire visuelle primaire) :

```

;M odèle de zone générique. (fichier Cortex.bio) (révision 01,00)
(interface ; section de description de l'interface du modèle
(Parameters ; définition des paramètres génériques
  TPT ; Table de préférence de type
  BasicFT ; Valeur maximale du Seuil d'émission
  Again ; Gain statique de l'influence IN → AN ;
  Aresp ; Temps de réponse du transfert de l'information à AN
  Igain ; Gain statique de l'influence d'auto-inhibition.
  OGresp ; Temps de réponse de l'influence d'émission (AN → OGN)
  FTgain FTresp ; Gain et temps de réponse de la connexion Thalamus → FT.
) ; Fin de la définition des paramètres génériques

(In-Influences ; Définition de schémas de connexions en entrée
((input-gating stim input-node delay) ; Connexion de propagation
(Influence stim IGN True IGgain delay 0))
((firing-threshold thal-input FT-node delay-thal) ; Connexion de retour du thalamus
(Influence thal-input FT True FTgain delay-thal FTresp))
) ; fin de la définition des schémas de connexion en entrée.

(Out-Node OGN) ; définition d'un unique point de sortie, l'atome OGN.
) ; fin de l'interface du modèle.

(Atoms ; Section de description des Atomes du graphe
(IGN display (space (0 inf))) ; Atome de filtrage d'entrée (Input Gating Node)
(AN display) ; Atome d'Activation
(FT display) ; Seuil d'émission dynamique
(OGN display (space (0 inf))) ; Atome de sortie (Output Gating Node)
(IN display (space (0 inf))) ; Inhibition
) ; fin de la section de description des Atomes du graphe

(Conditions ; Section de définition de conditions d'activations.
(firing (((before 1 AN) >= FT) ; Condition d'émission
and ((before 1 AN) > 0)
and ((before 1 AN) > (before 1 FT))))
) ; fin de la section Conditions.

(define dynamic-gain IGgain X ; définit le gain dynamique de l'influence dirigé sur IGN,
((maxp (in-var *T TPT) * X))

(Connexions ; Section de description des connexions du graphe
(Influence IGN AN ((before 1 OGN) = 0) Again 1 0 Aresp)
(Influence IN AN True -1 1 0 Aresp)
(Constant FT True basicFT nil)
(Influence AN OGN firing 1 1 0 OGresp)
(Influence OGN IN True Igain 1 1 0)
) ; fin de la section de description des connexions.
; fin de la description du modèle.

```

b) Modèle fonctionnel générique de la structure thalamique :

```

;M odèle de zone générique. (fichier Thalamus.bio) (révision 01,00)
(interface ; section de description de l'interface du modèle
(Parameters ; définition des paramètres génériques
  Ogresp ; Temps de réponse de l'influence d'émission (AN → OGN)
) ; Fin de la définition des paramètres génériques

(In-Influences ; Définition de schémas de connexions en entrée
((input-thal information input-node delay)
(Influence information IGN True 1 1 delay 0))
)

(Out-Node OGN) ; définition d'un unique point de sortie, l'atome OGN.

```

```

) ; fin de l'interface du modèle.

(Atoms ; Section de description des Atomes
 (IGN display) ; Nœud d'entrée
 (OGN display) ; Nœud de sortie
)

(Connexions ; Section de description des connexions
 (Influence IGN OGN True 1 1 0 OGresp))
; fin de la description du modèle.

```

c) Niveau structurel :

```

; Modèle de zone générique. (fichier CtxThalLoop.bio) (révision 01/00)
(Config ; Section de définition des paramètres généraux du système.
 (checkinterval 1 msc) ; définition du pas d'échantillonnage (1 milliseconde).

(Atoms (Stm stimulus) ; Section de description des Atomes

(zones ; Section de description des composants
 (Striate-Ctx "Cortex.bio" (Type 'flash) 1 1.5 8 0.8 10 (-90 -80) (192 256))
 (Thalamus "Thalamus.bio" 1))

(Connexions ; Section de description des connexions
 (Input-Gating Stm Striate-Ctx 1)
 (Firing-Threshold Thalamus Striate-Ctx 0)
 (Input-Thal Striate-Ctx Thalamus 1))
; fin de la description du modèle.

```

Marc LAFON

MODELISATION OF CEREBRAL
INFORMATION PROPAGATION IN CAUSAL
QUALITATIVE NETWORKS

Ph.D. Advisor : Doctor Josette PASTOR
INSERM Unité 455 – Toulouse
Neuro-Imagerie fonctionnelle, plasticité cérébrale et pathologie neurologique.

Ph.D. Co-advisor : Doctor Louise TRAVÉ-MASSUYES
LAAS/CNRS – Toulouse
Groupe Automatique et Symbolique.

- Abstract -

In the domain of neuroscience, activation studies use neuro-imaging techniques to provide measures of human brain activity. These studies highlight large-scale networks of cerebral areas involved in the processing of specific cognitive functions. The cerebral information processing complexity and the incompleteness of neuro-imaging data blur the understanding of the links between the network and the cognitive function. This arises the necessity of a formal modelling tool and computer simulation. Moreover, understanding of functional mechanisms must be achieved through an explicit representation. We propose an original formalism based on causal graphs associated to a qualitative representation of the information. The graphs fulfil the need for a twofold representation of the brain: structural and functional. We propose a hierarchical organisation, based on components, that reduces the computing complexity and extends the reusability of the models. This organisation is described by a homogeneous formalism, at both structural and functional levels. We propose a qualitative abstraction of the cerebral information, by a couple: (interval-symbol). With this dual representation, we can deal with the information magnitude, which represents the level of cerebral activity and can be measured by imaging tools, and the category of the information. In the BIOCAEN simulator, the main simulation mechanism consists in the propagation of influences through the causal graph. This mechanism allows to express the brain's stimulus-driven processes, which are the major part of the cerebral activity, and the source of brain's efficiency. In the field of the computational neurosciences, our contribution offers the neuroscientists, a formal tool for the expression of hypotheses on cerebral information propagation, at a high level of abstraction. Furthermore, the simulation can give a validation of the expressed hypotheses by a comparison of the simulation results and neuro-imaging data.

Keywords : Cognitive modeling, Hierarchical causal graph, Qualitative simulation.

INSERM Unité 455,
Services de Neurologie, CHU Purpan,
F-31059 TOULOUSE CEDEX.