Réseaux causaux probabilistes à grande échelle : un nouveau formalisme pour la modélisation du traitement de l'information cérébrale

Vincent Labatut

RESEAUX CAUSAUX PROBABILISTES A GRANDE ECHELLE :
UN NOUVEAU FORMALISME POUR LA MODELISATION DU TRAITEMENT DE L'INFORMATION CEREBRALE

THESE

Pour obtenir le grade de
Docteur en Informatique de l'Université Toulouse III
Présentée et soutenue le 18 décembre 2003 par

VINCENT LABATUT

JURY

Claudette Cayrol, Professeur des Universités, *Univesité Toulouse III – Paul Sabatier*	Présidente du jury
Salem Benferhat, Professeur des Universités, *Université d'Artois, Lens*	Rapporteur
José M. Bernardo, Professeur des Universités, *Université de Valence, Espagne*	Rapporteur
Emmanuel Guigon, CR CNRS, *INSERM U483, Paris*	Examinateur
Henri Prade, DR CNRS, *Institut de Recherche en Informatique de Toulouse*	Examinateur
Josette Pastor, IR INSERM & HDR, *INSERM U455, Toulouse*	Directrice de recherche
Pierre Celsis, DR INSERM, *INSERM U455, Toulouse*	Invité

UNIVERSITE TOULOUSE III - PAUL SABATIER
U.F.R. Mathématiques, Informatique et Gestion.

UNIVERSITE TOULOUSE III - PAUL SABATIER

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

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE TOULOUSE III

Discipline : INFORMATIQUE

Présentée et soutenue par

Vincent Labatut

Le 18 décembre 2003

RESEAUX CAUSAUX PROBABILISTES A GRANDE ECHELLE : UN NOUVEAU FORMALISME POUR LA MODELISATION DU TRAITEMENT DE L'INFORMATION CEREBRALE

JURY

Claudette Cayrol, Professeur des Universités, *Univesité Toulouse III – Paul Sabatier,*	Présidente
Salem Benferhat, Professeur des Universités, *Université d'Artois, Lens,*	Rapporteur
José M. Bernardo, Professeur des Universités, *Université de Valence, Espagne,*	Rapporteur
Emmanuel Guigon, CR CNRS, *INSERM U483*, Paris	Examinateur
Henri Prade, DR CNRS, *Institut de Recherche en Informatique de Toulouse*,	Examinateur
Josette Pastor, IR INSERM & HDR, *INSERM U455*, Toulouse	Directrice
Pierre Celsis, DR INSERM, *INSERM U455*, Toulouse	Invité

Vincent Labatut

RESEAUX CAUSAUX PROBABILISTES A GRANDE ECHELLE : UN NOUVEAU FORMALISME POUR LA MODELISATION DU TRAITEMENT DE L'INFORMATION CEREBRALE

Directeur de thèse :
Josette Pastor, ingénieur de recherche et habilitée à diriger des recherches
INSERM U455

- Résumé -

La compréhension du fonctionnement cérébral passe par l'étude des relations entre les structures cérébrales et les fonctions cognitives qu'elles implémentent. Les études en activation, qui permettent d'obtenir, grâce aux techniques de neuroimagerie fonctionnelle, des données sur l'activité cérébrale pendant l'accomplissement d'une tâche cognitive, visent à étudier ces liens. Ces études, ainsi que de nombreux travaux chez l'animal, suggèrent que le support neurologique des fonctions cognitives est constitué de réseaux à grande échelle d'aires corticales et de régions sous-corticales interconnectées. Cependant, la mise en correspondance simple entre réseaux activés et tâche accomplie est insuffisante pour comprendre comment l'activation découle du traitement de l'information par le cerveau. De plus, le traitement cérébral est très complexe, et les mesures fournies par la neuroimagerie sont incomplètes, indirectes, et de natures différentes, ce qui complique grandement l'interprétation des données obtenues. Un outil de modélisation explicite des mécanismes de traitement et de propagation de l'information cérébrale dans les réseaux à grande échelle est nécessaire pour palier ces défauts et permettre l'interprétation des mesures de l'activité cérébrale en termes de traitement de l'information.

Nous proposons ici un formalisme original répondant à ces objectifs et aux contraintes imposées par le système à modéliser, le cerveau. Il est basé sur une approche graphique causale et probabiliste, les réseaux bayésiens dynamiques, et sur une représentation duale de l'information. Nous considérons le cerveau comme un ensemble de régions fonctionnelles anatomiquement interconnectées, chaque région étant un centre de traitement de l'information qui peut être modélisé par un noeud du réseau bayésien. L'information manipulée dans le formalisme au niveau d'un noeud est l'abstraction du signal généré par l'activité de la population neuronale correspondante. Ceci nous conduit à représenter l'information cérébrale sous la forme d'un couple numérique/symbolique, permettant de tenir compte respectivement du niveau d'activation et de la configuration des neurones activés.

Ce travail se situe dans le prolongement d'un projet visant à développer une approche causale originale pour la modélisation du traitement de l'information dans des réseaux cérébraux à grande échelle et l'interprétation des données de neuroimagerie. L'aspect causal permet d'exprimer explicitement des hypothèses sur le fonctionnement cérébral. Notre contribution est double. Au niveau de l'intelligence artificielle, l'utilisation de variables aléatoires labellisées dans des réseaux bayésiens dynamiques nous permet de définir des mécanismes d'apprentissage non-supervisés originaux. Sur le plan des neurosciences computationnelles, nous proposons un nouveau formalisme causal, plus adapté à la représentation du fonctionnement cérébral au niveau des réseaux d'aires que les réseaux de neurones formels, et présentant plus de plausibilité biologique que les autres approches causales, en particulier les réseaux causaux qualitatifs.

Mots Clés : neurosciences computationnelles – réseau bayésien dynamique – filtre de Kalman – neuroimagerie fonctionnelle – réseau cérébral à grande échelle – apprentissage adaptatif non-supervisé.

INSERM U455
Pavillon Riser, CHU Purpan, 31059 Toulouse Cedex 3

TABLE DES MATIERES

Table des matières 1
Illustrations 4
Tableaux 6
Remerciements 7
Introduction 11
I. Neurosciences Intégratives 15
1. Cerveau 15
1.1. Description anatomique 16
1.2. Description fonctionnelle 21
1.3. Plasticité cérébrale 28
2. Neuroimagerie fonctionnelle 32
2.1. Techniques de surface 33
2.2. Techniques tomographiques 35
2.3. Autres Techniques 38
2.4. Etudes en activation 39
3. Contraintes 40
3.1. Architecture en réseau 40
3.2. Temporalité 41
3.3. Non-linéarité 42
3.4. Incertitude et imprécision 42
3.5. Causalité 43
3.6. Information cérébrale au niveau intégré 44
3.7. Plasticité 44
II. Modélisation cérébrale 47
1. Neuroimagerie 48
1.1. Localisation 48
1.2. Coactivation 50
1.3. Liens anatomiques 52
1.4. Bilan 53
2. Approche cognitive 54
2.1. Modèles symboliques 55
2.2. Réseaux bayésiens 56
2.3. Bilan 57
3. Neurosciences computationnelles 58
3.1. Réseaux de neurones 59
3.2. BioCaEn 63
3.3. Bilan 65
4. Formalisation des contraintes 67
4.1. Causalité 67
4.2. Autres contraintes 69
III. Modélisation Causale 73
1. Caractéristiques des formalismes existants 73
1.1. Réseaux de neurones formels 74
1.2. Simulation qualitative 84
1.3. Formalismes de l'incertain et de l'imprécis 90
1.4. Conclusion 101
2. Algorithmes pour l'inférence et l'apprentissage 103
2.1. Inférence dans les modèles d'espace d'états non-linéaires 103
2.2. Apprentissage dans les réseaux ART 112

IV. Concepts et définitions 117
1. Représentation en réseau 117
1.1. Réseau statique 118
1.2. Réseau dynamique 123
2. Représentation de l'information cérébrale 125
2.1. Magnitude 126
2.2. Type 126
2.3. Complémentarité des deux composantes 130
3. Traitement et propagation de l'information cérébrale 132
3.1. Définitions 132
3.2. Description des mécanismes 134
4. Processus d'apprentissage 139
4.1. Définitions 139
4.2. Description des mécanismes 140
5. Conclusion 143

V. RAGE : définition du formalisme 145
1. Représentation de l'information 146
1.1. Magnitude 146
1.2. Type 146
2. Description des nœuds 152
2.1. Etat d'un nœud 152
2.2. Table de préférence des types 152
3. Propagation et traitement de l'information 155
3.1. Activation 156
3.2. Emission 157
3.3. Conclusion 166
4. Mécanismes d'apprentissage 167
4.1. Renforcement, introduction et oubli 167
4.2. Glissement et fusion 169
4.3. Conclusion 171
4.4. Exemple 172
5. Discussion sur le formalisme 177
5.1. Propriétés du formalisme 177
5.2. Respect des contraintes 184
5.3. Comparaisons avec d'autres formalismes 184
6. Etude du comportement d'un noeud 187
6.1. Paramètres 187
6.2. Type d'activation constant 189
6.3. Type d'activation variable 193
6.4. Apprentissage à partir d'une TPT vierge 196

VI. Applications 203
1. Modèle de la réponse de l'aire visuelle primaire à un stimulus simple 204
1.1. Etude en activation 204
1.2. Modèle de boucle thalamo-corticale 206
1.3. Simulation et discussion 210
2. Modèle de la réponse du gyrus temporal supérieur droit lors d'une tâche passive de categorisation 212
2.1. Etude en activation 213
2.2. Modèle de gyrus temporal supérieur droit 216
2.3. Simulation et discussion 222
3. Conclusion 227

Conclusion 229
1. Un formalisme dédié à la modélisation cérébrale à un niveau intégré 229
2. Position de notre approche 232
3. Perspectives 234

Bibliographie 239

Index 253

A. Rappels et notions théoriques 257
1. Probabilités et statistiques 257
1.1. Notions de probabilités 257
1.2. Estimation en statistiques 266
2. Ensembles flous et théorie des possibilités 269
2.1. Ensembles flous 269
2.2. Théorie des possibilités 274
3. Calculs qualitatif et semi-qualitatif 278
3.1. Calcul qualitatif 278
3.2. Calcul semi-qualitatif 281
4. Neurone formel 284
B. Algorithmes et implémentation 287
1. Organisation générale 287
2. Représentation des données 289
2.1. Information 289
2.2. TPT 290
2.3. Fonctions pour le traitement et l'apprentissage 291
3. Algorithmes 292
3.1. Fonctions auxiliaires 293
3.2. Traitement principal 295
4. Implémentation 298
4.1. Interface 298
4.2. Moteur 302

ILLUSTRATIONS

Figure I.1.1 : anatomie externe de l'encéphale, découpage en lobes. 16
Figure I.1.2 : anatomie du neurone biologique. 18
Figure I.1.3 : découpage anatomique du cerveau en aires de Brodmann. 19
Figure I.1.4 : faisceaux d'axones entre des aires corticales..... 21
Figure I.1.5 : découpage fonctionnel du cortex cérébral. 22
Figure I.2.1 : exemple de reconstruction d'une image de l'activation cérébrale par EEG [Franceries *et al.* '03]. 34
Figure I.2.2 : coupes réalisées par TEP (© Inserm u455 2003). 35
Figure I.2.3 : exemples d'images de l'activation cérébrale par IRMf [Ruff *et al.* '03]. 36
Figure III.1.1 : exemples de RNF feed-forward. 75
Figure III.1.2 : exemples de RNF récurrents. 75
Figure III.1.3 : diagramme causal d'un réseau qualitatif simple. 85
Figure III.1.4 : réseau bayésien simple représenté sous la forme d'un graphe orienté acyclique..... 92
Figure III.1.5 : deux tranches temporelles d'un réseau bayésien dynamique simple. 96
Figure III.1.6 : une chaîne de Markov d'ordre 1. 97
Figure III.1.7 : exemple de modèle de Markov caché/modèle d'espace d'états. 98
Figure III.1.8 : modèle de Markov caché/modèle d'espace d'états avec une entrée. 99
Figure III.2.1 : structure d'un modèle ART1. 113
Figure IV.1.1 : exemple de réseau structurel. 119
Figure IV.1.2 : décomposition fonctionnelle du nœud structurel représentant le cortex cingulaire postérieur de la Figure IV.1.1. 120
Figure IV.1.3 : exemple de modèle générique d'une aire corticale (adapté de la représentation graphique propre à BioCaEn) [Lafon *et al.* '97]. 122
Figure IV.1.4 : deux instanciations du modèle générique décrit dans la Figure IV.1.3. 122
Figure IV.1.5 : résumé du processus menant d'un réseau structurel à un réseau dynamique. 124
Figure IV.1.6 : exemple de réseau dynamique correspondant au réseau fonctionnel de la Figure IV.1.4.a. 124
Figure IV.2.1 : décomposition correspondant à la description de l'information transitant par une population de neurones. 126
Figure IV.2.2 : illustration des concepts de magnitude et de type. 130
Figure IV.2.3 : illustration des concepts de magnitude et de type. 131
Figure IV.3.1 : décomposition d'un nœud du réseau dynamique due au traitement de l'information cérébrale. 133
Figure IV.3.2 : l'information décrivant l'état d'un nœud dynamique est constituée de quatre valeurs : magnitude et type d'activation, et magnitude et type d'émission. 134
Figure IV.3.3 : Calcul des valeurs d'activation d'un nœud B possédant *n* parents $A_1,\ldots,A_n$. 135
Figure IV.3.4 : Calcul des valeurs d'émission d'un nœud. 137
Figure IV.3.5 : La table de préférence des types (TPT) est un paramètre pouvant intervenir lors du calcul des valeurs d'émission. 138
Figure IV.4.1 : L'apprentissage est réalisé en modifiant un paramètre du nœud, appelé table de préférence des types (TPT). 140
Figure IV.4.2 : le mécanisme de renforcement. 141
Figure IV.4.3 : le mécanisme d'introduction. 142
Figure IV.4.4 : le mécanisme de glissement. 143
Figure IV.5.1 : ensemble des mécanismes d'apprentissage, de représentation et de traitement de l'information. 144
Figure V.3.1 : comparaison entre trois mesures de similitude. 159
Figure V.3.2 : comparaison entre trois mesures de similitude basées sur des sigmoïdes. 160
Figure V.3.3 : exemples de fonctions $f_{T_X}^{(1)}$ dans (eq. V.2.27). 162
Figure V.6.1 : nœud soumis à un type d'activation constant proche de l'archétype A_1, et à une magnitude aléatoire. 190
Figure V.6.2 : nœud soumis à un type d'activation constant différent de tous les archétypes de la TPT, et à une magnitude aléatoire. 191
Figure V.6.3 : nœud soumis à un type d'activation constant $\tilde{T}_{Bruit}$, représentant du bruit, et à une magnitude aléatoire. 193

Figure V.6.4 : nœud soumis à des type et magnitude d'activation aléatoires. 194
Figure V.6.5 : nœud soumis à des type et magnitude d'activation partiellement aléatoires..... 195
Figure V.6.6 : nœud soumis à des type et magnitude d'activation aléatoires, avec une TPT initiale vide..... 198
Figure V.6.7 : nœud soumis à des type et magnitude d'activation aléatoires, avec une TPT initiale vide..... 199
Figure V.6.8 : nœud soumis à des type et magnitude d'activation partiellement aléatoires, avec une TPT initiale vide..... 200
Figure V.6.9 : nœud soumis à des type et magnitude d'activation partiellement aléatoires, avec une TPT initiale vide..... 201
Figure VI.10 : résultats de l'expérience de Fox & Raichle utilisant des flashes lumineux comme stimuli [Fox & Raichle '84, '85]..... 205
Figure VI.1.11 : modèle de cortex visuel incluant une boucle thalamo-corticale. 207
Figure VI.1.12 : résultats des simulations obtenues par BioCaEn et RAGE, comparés aux valeurs expérimentales de Fox & Raichle [Fox & Raichle '84, '85]. 212
Figure VI.2.1 : organisation d'un run. 213
Figure VI.2.2 : disposition des différents mixages de syllabes utilisés pour la stimulation. 214
Figure VI.2.3 : activation moyenne du gyrus temporal supérieur droit pour chacun des 5 types de blocs [Ruff '00]. 215
Figure VI.2.4 : modèle de gyrus temporal supérieur droit destiné à reproduire un processus de catégorisation automatique. 217
Figure VI.2.5 : comparaison des activations expérimentales et simulées. 223
Figure VI.2.6 : évolution du niveau d'activation des nœuds d'entrée IGN_{pa} et IGN_{ta} dans les modèles témoin et dyslexique. 225
Figure A.1.1 : représentation graphique d'une fonction de densité normale (eq. A.1.42). 264
Figure A.2.1 : exemple de fonction d'appartenance d'un ensemble flou. 270
Figure A.2.2 : illustration des notions d' α -coupe, de support et de noyau d'un ensemble flou. 270
Figure A.2.3 : mesures de possibilité d'un évènement. 275
Figure A.2.4 : mesures de nécessité d'un évènement. 276
Figure A.4.1 : représentation graphique d'un neurone formel. 284
Figure B.1.1 : décomposition de la Figure IV.5.1 en deux sous-réseaux, l'un dédié à la magnitude et l'autre aux types et TPT. 288
Figure B.2.1 : exemple d'utilisation de la structure de données utilisée pour représenter un type. 289
Figure B.2.2 : exemple d'utilisation de la structure de données utilisée pour représenter une TPT. 290
Figure B.2.3 : exemple d'utilisation de la structure de données utilisée pour représenter une fonction. 292
Figure B.4.1 : définition du réseau statique. 299
Figure B.4.2 : définition des équations associées aux noeuds. 299
Figure B.4.3 : définition des stimuli. 300
Figure B.4.4 : mise en œuvre de la simulation. 301
Figure B.4.5 : résultats de la simulation. 302

TABLEAUX

Tableau V.3.1 : détail du calcul d'un type d'émission. 164
Tableau V.3.2 : récapitulatif des fonctions de traitement de l'information. 166
Tableau V.4.1 : récapitulatif des mécanismes d'apprentissage. 172
Tableau V.4.2 : TPT initiale. 173
Tableau V.4.3 : renforcement, introduction et oubli. 174
Tableau V.4.4 : glissement. 175
Tableau V.4.5 : fonction d'association. 175
Tableau V.4.6 : mesures de similitude entre les archétypes de $Arch_2$ 176
Tableau V.4.7 : TPT finale. 177
Tableau V.6.1 : TPT utilisée dans les exemples. 187
Tableau V.6.2 : récapitulatif des paramètres. 188
Tableau VI.1.1 : valeurs des paramètres du modèle de boucle thalamo-corticale. 210
Tableau VI.2.1 : types utilisés pour modéliser les 5 différents stimuli. 219
Tableau VI.2.2 : TPT des nœuds d'entrée (IGN) des deux processeurs. 219
Tableau VI.2.3 : valeurs des paramètres identiques pour le modèle témoin et le modèle dyslexique. 221
Tableau VI.2.4 : valeurs des paramètres différents pour le modèle témoin et le modèle dyslexique. 222

REMERCIEMENTS

Je tiens tout d'abord à remercier chaleureusement :

Salem Benferhat, José Bernardo, et Emmanuel Guigon pour avoir accepté d'être mes rapporteurs, pour la pertinence de leurs remarques, et pour avoir fait le déplacement jusqu'à Toulouse.

Claudette Cayrol et Henri Prade pour avoir accepté d'examiner et de commenter mon travail, et pour la qualité de leur enseignement.

-

Ma gratitude va également à tous les membres de l'unité Inserm 455. Faire une liste exhaustive des personnes à remercier serait bien évidemment impossible, mais je tiens à remercier tout particulièrement :

Josette Pastor pour avoir dirigé mon travail et m'avoir accordé énormément de temps et d'attention tout le long de mon DEA et de ma thèse ;

Pierre Celsis pour sa gentillesse, pour m'avoir accueilli dans son équipe, conseillé, et pour avoir accepté de faire partie de mon jury ;

François Chollet pour m'avoir accueilli au sein de l'unité 455 de l'Inserm ;

Jean-luc Nespoulous pour son soutien en tant que directeur de l'IFR des sciences du cerveau de Toulouse ;

Irène Delcroix pour être la pierre angulaire de l'unité ;

Jean-François Démonet, Dominique Cardebat, Bernard Doyon, Nicolas Chauveau, Florent Aubry, Kader Boulanouar, Chantal Blanchard, Gérard Viallard, Isabelle Loubinoux et pour leurs conseils avisés, leurs critiques constructives, et l'éclectisme des discussions du déjeuner ;

Patrice Perran, Sandra Lê et Serge Ruff pour les tournois mémorables de X-blast (et Oliver Vogel pour avoir créé ce jeu irremplaçable) ;

Xavier Franceries et Sébastien Basan pour être des sources inépuisables de mails humoristiques (de qualité très variable, certes) ;

Charlotte Momaur, Anne Léger, Élodie Goutines et Gaëlle Raboyau pour apporter la touche féminine du bureau (non, ça ne vise pas que la décoration dudit bureau) ;

Xavier de Boissezon, Angélique Gerdelat, Cyril Pernet, Emmanuelle Cassol, Sébastien Balduyck, Hélène Gros, Guillaume Thierry, Maria Traranino, Vanessa Camus, Sébastien Tresseras, Clara Martin, Jessica Faivre et tous les étudiants passés ou actuels de l'unité, pour ces contacts humains qui ne m'ont pas fait regretter d'avoir choisi un laboratoire aussi atypique pour une thèse d'informatique ;

Bertrand Lacotte et Marc Lafon pour le travail remarquable qu'ils ont effectué avant moi sur ce projet, et qui m'a servi de base solide.

-

Je remercie également toute ma famille et mes amis. Là aussi, faire une liste complète serait bien trop ardu, je me contenterai de citer :

Christian, Marie-Claude et Élise Labatut, qui seront toujours là ;

Stéphanie Monsérié, Jérôme Guilhem, Béatrice Pettes, Davy Capéra, Nicolas Idier, François Lartigau, David Duwou et Benjamin Gineste : leur amitié est un soutien de chaque instant ;

David Fauthoux, Jean-Pierre Georgé, Gauthier Picard, Laurent Prévôt, David Chemouil et Marc Boyer, pour leur amitié et leurs conseils plus professionnels...

A la mémoire de mon père

L'identification du rôle des organes qui composent le corps humain, et le problème inverse qui consiste à localiser anatomiquement le support biologique d'une fonction, occupent une partie des recherches de l'homme depuis l'antiquité. Suivant les évolutions de ces recherches, le siège de la pensée a successivement été placé dans divers endroits du corps humain (le cœur a, par exemple, connu un grand succès), avant de se stabiliser finalement, à l'époque de la renaissance, au niveau du cerveau. Il s'en est suivi une période d'intérêt pour cet organe, qui s'est traduite par de grandes avancées en ce qui concerne l'anatomie du système nerveux central. Ce n'est qu'au XIX^{ème} siècle, avec les débuts de la neuropsychologie, symbolisés notamment par les travaux de Paul Broca [Broca '61], que l'étude des relations entre l'anatomie du cerveau et sa fonction a réellement débuté. Les premiers outils de la neuropsychologie étaient uniquement d'ordre analytique, ils reposaient sur l'étude de cas cliniques. Le neuropsychologue tentait d'expliquer le rapport entre le comportement anormal observé chez un individu et la lésion neurologique présentée par celui-ci. Mais les technologies développées au cours du XX^{ème} siècle ont permis la définition d'une nouvelle méthode d'étude expérimentale : les études en activation. Les techniques de neuroimagerie fonctionnelle rendent désormais possible l'observation non-invasive (i.e. sans traumatisme) de l'activité du cerveau *in vivo*, pendant la réalisation d'une tâche cognitive.

Ces nouvelles techniques ont permis de mieux connaître l'organisation fonctionnelle du cerveau, mais elles ont avant tout exposé toute la complexité du traitement cérébral. On suppose aujourd'hui que la résolution d'une tâche cognitive donnée implique la mise en œuvre de tout un réseau d'aires cérébrales. Chaque aire est une population de neurones aux caractéristiques très variables. Dans un réseau donné, une aire possède un rôle bien particulier, indispensable à la bonne marche du réseau. De plus, une aire peut faire partie de plusieurs réseaux différents. Les études actuelles en neuropsychologie tentent de définir les liens entre les fonctions cognitives réalisées et les réseaux sous-jacents. Cependant, ce travail est rendu difficile par le fait qu'un réseau d'aires peut intervenir dans la résolution de plusieurs tâches différentes, et inversement, une tâche peut nécessiter l'activation de plusieurs réseaux.

La seule utilisation des techniques de neuroimagerie ne permet pas de résoudre ce type de problème, car les résultats obtenus ne reflètent que la conséquence de la mise en œuvre des mécanismes (l'activation de certaines aires cérébrales), et pas la façon dont ces mécanismes fonctionnent. Pour y remédier, une approche possible est d'interpréter ces mécanismes en termes de traitement de l'information cérébrale, au haut niveau qui est celui des réseaux d'aires cérébrales. On considère qu'à ce niveau, les mécanismes cérébraux sont l'intégration de mécanismes de plus bas niveau, qui sont étudiés par la neuroanatomie, la neurophysiologie et la neurobiologie, chez l'homme et chez l'animal. Toute la difficulté consiste à fusionner ces informations que l'on peut qualifier d'hétérogènes (car caractérisées par différentes échelles, espèces, etc.), de manière à les utiliser dans un but interprétatif. Cette intégration passe, à notre sens, par l'utilisation de modèles informatiques, qui permettraient en outre de compenser l'absence d'observation directe du fonctionnement cérébral.

La plupart des travaux en modélisation cérébrale sont répartis sur trois champs de recherche : la neuroimagerie, les sciences cognitives, et les neurosciences computationnelles. Dans le premier cas, les modèles s'appuient sur des techniques statistiques. Dans le deuxième cas, la plupart des travaux utilisent des techniques symboliques d'intelligence artificielle. Enfin, les neurosciences computationnelles sont largement dominées par les réseaux de neurones formels. Toutefois, aucun de ces modèles n'a pour objectif de résoudre le problème décrit auparavant, i.e. l'interprétation de données de neuroimagerie en termes de traitement de l'information cérébrale par le biais de modèles biologiquement plausibles.

Pour combler cette lacune, le projet de neurosciences computationnelles *MITIC* (*M*odélisation du *T*raitement de l'*I*nformation *C*érébrale) a été mis en place par l'unité 455 de l'Inserm, dans le but de définir un cadre formel de modélisation adapté. Cela signifie que le formalisme doit permettre de représenter explicitement l'information cérébrale et les mécanismes de traitement au niveau intégré des réseaux d'aires. De plus, le formalisme doit rendre possible l'intégration des différentes sources d'information précédemment décrites (neurophysiologie, etc.). Enfin, les connaissances en neurosciences étant sujettes à une perpétuelle évolution et à des remises en cause incessantes, le formalisme ne doit pas être rigide, et au contraire permettre de facilement adapter les modèles aux nouvelles découvertes.

La première version du formalisme, nommée BioCaEn, a fait l'objet de la thèse de Marc Lafon [Lafon '00]. BioCaEn repose sur une approche causale semi-qualitative des relations entre aires cérébrales, et sur une représentation duale et imprécise de l'information cérébrale. Toutefois, il est

restreint à la modélisation de mécanismes cérébraux automatiques, et ne contient pas de mécanismes d'apprentissage. De plus, il souffre de certaines limitations inhérentes à la méthode de la simulation semi-qualitative, utilisée pour la propagation de l'information, ce qui rend nécessaire la définition de nouveaux mécanismes de propagation. Le travail qui est présenté ici se situe à ce niveau du projet MITIC. Notre but est d'abord de choisir une autre méthode de propagation, afin de palier les défauts de la simulation qualitative rencontrés par BioCaEn. D'autre part, nous devons également modifier et développer le formalisme, aussi bien au niveau de la représentation de l'information qu'au niveau de son traitement, de façon à y inclure des mécanismes d'apprentissage.

En raison de la nature pluridisciplinaire de cette thèse, nous avons inclus plusieurs états de l'art, qui constituent les trois premiers chapitres. Nous avons tenté de les rendre aussi accessibles que possible aux personnes étrangères au domaine, ce qui explique la présence de certaines parties qui pourront paraître superfétatoires aux yeux du spécialiste. Le premier chapitre est dédié à la description anatomique et fonctionnelle du cerveau, ainsi que des techniques de neuroimagerie qui permettent d'étudier son fonctionnement. Le second chapitre fait l'état de l'art des techniques de modélisation cérébrale existantes, afin de situer notre approche par rapport à ces travaux. Le troisième chapitre fait l'état de l'art des formalismes causaux existants, et détermine quel est le plus adapté à nos contraintes. Les quatrième et cinquième chapitres constituent la description de RAGE (Réseaux Artificiels à Grande Echelle), notre outil de modélisation (son implémentation est donnée dans l'annexe B). Le sixième chapitre présente deux applications de notre outil à la modélisation de réseaux d'aires cérébrales. Enfin, nous concluons avec une critique de notre travail, et une discussion des futures évolutions possibles.

Chapitre I

NEUROSCIENCES INTEGRATIVES

Le cerveau humain est un système éminemment complexe et mal connu. On distingue principalement deux types d'approches pour le décrire : le niveau structurel et le niveau fonctionnel. Dans le premier cas, le cerveau est considéré comme un ensemble de structures biologiques interconnectées. Cela peut se faire à différentes échelles spatiales, de la région cérébrale au neurone. Dans le second cas, le cerveau est vu comme un ensemble de fonctions qui interagissent. Là aussi, l'étude peut être menée à différentes échelles, de la fonction cognitive très complexe à la primitive fonctionnelle. Les deux aspects sont liés, bien sûr, dans le sens où le niveau anatomique est le support physique du niveau fonctionnel. La neuroimagerie fonctionnelle est le principal outil permettant d'étudier ce lien entre fonction et structure. Toutefois, cet outil est imparfait. Tout d'abord, ses échelles spatiale et temporelle ne sont pas forcément compatibles avec l'étude de l'évolution rapide de l'activité neuronale, ou de sa localisation spatiale précise. D'autre part, ce n'est pas directement l'activité neuronale qui est directement observée grâce aux techniques de neuroimagerie, mais certaines de ses manifestations.

Afin de faire le lien entre ces connaissances hétérogènes du cerveau, l'utilisation d'un outil de modélisation s'impose. Dans ce chapitre, nous allons dans un premier temps décrire les principales propriétés anatomiques et fonctionnelles du cerveau. Puis, nous aborderons les caractéristiques des techniques de neuroimagerie. Grâce à ces informations, et en tenant également compte de nos objectifs de modélisation, il nous sera alors possible de déduire un certain nombre de contraintes qu'un outil de modélisation doit respecter.

1. CERVEAU

Dans cette partie, nous abordons les descriptions structurelle et fonctionnelle du cerveau. Elles n'ont pas pour but d'être exhaustives, mais de donner suffisamment d'informations pour justifier les choix de modélisation faits par la suite.

1.1. Description anatomique

1.1.1. Système nerveux central

Le système nerveux humain est subdivisé anatomiquement en un *système nerveux central* (SNC), comprenant l'*encéphale* et la *moelle épinière*, et un *système nerveux périphérique*, réunissant les nerfs qui parcourent le reste du corps. La description suivante est courte, et consacrée au seul système nerveux central, et plus particulièrement à l'encéphale, qui est la partie d'intérêt dans le cadre de ce travail de modélisation.

Le SNC est extrêmement important, puisqu'il s'agit en quelque sorte du centre de commande du corps humain. En conséquence, il est très protégé, au moyen de plusieurs couches anatomiques : une protection osseuse tout d'abord, puis une triple protection membranaire comprenant successivement la dure-mère, l'arachnoïde, et la pie-mère. Outre leur rôle protecteur, ces différentes couches constituent également des obstacles à la bonne observation de l'activité cérébrale via les techniques de neuroimagerie.

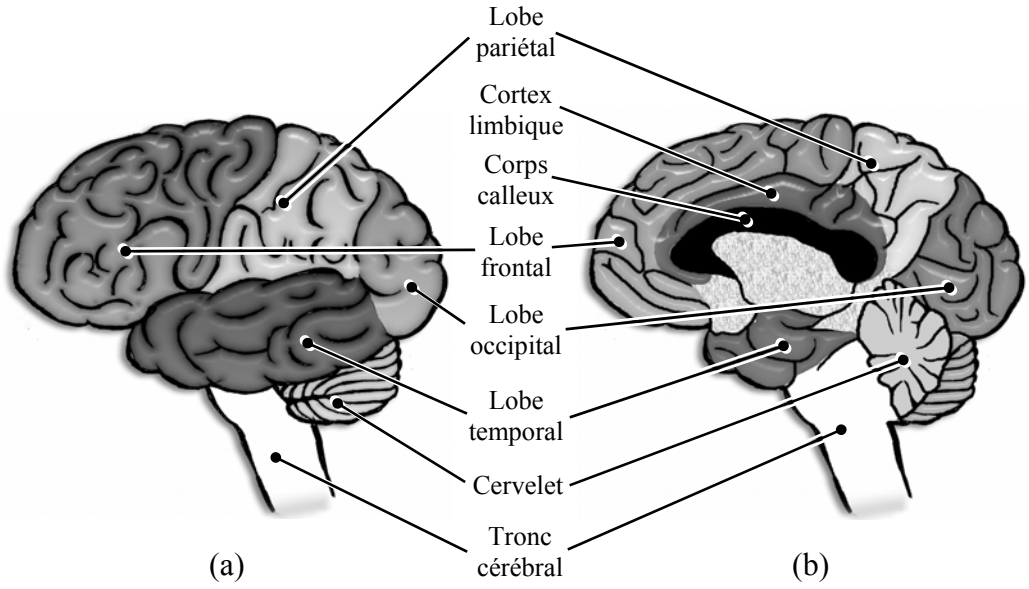

Figure I.1.1 : anatomie externe de l'encéphale, découpage en lobes.
(a) face externe gauche. (b) : face interne droite.

L'encéphale se découpe en trois parties : le cerveau, le tronc cérébral et le cervelet (Figure I.1.1). Le *cervelet* est situé en arrière de la boîte crânienne, et il est constitué de deux hémisphères cérébelleux. Il est relié (entre autres) aux muscles du corps via la moelle épinière. Il a essentiellement un rôle

dans la motricité, il permet de coordonner les mouvements volontaires, mais aussi involontaires : gestion de l'équilibre et de la posture. Il constitue un centre d'intégration de l'information motrice et de décision quant au mouvement à effectuer. Il est également impliqué dans l'apprentissage, en particulier l'apprentissage moteur [van Mier '00]. Le *tronc cérébral* est situé sous le cerveau et assure la liaison avec la moelle épinière.

Le *cerveau* constitue la plus grosse partie de l'encéphale. Il est creusé de cavités appelées ventricules cérébraux. Il se compose de deux *hémisphères cérébraux* qui concentrent l'essentiel de l'activité cérébrale en termes de fonctions cognitives.

La surface des hémisphères cérébraux est appelée *cortex cérébral*. Elle est extrêmement plissée, formant de nombreuses circonvolutions (les *gyri*) séparées par des sillons (les *scissures*). Ces séparations anatomiques permettent de découper chaque hémisphère cérébral en quatre lobes (Figure I.1.1.a) : *frontal* (au niveau du front), *temporal* (au niveau des tempes), *pariétal* (dessus du crâne), *occipital* (arrière du crâne). Ces quatre lobes constituent le *néocortex*. Sur la face interne de chaque hémisphère, on trouve une dernière structure corticale (Figure I.1.1.b), qui n'est pas considérée comme un lobe. Il s'agit du *cortex limbique*.

Les hémisphères cérébraux contiennent des structures neuronales, dites sous-corticales. On trouve notamment les *ganglions de la base* (comprenant entre autres la *substance noire* et le *corps strié*, lui-même composé du *striatum* et du *pallidum*), l'*amygdale*, le *thalamus*, et l'*hypothalamus*. Ces structures sont désignées globalement sous le nom de *noyaux gris centraux*. Les noyaux gris centraux sont impliqués dans la mémoire, le traitement d'informations sensorielles, et le contrôle de la motricité. L'*hippocampe* est également une structure sous-corticale, mais il ne fait pas partie des noyaux gris centraux. Le *système limbique* est un réseau de structures comportant : le cortex limbique, l'hypothalamus, l'amygdale, ainsi qu'une partie des ganglions de la base, une partie du thalamus, une partie de l'hippocampe, et une partie du cortex préfrontal.

1.1.2. Neurones

Le neurone ou cellule nerveuse est l'unité fondamentale du système nerveux, bien que celui-ci ne soit pas constitué uniquement de ce type de cellules. Il diffère beaucoup des autres cellules du corps humain par sa structure et son fonctionnement. Sa principale particularité est sa capacité à conduire les impulsions électriques qui constituent une des formes de l'information traitée par le système

nerveux. Il existe de nombreux types de neurones caractérisés par des formes, des fonctionnements, des interconnexions et des localisations variables.

On observe toutefois des propriétés communes. Un neurone se décompose anatomiquement en trois parties : le soma, les dendrites et l'axone (Figure I.1.2). Le *soma* est le corps cellulaire, contenant le noyau de la cellule. Les *dendrites* sont des prolongements ramifiés autour du corps cellulaire, qui permettent au neurone d'établir des connexions avec d'autres cellules (pas forcément des cellules nerveuses). Ces connexions sont appelées *synapses*. L'*axone* est également un prolongement du corps cellulaire, mais il est unique et beaucoup plus long que les dendrites. Il prend naissance au niveau d'une région particulière du soma appelée *cône d'emboîtement*. En général, il supporte des ramifications, notamment à son extrémité (opposée au corps cellulaire). Il se termine par des boutons synaptiques, qui constituent des connexions (synapses) avec les dendrites d'autres neurones. Un neurone est ainsi connecté à plusieurs milliers d'autres neurones.

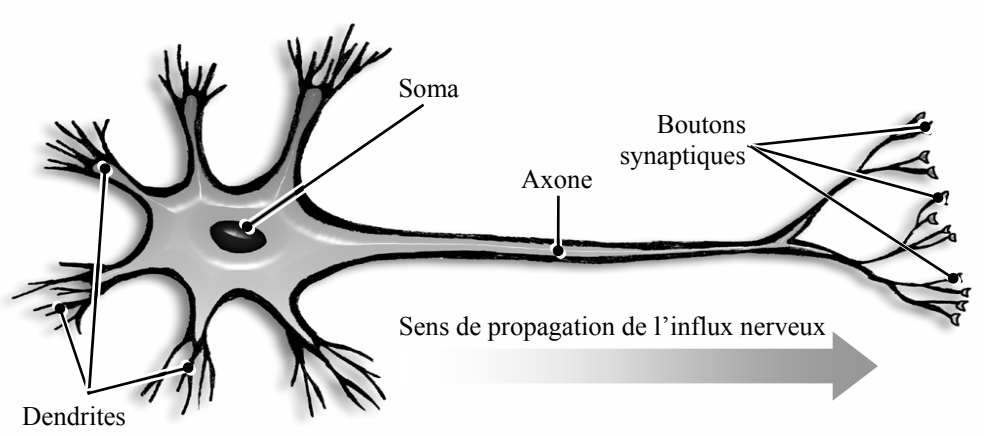

Figure I.1.2 : anatomie du neurone biologique.

L'influx nerveux, qui prend la forme d'une impulsion électrique appelée *potentiel d'action*, pénètre dans le neurone par les dendrites. En général, plusieurs influx nerveux arrivent à la fois. Le soma intègre spatialement et temporellement ces différents potentiels post-synaptiques, on parle *d'activité synaptique*. Si cette activation dépasse un certain seuil, l'influx nerveux se propage vers les terminaisons axonales via l'axone. On dit alors qu'il y a *décharge*. Au niveau des synapses, la transmission de l'influx nerveux se fait de façon chimique, via la libération de substances appelées neurotransmetteurs (ou neuromédiateurs). Les neurones peuvent se différencier par la capacité à être émetteurs ou récepteurs de neurotransmetteurs différents. Une synapse est soit excitatrice, soit inhibitrice, suivant qu'elle contribue à monter ou à baisser le potentiel d'action du neurone situé en aval.

Le corps cellulaire et l'axone présentent des différences de coloration : grisâtre pour le premier et blanchâtre pour le second. Or, les somas des neurones sont concentrés dans certaines zones telles que le cortex et le thalamus. On distingue donc dans le SNC la *substance grise* (les corps cellulaires) de la *substance blanche* (les liens entre les neurones). La substance blanche peut être organisée en faisceaux de fibres (d'axones) constituant des liaisons distales entre régions. Un exemple en est le corps calleux qui est une nappe de fibres assurant les liaisons entre les deux hémisphères. Cette organisation permet aux neurones d'être connectés à des neurones très éloignés de leur soma.

1.1.3. Populations de neurones

a. Découpage anatomique

A mi-chemin entre le niveau global et le niveau cellulaire, il est possible de considérer le cerveau au niveau du groupe de neurones. On appelle *modules* les populations élémentaires de neurones dont la structure et la fonction sont toutes les deux précisément définies [Arbib '85a]. Les colonnes corticales constituent un bon exemple de modules.

Le cortex peut être décomposé en six couches caractérisées par certaines organisations et densités neuronales, et par le type des neurones composant la couche : on parle de *cytoarchitecture* [Roland & Zilles '98]. Une *colonne corticale* est une structure perpendiculaire à la surface du cortex, dont la base possède une surface très réduite, et dont la hauteur est équivalente à l'épaisseur du cortex : la colonne s'étend donc sur les six couches corticales. Les colonnes ne sont pas toutes identiques, les densités des différentes couches varient suivant la localisation de la colonne sur le cortex.

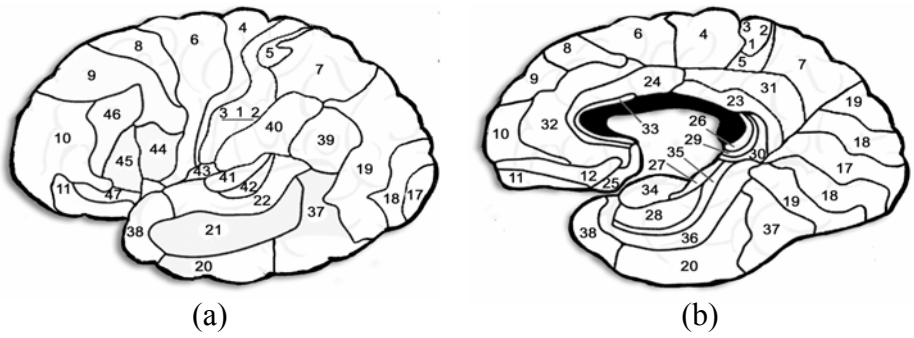

Figure I.1.3 : découpage anatomique du cerveau en aires de Brodmann.
(a) : vue externe. (b) : coupe sagittale médiane.

L'étude de ces variations a permis de découper le cortex en un ensemble de 52 aires anatomiques ou régions (Figure I.1.3), dites de *Brodmann* [Hedaen '72; Zilles '96]. Une aire corticale au sens anatomique est donc une population bien définie de colonnes corticales dont les stratifications possèdent des propriétés communes. Deux aires différentes peuvent avoir une même cytoarchitecture.

A noter que la cytoarchitecture n'est pas la seule propriété physique utilisée pour définir les variations corticales. Certaines études tiennent compte de la myéloarchitecure (propriétés de la couche de myéline des neurones), de l'architecture des récepteurs synaptiques, de la distribution de certaines composantes chimiques (enzymes, protéines,…), etc [Roland & Zilles '98].

b. Architecture en réseau

Le cerveau est un vaste réseau de neurones organisé en sous-réseaux cohérents, dont l'unité est fonctionnelle, à différents niveaux.

Les neurones impliqués dans la transmission d'un neuromédiateur particulier constituent des réseaux largement distribués dans l'ensemble du cortex, au sens où ces neurones sont présents dans de nombreux modules distincts et parfois très éloignés les uns des autres. Par exemple, les neurones *GABAergiques*, qui ont un rôle dans les mécanismes d'inhibition, et les neurones *dopaminergiques* qui jouent un rôle dans l'activation motrice. Il existe des liaisons anatomiques massivement parallèles [Arbib '95] entre les neurones de même type appartenant à des modules différents. C'est par exemple le cas pour les réseaux *sérotoninergiques* centraux [Loubinoux *et al.* '99; Pariente *et al.* '01] qui jouent un rôle actif dans la cognition.

L'examen du cortex au niveau modulaire, notamment en ce qui concerne les colonnes corticales, laisse également apparaître une organisation en réseau. D'une part, chaque couche d'une même colonne communique avec les couches directement voisines (on parle de relations intra-colonne). Et d'autre part, les colonnes communiquent entre elles au moyen de relations dites inter-colonne. Ces dernières connectent deux couches appartenant à deux colonnes distinctes mais situées au même niveau, ou bien à des niveaux différents [Felleman & Van Essen '91]. Ce type de liaison peut concerner aussi bien des colonnes très proches que des colonnes situées dans des aires très éloignées [Burnod '91; Guigon *et al.* '94]. Enfin, il existe également des liaisons entre des couches particulières des colonnes corticales et des structures sous-corticales telles que le thalamus.

A un niveau plus élevé, la combinaison d'études anatomiques et d'études en activation a également mis en évidence une organisation en réseau, utilisant des liaisons axonales comme liens d'interconnexion. Cette organisation existe au niveau des aires cérébrales [Bressler '95; Fischbach '92; Posner *et al.* '88; Pulvermüller '96], mais elle concerne également des structures intermédiaires, c'est-à-dire à la fois plus complexes que les modules, mais de taille inférieure à celle d'une aire cérébrale [Goldman-Rakic '88]. On désigne ces réseaux sous le nom de *réseaux à grande échelle*. Un réseau de se type se décompose en une multitude de réseaux parallèles, du type de ceux observés au niveau des colonnes corticales. Les aires constituant un réseau ne sont pas forcément proches les unes des autres. En fait on distingue deux types de liaisons : certaines sont locales, elles concernent deux aires contiguës, et d'autres sont distales, ce sont des faisceaux d'axones reliant des aires éloignées (Figure I.1.4).

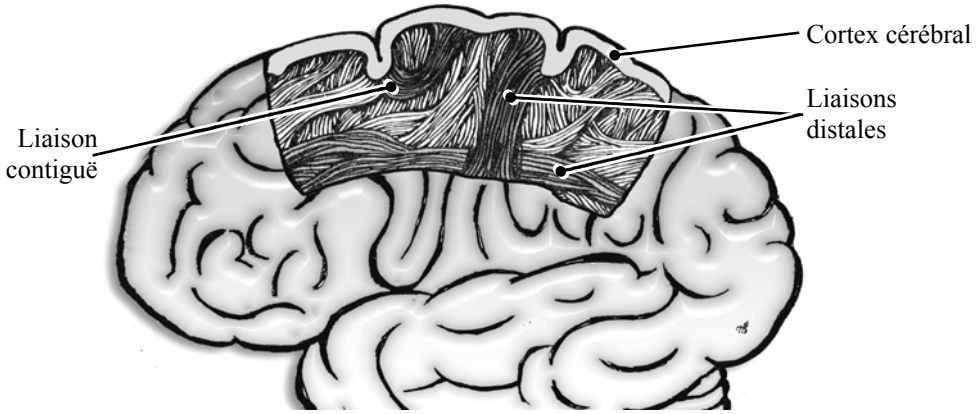

Figure I.1.4 : faisceaux d'axones entre des aires corticales.

Ces réseaux ne se limitent pas aux aires corticales, ils sont aussi connectés à des structures sous-corticales telles que le thalamus, l'hippocampe et les noyaux gris en général [Alexander *et al.* '92; Raichle '93]. De plus, il peut y avoir des recouvrements, c'est-à-dire qu'une aire peut appartenir à plusieurs réseaux différents [Ingvar & Petersson '00]. Finalement, le cerveau se présente sous la forme d'un ensemble de réseaux parallèles mais interconnectés d'aires cérébrales [Bressler '95].

1.2. Description fonctionnelle

Puisqu'il existe des différences anatomiques entre certains modules, il est logique de supposer que leur comportement peut aussi varier. Ainsi, en ce qui concerne les colonnes corticales, même si on distingue des traits communs tels que la présence de relations inter-colonne inhibitrices ou

excitatrices, la présence de mécanismes d'amplification ou une activité motivée par la recherche d'un état de stabilité [Burnod '91; Guigon *et al.* '94], la fonction spécifique implémentée par une colonne dépend de ses relations internes et externes, donc des propriétés de ses couches.

Une aire anatomique est un réseau de colonnes corticales. La fonction implémentée par un réseau dépend à la fois de la topologie du réseau, et des propriétés fonctionnelles des éléments qui le constituent. Par conséquent, si les fonctions des colonnes corticales varient suivant leurs caractéristiques anatomiques, alors on peut faire l'hypothèse que les fonctions d'aires cérébrales ayant des cytoarchitectures différentes varient également. De même pour les structures sous-corticales, dont les neurones présentent également des différences d'ordre anatomique.

1.2.1. Découpage fonctionnel

Les études en activation (c.f. paragraphe 2.4), qui sont venues compléter les études anatomiques, ont été utilisées pour valider cette hypothèse, en permettant un découpage fonctionnel partiel du cortex cérébral. Ce découpage tente d'associer des régions à des fonctions cognitives ou sensorimotrices spécifiques. Les régions fonctionnelles peuvent différer des régions obtenues par un découpage purement anatomique [Mesulam '90] tels que celui de Brodmann. Une des raisons est qu'il n'est pas possible de prendre en compte toutes les différences physiques [Roland & Zilles '98] lors du découpage anatomique (seule la cytoarchitecture est considérée dans le cas des aires de Brodmann). De ce fait, les frontières fonctionnelles sont imprécises, d'autant plus que le découpage anatomique est lui-même soumis aux variations anatomiques du cortex existant d'un individu à l'autre.

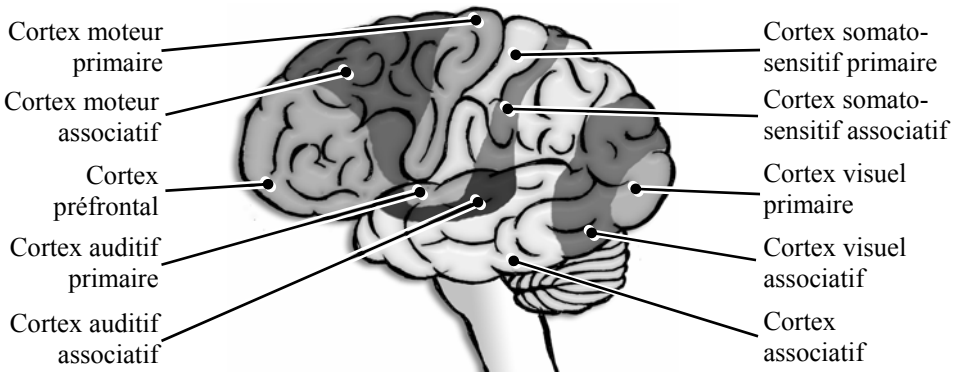

Figure I.1.5 : découpage fonctionnel du cortex cérébral.

Néanmoins, les études cliniques, animales, et en activation ont permis de réaliser une cartographie fonctionnelle de certaines parties du cortex, découpant celui-ci en aires primaires, aires associatives et aires pré-frontales (Figure I.1.5). Le fait que certaines de ces aires corticales soient anatomiquement reliées à des structures sous-corticales montre l'importance fonctionnelle de ces dernières.

a. Aires primaires

Les *aires primaires*, comme leur nom l'indique, sont les premières aires du cortex en relation avec l'extérieur du système nerveux : muscles (aires primaires motrices), récepteurs sensoriels (aires primaires sensorielles), elles constituent une sorte d'interface. En réalité, ce contact n'est pas direct puisqu'il se fait en général via le thalamus, qui réalise un premier traitement de l'information.

Les aires primaires sont caractérisées par une organisation *topique*, c'est-à-dire en correspondance avec l'organe sensoriel ou moteur qui leur est relié. Fait intéressant, cette organisation particulière se retrouve aussi, quoique de façon moins marquée, dans des aires situées en aval dans les circuits de traitement de l'information [Alexander *et al.* '92].

Les cortex moteur et sensitif primaires, qui participent au contrôle musculaire, possèdent une organisation *somatotopique* [Churchland & Sejnowski '92], c'est-à-dire qu'à chaque partie du corps pouvant être bougée ou pouvant ressentir le toucher, correspond une partie des cortex moteur et sensitifs primaires. De la même façon, on qualifie de *tonotopique* l'organisation des aires auditives primaires, car on peut réaliser un découpage en sous-populations, chacune réagissant à un certain intervalle de fréquences sonores (i.e. tons) [Ardila '93]. Les aires primaires visuelles, elles, reproduisent la configuration de la rétine, on parle d'organisation *rétinotopique* [Arbib *et al.* '98].

L'organisation *topique* des aires primaires montre l'influence des stimuli sur la spécialisation fonctionnelle des aires et permet de comprendre pourquoi les structures anatomiques et fonctionnelles ne sont pas complètement juxtaposables. Les aires primaires sont bien connues car elles sont en communication directe avec l'extérieur, c'est-à-dire qu'un stimulus va forcément avoir des effets sur les aires primaires associées, et qu'une lésion dans une de ces aires va obligatoirement avoir des répercussions sur la perception ou la motricité.

Les relations des aires primaires avec les aires associatives correspondantes sont bien connues. Cependant, des travaux récents montrent, chez le primate non humain, des connexions directes

entre les aires primaires visuelles et auditives et une intégration multimodale (i.e. à la fois visuelle et auditive) de l'information dans ces aires primaires [Falchier *et al.* '02].

b. Aires associatives

Les *aires associatives* prennent le relais des aires primaires dans le traitement de l'information. Elles représentent la majeure partie du cortex cérébral chez l'homme. L'information qu'elles reçoivent a déjà été traitée une première fois par les aires primaires. Les aires associatives assurent un degré plus élaboré de traitement des informations, dans le sens où elles permettent d'analyser, de reconnaître et de fusionner des messages sensoriels arrivant de différentes aires primaires.

De même qu'il y a des aires primaires motrices ou correspondant aux différents sens, on distingue plusieurs types d'aires associatives. Les aires associatives motrices sont chargées de gérer les mouvements volontaires et l'anticipation du mouvement [Duvernoy *et al.* '92]. Certaines aires associatives sensorielles sont dédiées à un seul type d'information : fusion d'informations visuelles pour la reconnaissance d'objets, interprétation tridimensionnelle de la nature d'objets par stimuli tactiles, etc. D'autres aires associatives sont multimodales, c'est-à-dire qu'elles intègrent des informations de plusieurs types. Par exemple, le cortex pariétal postérieur intègre des entrées sensorielles de toutes les modalités, afin de construire la représentation spatiale d'un objet [Andersen *et al.* '97; Bremmer *et al.* '01]. Les traitements complexes réalisés par les aires associatives incluent également des mécanismes de mémorisation : mémoire associée au traitement du langage, mémoire visuelle ou spatiale [Duvernoy *et al.* '92].

Il est plus difficile de localiser ce type d'aires que ce n'était le cas pour les aires primaires. En effet, elles sont impliquées dans des mécanismes plus complexes, permettant d'accomplir des tâches cognitives de plus haut niveau. De plus, une même aire associative peut participer à plusieurs fonctions différentes. Une autre propriété importante des aires associatives est leur grande plasticité (c.f. paragraphe 1.3).

Les aires associatives entretiennent des relations avec les aires primaires, mais elles sont également liées aux aires pré-frontales et à certaines structures sous-corticales. Les premières exercent un certain contrôle sur les aires associatives. Les secondes implémentent des fonctions automatiques pouvant servir de base ou venir compléter les fonctions des aires associatives. Par exemple, le cortex associatif moteur est relié au thalamus, qui implémente des mouvements réflexes.

c. Aires pré-frontales

Les aires pré-frontales sont parfois considérées comme des aires associatives particulières. Elles ont toutefois un rôle plus orienté vers le contrôle et la coordination. Elles reçoivent des informations en provenance des aires primaires et associatives, mais elles sont également en liaison avec des structures sous-corticales qui matérialisent des rythmes biologiques internes. Ceci rend possible l'organisation temporelle du comportement : planification, anticipation. Leurs liens avec les structures sous-corticales liées à l'émotion pourraient expliquer le contrôle qu'elles exercent sur la personnalité et la régulation de l'humeur et du comportement.

On leur attribue le contrôle de fonctions cognitives élevées : raisonnement, attention, mémorisation, concentration intellectuelle [Duvernoy *et al.* '92].

d. Structures sous-corticales

Fonctionnellement, les structures sous-corticales sont responsables de l'organisation de comportements instinctifs fondamentaux et de l'expression des émotions et des motivations, assurant la protection de l'individu et la survie de l'espèce. C'est le domaine de l'instinct et des mécanismes automatiques.

De manière générale, le *système limbique* est associé aux réponses émotionnelles et à la mémoire. L'*hippocampe* est impliqué dans les mécanismes de mémoire et de traitement des émotions, ce qui explique que certaines de ses parties appartiennent au système limbique. Il est connecté, entre autres, au thalamus et à l'hypothalamus. L'*amygdale* est une structure complexe intervenant dans les réponses émotionnelles, aussi bien innées qu'acquises. Elle est reliée notamment au thalamus et au cortex préfrontal.

Le *thalamus* participe à de nombreux aspects de la perception : vision, audition, perception de la douleur. En fait, à l'exception de l'olfaction, toutes les entrées sensorielles passent d'abord par le thalamus avant d'atteindre le cortex. De ce fait, le thalamus entretient des liens (dans les deux sens) avec les aires primaires dédiées à ces sens (cortex primaires visuel, auditif et sensoriel). Il participe à l'intégration sensorielle de ces différentes modalités. Il intervient dans les fonctions motrices automatiques (réflexes) via ses connexions avec le corps strié et le cervelet. Il est également supposé participer au contrôle de l'attention en ce qui concerne les processus sensoriels faisant intervenir ces modalités. Il est connecté avec le cortex préfrontal, qui intervient dans les processus de contrôle

attentionnel. Enfin, le thalamus intervient dans la mémoire à travers ses liens avec l'amygdale et le cortex préfrontal.

Les *corps striés* sont impliqués dans le contrôle moteur et la planification des mouvements en ce qui concerne les membres externes et les muscles des yeux. Ils sont connectés, directement ou via le thalamus, au cortex moteur. Ils se projètent dans la *substance noire*, qui est également impliquée dans le contrôle des mouvements.

L'hypothalamus est le principal centre des fonctions végétatives, il régit la soif, l'appétit, la température, etc. Il intervient également dans les émotions, et contrôle la production de nombreuses hormones.

1.2.2. Organisation fonctionnelle

a. Réseaux fonctionnels

Les études expérimentales [Alexander *et al.* '92; Damasio '89; Mesulam '90] et des travaux de modélisation [Arbib '95; Rumelhart & Mc Clelland '86] ont montré que la résolution d'une tâche cognitive ou sensorimotrice donnée impliquait la mise en œuvre d'un réseau spécifique de régions cérébrales interconnectées, appelé *réseau cérébral à grande échelle* (RCGE). Les réseaux de ce type reposent sur les connexions anatomiques déjà décrites (paragraphe 1.1.3.b), et ils peuvent donc impliquer aussi bien des régions corticales que des structures sous-corticales [Bressler '95].

Il n'y a pas de bijection entre les RCGE et les fonctions cognitives de haut niveau qui sont implémentées [Démonet *et al.* '94; Raichle '93]. En d'autres termes, un réseau n'est pas forcément dédié qu'à une seule tâche, et la résolution d'une même tâche cognitive peut nécessiter l'activation de plusieurs RCGE différents [Mesulam '90; Pulvermüller '96; Sergent '94]. De plus, on sait que ces réseaux sont massivement parallèles et interconnectés. Il semble que certains mécanismes de contrôle permettent de coordonner les réseaux, d'en activer ou désactiver certaines parties [Bressler '95]. Il s'agirait de mécanismes dynamiques d'origine à la fois corticale et sous-corticale (notamment le thalamus), capables d'influencer sur les entrées-sorties et le comportement de régions cérébrales.

Un RCGE est un ensemble de régions cérébrales interconnectées, et on peut dire que la fonction complexe qu'est la tâche cognitive émerge du réseau de fonctions plus simples implémentées par les régions ou leurs sous-régions. Pour comprendre comment une tâche cognitive donnée est résolue par un RCGE, il faut donc non seulement connaître les différentes régions qui composent le réseau

concerné ainsi que sa topologie (aspect anatomique), mais aussi le rôle de chaque région au sein de ce réseau, et la nature des interactions existant entre ces régions (aspect fonctionnel).

A noter qu'une même région peut appartenir à plusieurs réseaux différents, et que sa fonction peut varier suivant le RCGE (et donc la tâche cognitive) qui la sollicite. Le rôle d'une région dépend donc du contexte d'activation [McIntosh '00]. Dans certains cas, cela peut s'expliquer par le fait que la région est décomposable en un ensemble de sous-régions fonctionnellement différentes. Les différents réseaux ne sollicitent qu'une de ces sous- régions, c'est-à-dire une partie seulement de la région entière [Goldman-Rakic '88]. Néanmoins, il a également été observé, dans certains cas, qu'une région ne puisse pas remplir plusieurs rôles en même temps si elle est sollicité par plusieurs réseaux à la fois [Roland & Zilles '98].

b. Primitives fonctionnelles

Considérons la fonction d'une région cérébrale. Cette fonction dépend des propriétés et de l'agencement des différents types de neurones qui la composent. Ces types de neurones forment des sous-populations neuronales. Il est parfois possible de distinguer, au sein d'une région cérébrale, des sous-populations neuronales plus ou moins précisément définies, mais ayant un rôle fonctionnel bien déterminé. Ces sous-populations peuvent regrouper des neurones sensibles à certains neurotransmetteurs, comme par exemple, les neurones GABAergiques qui implémentent une fonction d'inhibition. Elles peuvent aussi concerner des modules cérébraux, tels les colonnes du cortex primaire qui influencent les muscles fonctionnant en synergie pour accomplir un mouvement particulier. Enfin, la fonction peut être une simple hypothèse, et dans de cas-là, la population qui l'implémente n'est pas précisément identifiée. Le traitement des phonèmes (i.e. des sons élémentaires de la langue) dans l'aire de Wernicke en est une illustration.

On peut donc réaliser une décomposition fonctionnelle de la région, en découpant la fonction globale en primitives fonctionnelles ou en fonctions plus élémentaires qui interagissent. On obtient un réseau de primitives fonctionnelles décrivant de façon plus détaillée le comportement de la région. Des différences dans l'interaction et l'implication de ces primitives permet d'expliquer le fait qu'une même région puisse implémenter plusieurs fonctions différentes, suivant le RCGE considéré [Ingvar & Petersson '00].

1.3. Plasticité cérébrale

On appelle plasticité cérébrale la faculté qu'a le cerveau de se réorganiser physiquement et fonctionnellement pour s'adapter à un environnement. Cette propriété est à la base des processus de mémoire et d'apprentissage, mais elle intervient également pour compenser partiellement les dégâts causés par des lésions cérébrales. La plasticité est un élément important de la dynamique cérébrale, qui doit, à ce titre, être pris en compte lors de la modélisation.

1.3.1. Récupération fonctionnelle

L'étude des processus de réorganisation qui suivent l'apparition d'une lésion cérébrale a permis de détailler plusieurs mécanismes de récupération fonctionnelle au niveau cortical. On distingue ainsi les processus de vicariation, de redondance, de récupération de diaschisis et de substitution [Chollet '00].

Le mécanisme de vicariation est le plus observé. Une ou plusieurs régions différentes de la région lésée prennent le relais, et implémentent la fonction perdue, alors qu'elles n'étaient auparavant pas impliquées dans sa réalisation. Par exemple, dans l'étude de Weiller et collaboratuers (1995), il est demandé à un patient dont l'aire de Broca (située dans l'hémisphère gauche) est lésée, d'effectuer une tâche liée au langage. On observe une récupération réalisée en partie par vicariation sur cette aire, ce qui se traduit par une activation dans l'homologue de l'aire de Broca, situé dans l'hémisphère droit. La récupération de déficit pour les fonctions concernant le langage, qui sont fortement latéralisées à gauche (c'est-à-dire implémentées en grande partie par des aires situées dans l'hémisphère gauche) repose souvent sur l'utilisation des aires homologues de l'hémisphère droit [Chollet '00]. On pourrait donc penser que les structures analogues des deux hémisphères peuvent implémenter les mêmes fonctions, avec des degrés d'efficacité différents, probablement dus à la dissymétrie hémisphérique observée chez les sujets sains [Rae *et al.* '02].

On observe également parfois un glissement de la fonction sur des populations neuronales aux caractéristiques fonctionnelles proches. Par exemple, dans [Weiller *et al.* '93], à la suite d'une lésion dans l'aire motrice primaire, le contrôle des doigts est effectué en partie par une zone qui est normalement dédiée au contrôle des muscles du visage. Dans [Sadato *et al.* '96], des patients aveugles de naissance présentent des activations au niveau de ce qui devrait être leur cortex visuel primaire, lorsqu'ils sont soumis à des stimuli tactiles. Ceci signifie que des populations corticales habituellement réservées à la vision peuvent traiter d'autres modalités sensorielles [Chollet '00].

L'existence de projections du cortex auditif vers le cortex visuel [Falchier *et al.* '02] sont également en faveur de cette hypothèse.

La récupération par *redondance* s'appuie sur des neurones de l'aire lésée ayant été épargnés par la lésion. Ce type de récupération passe par la création de nouvelles connexions neurales. Le phénomène de *diaschisis* désigne une désactivation fonctionnelle de régions saines mais connectées à la région lésée. La *récupération de diaschisis* consiste à progressivement relâcher ces inhibitions corticales distales. Enfin, la *substitution* est une récupération fonctionnelle, et donc de plus haut niveau. Elle consiste à mettre en place des stratégies de remplacement qui viennent palier la perte fonctionnelle causée par la lésion.

Cependant, la réorganisation n'est pas un phénomène purement cortical. Certaines réorganisations corticales sont la conséquence d'une réorganisation s'effectuant à un niveau inférieur (structures sous-corticales). La présence de mécanismes à ce niveau, représentant une réorganisation moins importante qu'au niveau cortical, est une hypothèse largement acceptée [Chollet '00].

Les mécanismes de la plasticité cérébrale, essentiellement la vicariation, mettent en lumière le fait que différentes régions peuvent avoir les mêmes capacités fonctionnelles. Ceci est certainement dû à une organisation physique similaire [Arbib *et al.* '98; Burnod '91]. La spécialisation observée pour chaque aire semble être due aux facultés d'apprentissage, et dépend de l'information contenue dans les stimuli reçus par l'aire. On peut cependant supposer qu'il est possible de regrouper les aires en classes fonctionnelles.

1.3.2. Apprentissage et habituation

L'apprentissage et la mémoire sont des fonctions fondamentales du cerveau, qui permettent à l'individu de conserver son expérience et de s'adapter à un monde en perpétuelle évolution. D'un point de vue biologique, l'apprentissage dans un réseau de neurones est la conséquence de la propagation de stimuli, combinée à la plasticité du réseau [Ingvar & Petersson '00]. A un plus haut niveau, on peut définir l'apprentissage comme l'ensemble des processus grâce auxquels le cerveau se réorganise fonctionnellement et révise les informations stockées, en fonction de l'expérience acquise [Petersson *et al.* '97]. La réorganisation fonctionnelle peut se faire à différents niveaux : celui du réseau implémentant une fonction complexe, et celui de l'aire cérébrale, qui implémente une fonction plus élémentaire. Dans tous les cas, ils reposent sur les propriétés d'apprentissage présentes au niveau du neurone même.

a. Niveau du neurone

Le rôle d'un neurone est de recevoir en entrée les stimuli reçus de plusieurs neurones afférents, et de les combiner pour éventuellement émettre en sortie un unique stimulus. Les mécanismes de mémorisation au niveau du neurone se traduisent par des changements de la fonction implémentée par le neurone. L'apprentissage neuronal peut se décomposer en quatre processus caractérisés par des échelles temporelles différentes [Burnod '91].

Les délais de transmission de l'information le long de l'axone et à travers les synapses constituent de facto un premier mécanisme de mémorisation (de l'ordre d'une à cent millisecondes).

La transmission de l'information passe par des échanges ioniques. Les changements dans les concentrations ioniques affectent l'efficacité de la transmission synaptique (la rendant plus rapide ou plus lente). Ce second mécanisme permet une mémorisation à court terme (de l'ordre de quelques secondes à quelques minutes).

Une modification des récepteurs synaptiques et des canaux ioniques permet un troisième type de mémorisation, à plus long terme (un jour ou plus). Ce mécanisme est à la base du phénomène d'*habituation* : lorsqu'un neurone est soumis au même stimulus de façon répétée, sa réponse a tendance à baisser au fil du temps, du moins en ce qui concerne ce stimulus. Par contre, sa réponse est d'autant plus grande que le stimulus reçu ne ressemble pas au stimulus appris [Miller *et al.* '91]. Ce phénomène peut s'interpréter comme un mécanisme privilégiant l'information nouvelle au détriment de celle qui est déjà connue.

Enfin, de fortes coactivations répétitives et la maturation cérébrale (stabilisation des contacts synaptiques) permettent une mémorisation permanente.

b. Niveau du module

Au niveau du module cérébral, l'apprentissage neuronal a des répercussions sur la topologie fonctionnelle du réseau constitué par le module. Par exemple, dans le cas de la colonne corticale, certaines configurations de stimuli vont renforcer l'aspect excitateur ou inhibiteur des connexions inter-colonne, ou modifier les connexions intra-colonne [Burnod '91].

A un niveau plus élevé, l'aire cérébrale se compose d'un grand nombre de réseaux parallèles de colonnes corticales. Basiquement, tous les réseaux implémentent plus ou moins la même fonction, propre à l'aire cérébrale, avec un grand nombre de redondances (plusieurs réseaux remplissant la

même fonction). A la suite d'un processus d'apprentissage, plusieurs réseaux fonctionnels vont se spécialiser dans le traitement d'une catégorie d'information en particulier. Par la suite, ces réseaux pourront encore se spécialiser pour traiter des sous-catégories [Burnod '91].

c. Niveau de l'aire

En termes d'activation (mesurée par les techniques de neuroimagerie, c.f. paragraphe 2), l'apprentissage influe sur trois facteurs : l'intensité de l'activation, sa localisation, et l'interaction entre les différentes aires impliquées par la tâche apprise. Le phénomène d'habituation observé au niveau neuronal a des répercussions au niveau de l'aire cérébrale. La répétition d'un stimulus invariant entraîne une baisse de l'activation correspondant à la réponse de l'aire. Ce phénomène d'habituation de l'aire correspond à une optimisation progressive de sa réponse, dictée par la tâche à réaliser [Büchel *et al.* '99].

Un déplacement de la zone activée peut survenir lorsque la durée d'exposition au stimulus est plus importante [van Mier '00]. Par exemple, dans [Raichle *et al.* '94], on observe ce type de glissement au bout d'une dizaine de minutes de stimulation (génération de mots). Une hypothèse qui peut expliquer en partie ce phénomène est que l'apprentissage sur un petit nombre de répétitions est guidé par des indices provenant de l'extérieur (informations sensorielles), alors que la réalisation d'une tâche qui a commencé à être apprise fait plutôt intervenir des indices internes, préalablement mémorisés [van Mier '00].

Lorsqu'une tâche complexe est tellement apprise qu'elle est réalisée de façon plus ou moins automatique, il est possible d'observer une augmentation de l'activation, qui suit la diminution de l'activation observée en début d'apprentissage. Cette augmentation peut s'expliquer par un accroissement de la partie de l'aire cérébrale dédiée à la réalisation de cette tâche, pour une meilleure réalisation [van Mier '00]. Par exemple, dans [Pascual-Leone *et al.* '95], la pratique quotidienne du piano accroît la partie du cortex moteur dédiée au contrôle des doigts.

Au niveau du réseau d'aires cérébrales, ces changements dans les activations régionales s'accompagnent d'une réorganisation fonctionnelle, c'est-à-dire d'un changement dans les interactions des aires [Büchel *et al.* '99]. Cela se traduit par une répartition différente des activations dans les aires qui composent le réseau. Par exemple, le cortex préfrontal est très impliqué dans le fonctionnement de la mémoire de travail. Par conséquent, il risque d'être fortement sollicité au début de l'apprentissage d'une tâche. Par contre, son influence sur les autres aires du réseau

fonctionnel concerné sera moins importante pour la suite de l'apprentissage (du moins en ce qui concerne la mémoire de travail). Ces changements de répartition ne se limitent pas aux aires corticales, les structures sous-corticales peuvent également être concernées, par exemple, lors de l'apprentissage de mouvements [Penhune & Doyon '02; Schiltz *et al.* '01].

2. NEUROIMAGERIE FONCTIONNELLE

La neuroimagerie fonctionnelle constitue un outil relativement récent, permettant de quantifier de façon plus ou moins directe l'activité cérébrale à un niveau macroscopique, c'est-à-dire dans les populations de neurones [Ingvar & Petersson '00].

En dehors des techniques invasives, qui impliquent d'accéder physiquement au cerveau (électrodes implantées dans le cortex, stimulation corticale directe) et ne sont utilisées sur l'homme qu'en situation préopératoire, il existe principalement deux méthodes pour mesurer l'activité cérébrale chez l'être humain. Ces techniques de neuroimagerie réalisent des mesures indirectes de l'activation cérébrale, car elles sont basées sur l'étude de certaines conséquences physiques ou biologiques de l'activité du cerveau, et non pas sur l'activité elle-même. Les relations entre l'activité cérébrale et les conséquences mesurées sont encore mal connues, ce qui peut rendre difficile l'interprétation de résultats [Ingvar & Petersson '00]. Néanmoins, les techniques de neuroimagerie constituent un outil inestimable de l'étude fonctionnelle du cerveau.

La première de ces méthodes indirectes consiste à étudier les effets produits par les variations du champ électrique créé par l'activité des neurones : c'est le cas des *techniques de surface*. La propagation du potentiel d'action neuronal consomme beaucoup d'énergie, et les neurones ne possèdent pas de grandes réserves énergétiques. Par conséquent, lors d'une activation soutenue, une augmentation du débit sanguin a lieu à proximité des neurones concernés, pour leur apporter l'énergie nécessaire. L'étude des effets dus aux variations du débit sanguin représente la seconde méthode de mesure indirecte de l'activité cérébrale, appliquée dans les *techniques tomographiques*.

2.1. Techniques de surface

2.1.1. Electroencéphalographie et Magnétoencéphalographie

L'électroencéphalographie (EEG) est la plus ancienne des techniques de surface, puisque le premier tracé EEG a été relevé en 1929 par Hans Berger [Berger '29]. Elle consiste à mesurer le champ électrique induit par la propagation des potentiels d'action. Ces mesures sont effectuées au moyen d'électrodes (de 20 à 256) disposées sur le scalp du sujet. Les valeurs mesurées sont très faibles, de l'ordre de la centaine de microvolts. Afin d'améliorer les mesures, il est possible d'implanter des électrodes directement dans le cerveau (dans ce cas l'EEG devient invasive), mais cette méthode n'est employée que dans un cadre opératoire.

La *magnétoencéphalographie* (MEG) est plus récente, puisque la première utilisation est due à David Cohen en 1972 [Cohen '72]. Cette fois, c'est le champ magnétique créé par le champ électrique qui est mesuré, grâce à des capteurs très sensibles, à base de supraconducteurs. A l'instar de l'EEG, les valeurs mesurées sont très faibles, de l'ordre du femtotesla ($10^{-13}T$) [Gernero '01a]. On pourrait penser que les résultats obtenus sont moins précis que pour l'EEG, puisque l'on obtient des mesures encore plus indirectes de l'activité cérébrale : la propagation du signal neuronal crée un champ électrique qui cause lui-même un champ magnétique, qui est mesuré par la MEG. Mais le champ magnétique a la propriété d'être beaucoup moins distordu et absorbé que le champ électrique lors de la traversée de la boîte crânienne, ce qui permet finalement d'obtenir des résultats plus précis [Pernier *et al.* '92].

Relativement à l'échelle de la dynamique neuronale, la précision temporelle des techniques de surface est bonne, de l'ordre de la milliseconde. En revanche, la précision spatiale souffre de certaines limitations. D'une part, en raison de la nature des champs mesurés, il est impossible d'étudier l'activation des structures sous-corticales (thalamus, par exemple) [Gernero '01a]. D'autre part, en raison de le diffusion et de l'atténuation des signaux électromagnétiques, chaque capteur réalise une mesure locale (depuis sa propre position) de l'ensemble de l'activité corticale. Ceci rend nécessaire un certain traitement des données pour localiser les aires activées.

2.1.2. Traitement des données

Les techniques de surface enregistrent un signal correspondant à l'activité spontanée du cerveau, c'est-à-dire une activité qui a plusieurs origines : rythmes cérébraux physiologiques, activité

musculaire parasite, etc. L'activité que l'on veut vraiment mesurer, qui est souvent celle liée à une tâche cognitive précise, est noyée dans l'activité spontanée. Pour l'extraire, il est nécessaire de disposer d'un grand nombre de relevés (plusieurs dizaines).

La technique des *réponses évoquées (potentiels évoqués pour* l'EEG et *champs magnétiques évoqués* pour la MEG) consiste à calculer la moyenne de toutes les activités mesurées, après avoir synchronisé ces mesures. Cette superposition permet de distinguer l'activité ciblée du bruit que constitue l'activité spontanée, que l'on suppose plus aléatoire. Cette technique repose sur une hypothèse importante, qui est que les réponses cérébrales sont très reproductibles d'un essai à l'autre (voire d'un sujet à l'autre).

Après avoir appliqué ce traitement à chaque point de mesure (électrode ou capteur), il est possible d'avoir une représentation graphique des variations de champ (électrique ou magnétique) à la surface du crâne. L'étape suivante consiste à localiser spatialement les aires qui génèrent les champs observés, ce qui est un problème très complexe. On utilise pour ce faire un modèle plus ou moins réaliste du crâne, prenant en compte la géométrie et les propriétés de conduction des tissus cérébraux (Figure I.2.1). On place dans ce modèle un certain nombre de dipôles (i.e. des résistances électriques) qui représentent chacun une aire activée. La difficulté est de savoir, a priori, combien de dipôles placer et où les placer.

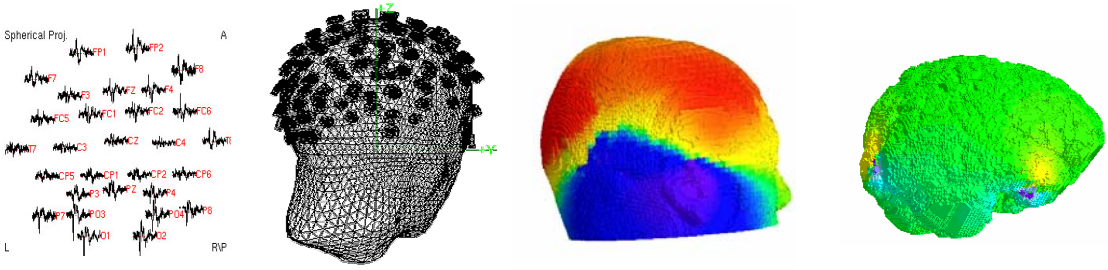

Figure I.2.1 : exemple de reconstruction d'une image de l'activation cérébrale par EEG [Franceries *et al.* '03].

Le problème de localisation se décompose alors en deux parties : le problème direct et le problème inverse. Le *problème direct* consiste à calculer le champ résultant du modèle précédemment spécifié [Gernero '01b]. Le *problème inverse*, ou reconstruction, consiste à ajuster les paramètres des dipôles du modèle (orientation, position, amplitude), de façon à faire correspondre les champs mesurés dans le modèle à ceux mesurés dans la réalité. En général, seule une petite partie des sources peut être

identifiée, ce qui explique pourquoi malgré une bonne résolution, les techniques de surfaces sont considérées comme moins efficaces que les techniques tomographiques en ce qui concerne la localisation spatiale [Horwitz *et al.* '99].

2.2. Techniques tomographiques

2.2.1. Tomographie par émission de positrons et Imagerie par résonance magnétique

La *tomographie par émission de positrons* (TEP) est une technique datant des années 60 [Rankowitz '62]. Elle consiste à injecter dans l'organisme un traceur radioactif sélectionné pour étudier une fonction précise de l'organisme. Il est injecté à une dose suffisamment faible pour ne pas constituer de danger. En se désintégrant, les noyaux radioactifs émettent des positrons, c'est-à-dire des particules de même masse qu'un électron, mais chargées positivement. Un dispositif spécifique appelé caméra à positrons permet de les capter, et d'obtenir des données brutes. Après traitement mathématique, il est possible de déterminer les variations de radioactivité dans l'organisme, et ainsi de connaître les zones possédant une forte concentration en traceur.

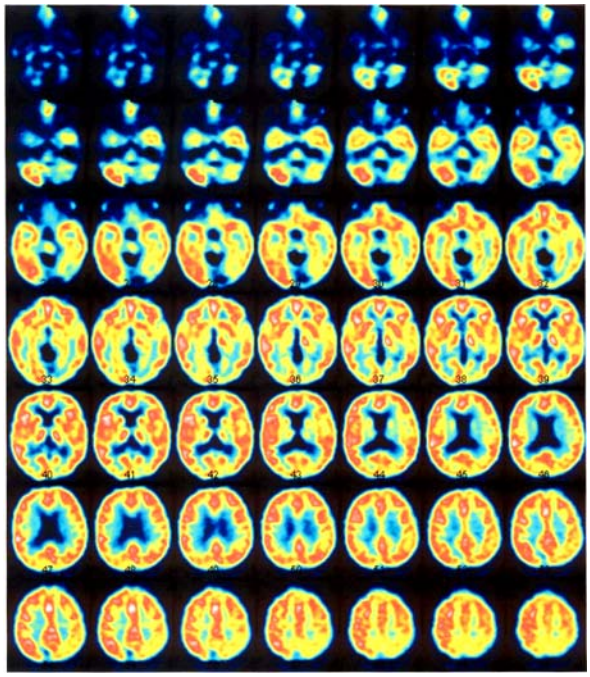

Figure I.2.2 : coupes réalisées par TEP (© Inserm u455 2003).

Dans le cadre de la neuroimagerie fonctionnelle, on utilise généralement de l'eau marquée à l'oxygène 15 (H_2O^{15}). Ainsi, il est possible de déterminer des débits sanguins régionaux, et donc d'étudier indirectement la dynamique sanguine. La résolution temporelle, qui dépend de l'isotope utilisé, est très mauvaise, de l'ordre de 40 secondes. La résolution spatiale est de 4-5 millimètres [Defrise & Trebossen '02].

Les premières acquisitions utilisant l'*imagerie par résonance magnétique* (IRM) eu lieu en 1973 [Lauterbur '73]. L'IRM est basée sur le principe de la résonance magnétique nucléaire (RMN), qui veut qu'il soit possible, en présence d'un champ magnétique, de faire entrer en résonance des atomes au moyen d'une onde radio. Chaque atome correspond à une fréquence d'onde spécifique. Lorsqu'un atome entre en résonance, il absorbe l'énergie de l'onde, et il la restitue à l'arrêt du phénomène de résonance. Ceci permet de calculer l'énergie absorbée, et de là, la concentration pour ce type d'atomes.

Quand cette technique est utilisée pour obtenir des images des tissus, on parle d'IRM anatomique (IRMa). L'IRM fonctionnelle (IRMf) repose sur la technique BOLD (*B*lood *O*xygenation *L*evel *D*ependant). Elle consiste à étudier les modifications de signal liées aux variations d'oxygénation sanguine [Cottier & Destrieux '01] en appliquant le principe de la RMN aux propriétés magnétiques de l'hémoglobine.

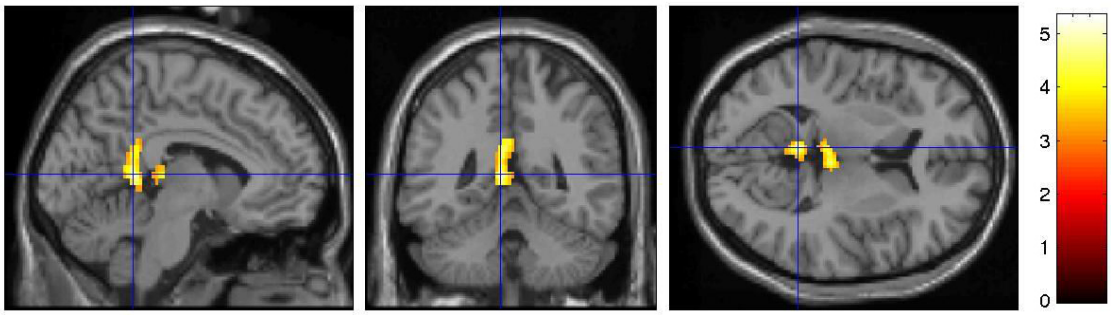

Figure I.2.3 : exemples d'images de l'activation cérébrale par IRMf [Ruff *et al.* '03].
(a) : coupe sagittale. (b) : coupe coronaire. (c) : coupe horizontale.

La résolution spatiale est comparable à celle de la TEP, mais la résolution temporelle est bien supérieure, de l'ordre de la centaine de millisecondes. L'IRMf est plus invasive que l'EEG ou la MEG, puisqu'il est nécessaire de soumettre le sujet à un rayonnement magnétique. En revanche, elle est moins invasive que la TEP, qui nécessite une injection de traceur radioactif. D'un autre côté,

le dispositif utilisé pour produire le champ magnétique émet un bruit très important, qui peut gêner le sujet lors de l'acquisition de son signal cérébral.

2.2.2. Traitement des données

Les techniques tomographiques permettent de réaliser des images du cerveau prenant la forme de coupes [Defrise & Trebossen '02]. A la différence des techniques de surface, il est donc possible d'étudier l'activation des structures sous-corticales. La définition temporelle dépend du nombre de coupes réalisées : plus il est important, moins la définition temporelle est bonne.

Il existe une autre différence essentielle entre ces deux familles de neuroimagerie fonctionnelle : la mesure des variations du débit sanguin est un indicateur de l'activité cérébrale bien moins précise et sûre que la mesure du champ électrique. On ne connaît ni la distance séparant le neurone du vaisseau sanguin, ni le rapport entre l'activité du neurone et l'augmentation du débit, ni même le délai nécessaire à l'organisme pour répondre aux besoins des neurones [Mazoyer & Belliveau '96; Sergent '94]. En fait, si les techniques de surfaces mesurent l'activité *neuronale*, on suppose par contre que les techniques tomographiques mesurent plutôt l'activité *synaptique* [Horwitz *et al.* '99]. Or, l'activité synaptique peut correspondre aussi bien à une activation qu'à une inhibition. Enfin, le débit sanguin cérébral subit l'influence, mal connue, d'autres paramètres physiologiques indépendants de l'activité cérébrale [Horwitz *et al.* '00].

Comme pour les techniques de surface, on rencontre ici aussi le problème de l'activité spontanée du cerveau, qui se répercute sur le débit sanguin cérébral. Toutefois, ici, les difficultés techniques (reproductibilité, confort du sujet) et le coût des acquisitions rendent plus difficile l'obtention de nombreux résultats nécessaires au moyennage.

La localisation des zones activées suit un processus différent de celui employé en MEG/EEG. L'image anatomique de chaque sujet, obtenue par exemple par IRMa, est normalisée de façon à correspondre à un atlas anatomique standard du cerveau. Il sera ainsi possible de comparer des images de sujets différents. Puis, les mêmes modifications sont appliquées aux données fonctionnelles, et les images fonctionnelles sont superposées aux images anatomiques. Pour cette raison, les techniques tomographiques demandent l'immobilité la plus parfaite du sujet.

2.3. Autres Techniques

Il ressort clairement que les techniques présentées précédemment possèdent des propriétés qui, principalement en termes de définitions spatiale et temporelles, les rendent complémentaires. Il existe des équipements permettant l'utilisation conjointe de deux techniques, comme par exemple l'EEG et la MEG. Mais la plupart du temps, les contraintes techniques rendent impossible ce type d'étude. Pour contourner cette difficulté, les différentes modalités d'acquisition sont utilisées successivement, ce qui demande une bonne reproductibilité de l'expérience. En général, on combine une technique tomographique et une technique de surface, comme par exemple : EEG et IRMf [Alary *et al.* '98], MEG et TEP [Joliot *et al.* '98], ou EEG et TEP [Thierry *et al.* '98]. La bonne résolution spatiale de la technique tomographique est utilisée pour faciliter le placement des dipôles lors du traitement des données issues de la technique de surface. Des difficultés sont également rencontrées lors de la fusion de données obtenues par différentes techniques, essentiellement en raison des différences de résolution (spatiale et temporelle) et de la nature de l'activité mesurée (neuronale ou synaptique) [Horwitz *et al.* '99].

Il existe également des techniques nouvelles, moins appliquées que les techniques de surfaces et tomographiques présentées. La *topographie optique* est une technique récente [Koizumi *et al.* '99] considérée comme une alternative à l'IRMf et à la TEP [Franceschini *et al.* '00]. Elle est non-invasive, et repose sur le principe de la spectroscopie en proche infrarouge qui est ici utilisé pour étudier la dynamique sanguine à travers les variations d'oxygénation (à l'instar de l'IRMf). La spectroscopie en proche infrarouge s'appuie sur la propriété qu'ont les tissus vivants d'absorber différemment ce type de rayonnements suivant la concentration en hémoglobine et le niveau d'oxygénation [Kennan *et al.* '02]. Des dispositifs appelés optodes, et répartis sur le scalp à la manière des électrodes d'un bonnet EEG, sont utilisés pour obtenir une carte spatiale de la lumière réfractée par le cortex cérébral. Les optodes comprennent des fibres optiques destinées à l'émission et des fibres optiques destinées à la mesure de la réfraction. Avec cette technique, le sujet est relativement libre de ses mouvements. La définition temporelle (pour des utilisations sur l'humain) peut aller jusqu'à 3 ms. Comme l'EEG et la MEG, cette technique d'imagerie est limitée à l'observation de l'activité corticale, mais les fibres optiques rendent possible l'utilisation conjointe de techniques basées sur les champs électromagnétiques.

La *stimulation magnétique transcranienne* (SMT) n'est pas à proprement parler une technique de neuroimagerie, mais il s'agit néanmoins d'un outil d'investigation en ce qui concerne les réseaux d'aires corticales. La SMT consiste à induire, via une électrode en contact avec le scalp, un champ

magnétique qui va inhiber l'activité neuronale, un peu comme si le sujet était un patient cérébrolésé, ou au contraire la déclencher, suivant l'amplitude et l'intensité du champ utilisé [Ilmoniemi '02]. Elle fonctionne donc à l'envers des techniques d'imagerie classiques, puisqu'au lieu d'observer l'activation cérébrale, elle permet de l'influencer. A cause de cette particularité, elle est utilisé la plupart du temps dans un contexte moteur, puisqu'il est plus aisé pour un observateur extérieur de relever une réponse motrice qu'un effet sensoriel. L'utilisation conjointe de la TMS et d'une technique d'imagerie classique offre des perspectives intéressantes, mais pose des problèmes techniques puisque le champ magnétique produit vient parasiter le champ mesuré (pour l'EEG, la MEG et l'IRMf).

L'*IRM de diffusion* (IRMd) applique la RMN à la mesure de la diffusion des molécules d'eau. Jusqu'à présent, elle était surtout utilisée dans des buts clinique (détection d'infarctus) et anatomique, notamment pour déterminer les connexions anatomiques entre aires cérébrales et orienter ces liaisons. Mais des résultats très récents ont montré que l'activation d'un neurone fait varier son volume, ce qui a une incidence directe sur la diffusion de l'eau (baisse de la diffusion) et peut donc être mesuré par IRMd [Le Bihan '00]. L'avantage sur les techniques tomographiques est que le rapport entre l'activation neuronale et les variations de diffusion est plus direct que ce n'est le cas pour les variations hémodynamiques.

2.4. Etudes en activation

Le principe des études en activation est d'étudier les liens entre une fonction cognitive et un RCGE en mesurant, grâce à une technique de neuroimagerie, l'activité cérébrale pendant la réalisation d'une tâche supposée mettre en œuvre la fonction cognitive en question. Les hypothèses sous-jacentes sont (1) que les valeurs mesurées expriment les comportements émergeant de populations neuronales (interactions synaptiques, transmission de potentiels d'action), (2) que ces comportements sont à la base des fonctions cognitives de haut niveau [Ingvar & Petersson '00] et (3) que la tâche met bien en œuvre la fonction étudiée.

Une étude en activation consiste à faire effectuer une tâche cognitive ou sensorimotrice précise à un sujet, tout en enregistrant des mesures indirectes de l'activité cérébrale. La tâche peut être complexe, par exemple : analyser une image ou bien très simple, par exemple : lire un mot, répéter un mouvement. Elle peut même être passive, c'est-à-dire ne pas appeler de réponse volontaire de la part du sujet : entendre un son, visionner une image. Mais dans tous les cas, elle est très contrainte

et contrôlée, dans le but de cibler une seule fonction cognitive ou sensorimotrice, et donc d'éviter de solliciter d'autres fonctions.

L'interprétation de données de neuroimagerie relatives à l'accomplissement d'une tâche cognitive pose donc de multiples problèmes. Le *premier* est le lien entre le signal mesuré par les techniques de neuroimagerie et l'activité neuronale. Il demande la modélisation de l'interface entre activité neuronale et signal de neuroimagerie [Arbib *et al.* '95; Taylor *et al.* '00]. Le *second* problème est le lien entre l'activité neuronale des réseaux à grande échelle d'aires cérébrales décrits au paragraphe 1.1.3, qui sont supposés être le support neurologique des fonctions cognitives complexes [Alexander *et al.* '92; Bressler '95; Goldman-Rakic '88], et les fonctions cognitives elles-mêmes. Le *troisième* concerne le lien entre les fonctions cognitives et les tâches qui sont censées les mettre en œuvre, c'est-à-dire entre la cognition et le comportement. Nous pensons que le traitement de ces deux derniers points passe par la modélisation de l'activité des RCGE en termes de traitement de l'information. Notre but est de développer un outil permettant de réaliser de tels modèles.

3. CONTRAINTES

D'après les descriptions qui viennent d'être faites, on peut déduire un certain nombre de contraintes qui devront être respectées par l'outil que nous voulons développer [Labatut '00]. Par souci de plausibilité biologique, certaines contraintes sont dictées par les propriétés du système à modéliser, le cerveau humain. De plus, les contraintes peuvent aussi avoir une origine technique, c'est-à-dire découlant des objectifs que nous avons fixés ou des techniques de neuroimagerie que nous allons utiliser pour obtenir les données qui nous permettront de construire ou valider les modèles.

3.1. Architecture en réseau

Du niveau cellulaire au niveau le plus intégré, le cerveau peut être considéré comme un réseau dont les nœuds (neurones, modules, aires) sont reliés par des arcs orientés (axones, paquets d'axones, faisceaux). La neuroimagerie fonctionnelle fournit des informations portant sur l'activité de populations neuronales. A ce niveau d'intégration, le cerveau peut être considéré comme un réseau orienté, dans lequel les nœuds sont des régions cérébrales fonctionnellement homogènes et

anatomiquement bien définies et les liens des faisceaux d'axones orientés. Nous avons vu que la fonction implémentée par un réseau dépendait en partie de sa structure [Friston '94; Lafon *et al.* '99; Svensén *et al.* '02]. La topologie du réseau est donc un élément déterminant dans la définition d'un modèle, qui doit être pris en compte lors du développement de l'outil de modélisation.

De plus, nous savons également (paragraphe 1.2) que la fonction implémentée par un réseau dépend des aires qui le composent, et plus précisément du rôle de ces aires dans le réseau. En effet, chaque région cérébrale possède ses propres caractéristiques fonctionnelles. Parfois, il est possible de décomposer ces caractéristiques sous la forme d'un réseau de primitives fonctionnelles (paragraphe 1.2.2). Des études sur la plasticité cérébrale (paragraphe 1.3) ont d'ailleurs montré que certaines aires présentaient des similitudes fonctionnelles, qui peuvent s'expliquer par des similarités dans l'organisation de ce réseau de primitives fonctionnelles.

Finalement, un modèle à grande échelle prend la forme d'un réseau orienté relativement hétérogène, puisque certains nœuds représentent directement des populations neuronales, alors que d'autres sont des primitives fonctionnelles issues de la décomposition d'une aire.

3.2. Temporalité

Le cerveau est un système dynamique, ce qui signifie que les mécanismes que nous étudions se déroulent dans le temps. L'organisation physique du cerveau, c'est à dire la façon dont chaque zone est reliée (ou pas) aux autres zones, et le délai que met un stimulus à parcourir une liaison impliquent forcément un ordre d'activation et une dynamique temporelle.

Le cerveau est un système non-stationnaire : les relations peuvent évoluer relativement rapidement (mécanismes d'apprentissage, plasticité et réorganisation cérébrale), et l'activité d'une région cérébrale est, elle aussi, variable dans le temps. L'échelle temporelle de cette dynamique cérébrale n'est pas la même à tous les niveaux. Par exemple, elle peut être de l'ordre de la milliseconde pour les traitements neuronaux, près d'une demi-seconde si on se place à un niveau plus cognitif, et plusieurs secondes en ce qui concerne le débit sanguin cérébral. Or, les nœuds qui composent un modèle sont hétérogènes et peuvent nécessiter la représentation de ces différentes échelles au sein d'un même modèle. Il est donc nécessaire d'utiliser un outil dans lequel le temps est représenté de façon explicite.

De plus, cette représentation du temps doit aussi être adaptée aux données expérimentales dont nous disposons pour la validation des modèles, et qui proviennent essentiellement d'études en activation. En ce qui concerne l'EEG, la MEG et l'IRMf, ces données sont constituées de séries temporelles. Il s'agit donc de mesures concernant les zones étudiées, répétées à intervalles réguliers sur la durée de la stimulation (les données de TEP comportent également des informations temporelles, mais leur définition temporelle est encore très peu précise, de l'ordre de quarante secondes).

3.3. Non-linéarité

Les relations entre les aires cérébrales ou les primitives fonctionnelles peuvent être de nature non-linéaire. On peut distinguer plusieurs causes à cela : d'une part, les relations entre les neurones sont généralement considérées comme non-linéaires [Guigon *et al.* '94]. Les neurones formels en sont l'illustration, leurs fonction de seuil étant la plupart du temps une sigmoïde (c.f. annexe A.4). Le comportement d'une population neuronale dépend du comportement des neurones qui la composent. Si les relations entres les neurones de deux aires cérébrales, pris individuellement, sont non-linéaires, nous avons tout lieu de penser que les relations liant les deux aires sont globalement non-linéaires. D'autre part, il existe au niveau des réseaux à grande échelle des processus de contrôle d'une aire sur une autre, qui introduisent des discontinuités dans la propagation de l'information, et donc de la non-linéarité.

3.4. Incertitude et imprécision

D'une part, à un haut niveau de modélisation (anatomiquement parlant), la transmission de l'information cérébrale semble soumise à des mécanismes non-déterministes. Par exemple, le même stimulus ne va pas provoquer systématiquement la même réponse de la part de l'aire qui le reçoit. En fait, au niveau cellulaire même, le comportement des neurones paraît en partie aléatoire. On suppose que dans certains cas, cet aspect aléatoire est essentiel à l'obtention d'une certaine stabilité [Woodburn *et al.* '00]. D'autre part, la variété (échanges électriques, chimiques), et la complexité des interactions (réseaux de neurones massivement interconnectés) entre neurones sont telles que le comportement émergeant au niveau des aires cérébrales paraît soumis à un mécanisme en partie aléatoire.

D'autre part, les observations obtenues au moyen des techniques de neuroimagerie sont des valeurs statistiques. On manipule en général des moyennes auxquelles on associe des valeurs de dispersion (e.g. variance), ou bien d'autres valeurs statistiques calculées à partir de ce couple, comme par exemple des z-scores (divergence d'un item par rapport à la moyenne de la population). Les valeurs obtenues sont dépendantes des inévitables erreurs expérimentales liées à l'imperfection des techniques d'acquisition. De plus, ce ne sont pas des mesures expérimentales directes de l'activité cérébrale, mais des variations de débits sanguins (techniques tomographiques) ou de champs électriques (techniques de surface) reflétant elles-mêmes l'activité cérébrale. Les valeurs obtenues présentent donc une certaine imprécision, spatiale ou temporelle, suivant la technique d'imagerie utilisée. Cette imprécision dans les données manipulées vient renforcer l'incertitude due aux relations non-déterministes [Dubois & Prade '94].

Enfin, la présence de relations non-linéaires dans le modèle a pour conséquence d'accroître fortement l'imprécision et l'incertitude de l'information au cours de sa propagation.

3.5. Causalité

La transmission de l'information cérébrale se réalise par le cheminement d'un stimulus à travers une population de neurones interconnectés. Les neurones relaient l'information, en d'autres termes : c'est l'activation d'un neurone par un influx électrique qui va *causer* l'activation du neurone suivant, et ce mécanisme apparaît aussi bien que l'on s'intéresse au fonctionnement cérébral au niveau du neurone, qu'au niveau des aires cérébrales. On peut donc considérer le cerveau comme un système causal.

De plus, notre objectif est de créer un outil utilisable en neurosciences pour tester diverses hypothèses sur le fonctionnement cérébral. Ces hypothèses doivent pouvoir être exprimées à la fois de façon très complète, et aussi très intuitive, puisque l'outil ne s'adresse pas à des experts en programmation. Or, dans le milieu scientifique, les hypothèses sont la plupart du temps exprimées en termes de causes et d'effets, d'où l'intérêt d'avoir une approche causale lors de la modélisation. D'un point de vue formel, la nécessité d'une modélisation causale sera plus détaillée dans le chapitre II.4.1.

3.6. Information cérébrale au niveau intégré

Au niveau intégré d'une population neuronale, l'information cérébrale peut être considérée comme l'abstraction de l'activité intégrée des cellules individuelles. A un tel niveau, le degré sémantique de l'information traitée est supérieur à celui de l'information manipulée au niveau du simple neurone. L'information manipulée par un neurone est représentée par son activité. L'information dans une aire cérébrale est caractérisée en partie seulement par l'activité de l'aire, qui correspond à l'intégration des activités des neurones pris individuellement. Elle est également caractérisée par la façon dont ces activations individuelles sont réparties dans la population qui la compose [Arbib '85b].

On retrouve cette idée de configuration dans l'organisation topique des cortex primaires. Cette propriété est également présente au niveau de la propagation de l'information entre aires cérébrales, à travers la configuration et le nombre de fibres activées dans un faisceau d'axones [Leiner & Leiner '97], qui correspondent à la configuration et au nombre de neurones activés dans le nœud cérébral émetteur. De ce fait, cette organisation topique des aires primaires se répercute dans certaines aires associatives et préfrontales [Goldman-Rakic '88] (paragraphe 1.2). Mais cette propriété n'est pas exclusivement corticale, elle est également présente à un niveau sous-cortical [Alexander *et al.* '92].

Si le nombre de neurones activés ou le niveau d'activation global de l'aire sont des données de nature quantitative, pouvant se rapprocher du signal émis par un neurone seul, en revanche la configuration des neurones activés est d'ordre plus qualitatif, et n'existe pas au niveau du neurone seul.

3.7. Plasticité

Notre but étant de modéliser des fonction cognitives de haut niveau, il est nécessaire d'inclure dans notre formalisme les mécanismes qui donnent au cerveau son adaptabilité, qui confèrent aux aires cérébrales des propriétés de réorganisation, d'habituation et d'apprentissage.

L'habituation est une diminution temporaire de l'activation qui survient quand une population neuronale reçoit plusieurs fois le même stimulus consécutivement, et qui disparaît quand quelque chose de différent lui est présenté. L'apprentissage est une modification à plus long terme, voire quasi-permanente de l'état du cerveau, qui apparaît quand une population neuronale reçoit régulièrement le même type d'information. La réponse de la population devient plus efficace, c'est-

à-dire que moins de neurones émettent, et que ceux-ci se spécialisent dans le traitement de cette information. Cela se traduit par une modification fonctionnelle de l'aire, qui ne traite plus l'information tout à fait comme avant.

Il est important de retenir que les éventuels changements fonctionnels d'une aire ne dépendent que de l'information reçue des aires en amont. L'habituation aussi bien que l'apprentissage ne font pas intervenir les aires en aval de l'aire concernée, à moins qu'il n'existe un lien en sens inverse, ou une boucle de retour. Mais, dans ce cas, l'information met un certain temps pour atteindre de nouveau l'aire concernée.

Chapitre II

MODELISATION CEREBRALE

Il existe à l'heure actuelle tout un pan de la recherche en neuroimagerie dédié à la modélisation. Le but est de répondre à certaines questions que nous avons soulevées dans le chapitre précédent, en particulier en ce qui concerne le lien entre les données d'activation et l'activité neuronale. Le modèle est alors un outil dont le rôle est d'expliquer ces résultats obtenus in vivo, via des techniques de neuroimagerie.

Il existe également deux autres grandes approches en modélisation cérébrale, issues de champs bien distincts : l'approche cognitive et celle des neurosciences computationnelles. La première est issue de l'intelligence artificielle (IA). La modélisation cérébrale y est vue comme un moyen d'imiter, afin de l'expliquer, le comportement du cerveau. Elle est abordée de façon plus ou moins plausible, biologiquement (et, dans une moindre mesure, fonctionnellement) parlant. Les neurosciences computationnelles tirent leur origine de la cybernétique et des mathématiques. Le but général est de modéliser et d'expliquer les liens entre la fonction donnée et l'activité du réseau qui l'implémente. Ici aussi, la plausibilité biologique du modèle peut varier. Les neurosciences computationnelles reposent sur des modèles formels que nous détaillerons davantage au chapitre III. Nous aborderons ici ces modèles seulement sous l'angle de leurs applications. Ce n'est que très récemment que des travaux situés à la jonction de ces approches ont commencé à apparaître [Arbib *et al.* '95; Pastor *et al.* '00; Taylor *et al.* '00], mettant en rapport dans un même modèle la structure biologique sous-jacente, les fonctions cognitives et des valeurs comparables à des mesures d'activation.

Dans ce chapitre nous allons décrire des travaux issus de ces différents domaines de la modélisation cérébrale. Nous analyserons les propriétés de ces diverses approches, relativement à nos objectifs et à nos contraintes, afin de montrer l'originalité de notre démarche. Puis, nous préciserons nos contraintes en les exprimant de façon formelle, dans l'optique de choisir le formalisme le plus adapté, afin de l'utiliser comme base de notre propre formalisme.

1. NEUROIMAGERIE

L'interprétation du signal de neuroimagerie au moyen d'un modèle se déroule en trois temps : traitement des données brutes (réalignement temporel, normalisation anatomique, filtrage…), expression d'un modèle, et ajustement statistique des paramètres du modèle aux données traitées. La modélisation a évolué avec les techniques de neuroimagerie, passant d'une simple description des aires activées à une représentation prenant en compte des informations plus complexes comme les coactivations ou les liens anatomiques.

1.1. Localisation

Les méthodes de modélisation les plus basiques s'appuient sur une analyse statistique des données de neuroimagerie pour localiser les aires activées lors de la réalisation d'une tâche cognitive étudiée. Cette *localisation* est soit *spatiale*, pour les techniques tomographiques, soit *temporelle*, pour les techniques de surface, à travers la chronologie et la datation d'événements cérébraux. Elle est soumise à une certaine imprécision, en raison de la nature des techniques utilisées, qui ne reflètent qu'indirectement l'activité cérébrale (c.f. chapitre I.2).

En ce qui concerne la TEP et l'IRMf, la méthode la plus répandue pour la localisation est la *méthode ségrégative* [Fox & Raichle '84, '85]. Cette méthode repose sur l'hypothèse (décrite dans le chapitre I.1.2) qui veut qu'une aire cérébrale soit spécialisée, c'est-à-dire qu'elle implémente une fonction donnée en termes de traitement de l'information cérébrale [Ingvar & Petersson '00]. De plus, l'activité cérébrale est considérée comme sommative, c'est-à-dire que l'activité observée est supposée être la somme des activités correspondant à chaque fonction en cours d'utilisation.

Il faut noter que même au repos, le cerveau a une activité, due aux divers processus inconscients et incontrôlables. Par conséquent, pour une tâche ciblant une fonction donnée, l'activation mesurée est en fait la résultante de l'activation cérébrale au repos et de celle de la fonction proprement dite. C'est pourquoi des mesures de l'activation correspondant à une *tâche de référence*, la plus proche possible du repos (en général : ne rien faire, ne penser à rien, fermer les yeux, etc.) sont réalisées, afin d'être soustraites (puisque l'on se place dans une hypothèse d'activation sommative) à l'activité mesurée pour la tâche.

Parfois, il est impossible de définir une tâche ne nécessitant que la fonction ciblée : la tâche implique l'utilisation d'autres fonctions cognitives, et donc l'activation d'aires qui vont parasiter l'identification de l'aire implémentant la fonction ciblée. Dans ce cas, la tâche de référence est choisie de façon à faire intervenir les fonctions superflues, mais pas la fonction ciblée. On retrouve ainsi, par soustraction, l'activation de la fonction ciblée.

La technique habituellement utilisée pour analyser les données est celle des modèles linéaires généraux [Friston *et al.* '95], mais un grand nombre d'autres techniques, en général linéaires, sont également employées [Horwitz *et al.* '99], par exemple des filtres de Kalman [Gössl *et al.* '00]. Parfois, un traitement supplémentaire permet de déterminer la relation entre les variations hémodynamiques mesurées par les techniques tomographiques et l'activité neuronale [Gössl *et al.* '01].

Un *modèle linéaire général* (MLG) est un outil statistique de régression multiple. Il permet d'étudier une relation linéaire potentielle entre deux groupes de variables pour lesquelles on dispose de séries observations. Ici, le MLG est utilisé pour déterminer les zones dont les activations sont statistiquement différentes suivant qu'on les observe pour la tâche de référence ou pour la tâche étudiée. Pour cela, les *voxels* (équivalent du pixel pour une image en 3D) composant les images sont comparés un à un. Lorsque l'activation d'une certaine aire pour la tâche étudiée est supérieure à celle observée pour la même aire lors de la tâche de référence, on parle d'*hyperactivation* de l'aire. Dans le cas contraire, on parle d'*hypoactivation*.

Le choix de la tâche de référence est très important, puisqu'il peut grandement faire varier la localisation des aires d'intérêt. De plus, le fait de considérer qu'une fonction cognitive peut être décomposée en une combinaison linéaire de fonctions plus simples constitue une approximation très grossière. En général, une fonction complexe est, au contraire, le résultat d'interactions non-linéaires [Ingvar & Petersson '00].

En ce qui concerne l'EEG et la MEG, la localisation temporelle se fait de façon beaucoup plus directe. Deux indices permettent d'effectuer la datation des évènements cérébraux à partir des ondes électromagnétiques mesurées : les variations d'amplitude et les différences de latence. En effet, un pic de potentiel (positif ou négatif) permet de déterminer le moment où une aire cérébrale est activée [Horwitz *et al.* '99]. Une latence observée entre deux ondes de même forme mesurées sur deux tâches cognitives différentes peut également être un trait caractéristique de l'activation d'une

aire cérébrale. La localisation spatiale se fait après une étape de modélisation qui vise à retrouver les sources des variations électromagnétiques mesurées (c.f. chapitre I.2.1).

1.2. Coactivation

Le fait de localiser les aires impliquées dans la mise en œuvre d'une fonction donnée n'est pas satisfaisant, car il ne permet pas d'expliquer les résultats obtenus. L'étape suivante dans l'évolution des modèles en neuroimagerie est la méthode de la *connectivité fonctionnelle* [Herbster *et al.* '96], qui consiste à localiser le réseau des aires impliquées dans le traitement de la tâche cognitive. Pour cela, on va identifier les aires activées, mais également leurs relations fonctionnelles.

On parle de relation fonctionnelle entre deux aires quand celles-ci s'activent conjointement, en même temps ou de façon décalée dans le temps. Formellement, la coactivation se traduit par une corrélation (ou une covariance) de l'activation des aires cérébrales [Friston '94]. La force de la relation est déterminée par cette valeur de corrélation.

Pour les techniques tomographiques, l'étude est réalisée au moyen d'outils d'analyse statistique factorielle, du type *analyse en composantes principales* (ACP) ou *analyse en composantes indépendantes* (ACI). L'ACP permet de traiter un ensemble complexe de variables corrélées pour construire un nouvel ensemble, plus restreint, de variables décorrélées (les composantes principales), tout en évitant une trop grande perte d'information. L'hypothèse principale est que les variables initiales correspondent à une combinaison linéaire des composantes principales. On suppose également que les composantes principales suivent une distribution normale. L'ACI est une généralisation dans laquelle cette dernière hypothèse est relâchée.

L'ACP est utilisée sur les données de neuroimagerie d'un sujet pour extraire des composantes principales correspondant à des configurations spatiales (spatio-temporelles pour de l'IRM [Kherif *et al.* '02]), et des scores (ou valeurs propres) qui leurs sont associés. Ces configurations correspondent aux aires cérébrales activées par la tâche cognitive, et les scores quantifient l'intensité de leurs relations fonctionnelles. Une fois que les aires et leurs relations ont été obtenues, on obtient un RCGE, et il est possible de tenter d'interpréter ce résultat en termes de fonctions cognitives [Horwitz *et al.* '99]. Dans une étude de Svensén et coauteurs ['02], l'ACI est utilisée sur les mesures d'IRMf d'un groupe de sujets, afin de faciliter l'interprétation des composantes principales

obtenues (essentiellement en atténuant certains bruits individuels et en renforçant les composantes correspondant à des configurations présentes chez plusieurs sujets).

On trouve également des méthodes d'analyse qui ne sont pas entièrement guidées par les données, et qui nécessitent de spécifier un modèle statistique à priori, comme les *modèles linéaires multivariés* (MLM) [Kherif *et al.* '02].

En qui concerne les techniques de surface, la connectivité fonctionnelle peut être étudiée à travers la corrélation existant entre les différentes électrodes. Il existe pour cela plusieurs méthodes, qui diffèrent suivant les caractéristiques du signal électromagnétique retenues pour déterminer la corrélation. Par exemple, la mesure appelée *cohérence* est la corrélation entre des signaux d'EEG dans le domaine fréquentiel [Horwitz *et al.* '00].

Par rapport aux techniques tomographiques, les techniques de surface ont l'avantage de décrire la dynamique du traitement cérébral. En réalisant une analyse statistique des corrélations existant entre les mesures provenant de différents canaux (i.e. différentes électrodes, pour l'EEG), il est possible de déterminer la force des relations entre les canaux, un peu comme en tomographie. Mais, en utilisant en plus la dimension temporelle des mesures, il est également possible d'orienter ces relations. Cette analyse s'appuie sur des mesures de causalité interprétables au sens de Granger [Kaminski *et al.* '01; Liang *et al.* '00]. La définition de la causalité de Granger [Granger '69] veut que, pour deux séries temporelles de mesures, la première est la cause de la seconde si l'erreur sur la prédiction de la valeur de la seconde à l'instant présent est réduite, en tenant compte des valeurs passées de la première lors de la régression. A noter que la régression est de type linéaire. Ce type d'analyse est également utilisée lors d'expériences impliquant des mesures de l'activité neuronale par électrodes implantées [Liang *et al.* '00].

Les relations fonctionnelles n'ont pour fondement que le seul traitement statistique effectué sur les données d'imagerie. La présence d'une relation fonctionnelle laisse supposer l'existence d'un lien anatomique direct, mais, même si c'est souvent le cas [Sporns *et al.* '02], elle ne la prouve pas [Ingvar & Petersson '00; Liang *et al.* '00]. En effet, la covariance observée entre les activations de deux régions peut être due à l'influence d'une troisième région sur les deux premières, ou bien à une quelconque combinaisons d'effets directs et indirects. Ceci pose un important problème en ce qui concerne l'interprétation du réseau fonctionnel obtenu.

La même critique peut s'appliquer à l'orientation d'une relation suivant la causalité de Granger, puisque celle-ci se base exclusivement sur des critères statistiques, et peut donc aller à l'encontre de la réalité anatomique. En fait, de façon plus générale, le calcul de la mesure de causalité suppose ici que les deux éléments sont liés par une relation causale [Kaminski *et al.* '01], ce qui n'est pas forcément le cas.

1.3. Liens anatomiques

Pour éviter autant que possible d'obtenir des relations fonctionnelles ne correspondant pas à des relations anatomiques existantes, il est nécessaire d'avoir la possibilité de tenir compte de l'information anatomique disponible lors de l'analyse des données. Pour cela, une autre méthode d'interprétation appelée *connectivité effective* a été développée [Büchel & Friston '97; Friston '94].

La connectivité effective repose sur les *modèles d'équations structurelles* (MES). Un tel modèle est un outil statistique reposant sur l'analyse des variances et des covariances d'un ensemble de variables. Il permet de savoir si ces variables peuvent être considérées comme les observations des entrées et des sorties d'un modèle linéaire. L'intérêt est de pouvoir spécifier la nature des relations entre entrées et sorties, sous la forme d'un système d'équations.

En neuroimagerie, les modèles portent sur les activations des régions d'intérêt relatives à la tâche cognitive. Les relations sont définies sous forme fonctionnelle, entre des régions que l'on sait être connectées anatomiquement, suivant les données de neuroanatomie disponibles, ou que l'on suppose connectées, suivant les hypothèses faites par le modélisateur sur l'existence de liens anatomiques. Ceci permet d'exprimer de façon explicite des contraintes sur les relations, par exemple en fixant certains paramètres. Des études portant sur la nature des connexions (densité, complexité) entre les aires cérébrales peuvent également venir compléter le modèle [Sporns *et al.* '00a, b, '02].

Après l'analyse statistique des données, on obtient des valeurs qui quantifient l'importance de chaque lien [Ingvar & Petersson '00]. On compare les valeurs de ces poids pour différentes tâches cognitives, ou bien pour différents groupes de sujets effectuant la même tâche cognitive. Certaines études ont ainsi montré que l'apprentissage [Büchel *et al.* '99; Büchel & Friston '00] ou des variations attentionnelles [Büchel & Friston '97] provoquaient des changement dans la connectivité effective d'aires impliquées dans la réalisation de ces fonctions. Dans [Krause *et al.* '00], les MES

sont utilisés sur des données de TEP et d'IRMf, en conjonction avec d'autres techniques d'interprétation des données, pour étudier l'évolution, due au vieillissement, des mécanismes de la mémoire de travail.

En résumé, la connectivité effective revient à utiliser des données ou des hypothèses anatomiques pour contraindre l'analyse des corrélations. Comme dans tout modèle, il est toujours possible d'avoir deux aires corrélées sans que, pour autant, l'éventuel lien anatomique qui les unit soit utilisé : la relation peut être indirecte, et passer par des aires absentes du modèles [Ingvar & Petersson '00]. Néanmoins, par rapport à la connectivité fonctionnelle, la connectivité effective permet de renforcer la plausibilité anatomique des liens déterminés de façon statistique.

Une autre particularité provoquant des problèmes d'interprétation est l'utilisation purement statistique et non-causale des MES : une connexion effective peut être renversée mathématiquement [Büchel *et al.* '99]. Or, si les relations fonctionnelles étudiées ne sont pas orientées, les liens anatomiques, eux, le sont. De plus, les MES permettent uniquement de modéliser des relations linéaires. Il est possible d'introduire certaines formes de non-linéarité [Büchel & Friston '00; Friston *et al.* '98], mais l'expression de relations de ce type reste limitée à la modulation de l'activité d'une aire par une autre, représentée par un paramètre multiplicatif [Büchel *et al.* '99].

1.4. Bilan

Les techniques de modélisation dédiées à l'interprétation de données de neuroimagerie ont été définies dans le but d'étudier les conséquences, en termes de mesures d'activation, de la mise en œuvre de fonctions cognitives. L'objectif premier est d'identifier les structures impliquées dans la réalisation de la fonction, et plus tard, les liens entre ces structures.

La méthode de la localisation utilisée sur les données issues des techniques tomographiques permettent de savoir *où* sont les foyers d'activation, et donc de savoir quelles sont les aires activées. L'application de la même méthode aux techniques de surface permet d'obtenir des indices quant à la localisation temporelle de certains évènements cérébraux, et donc de savoir *quand* la fonction est réalisée. L'étude des coactivations (connectivité fonctionnelle) permet d'avoir une idée de la nature des relations fonctionnelles entre les différentes aires identifiées. Elle permet donc de savoir *quel* est le réseau fonctionnel, et en particulier, d'obtenir des renseignements sur sa topologie. En prenant

en compte les liens anatomiques (connectivité effective), on tente de savoir *pourquoi* l'activation d'une aire affecte une autre aire.

Les méthodes utilisées en neuroimagerie permettent de répondre à des questions de certains types (où, quand, quoi, et pourquoi). Ces questions portent exclusivement sur le rapport existant entre l'activation (mesurée par la neuroimagerie) d'un certain réseau d'aires cérébrales correspondant à l'implémentation d'une fonction cognitive donnée, et les propriétés structurelles de ce réseau (étudiées par les méthodes d'interprétation) : aires impliquées, topologie, etc. Or, la bonne connaissance du substrat cérébral d'une fonction cognitive est nécessaire, mais pas suffisante pour diagnostiquer précisément un déficit fonctionnel, prévoir les conséquences d'une lésion, ou simplement étudier des mesures d'activation contradictoires. Il est également essentiel d'interpréter les données de neuroimagerie comme le résultat du traitement de l'information cérébrale à un niveau intégré. En d'autres termes, il faut répondre à la question de savoir *comment* la mise en œuvre d'un réseau provoque l'activation observée.

Les méthodes interprétatives classiques utilisées en neuroimagerie partagent plusieurs points communs avec notre approche, dans le sens où elles respectent certaines de nos contraintes. En particulier, les approches les plus récentes ont introduit la notion de causalité dans les modèles. Elles emploient différentes définitions de la causalité, qui ont en commun de reposer sur des principes statistiques. Toutefois, la différence la plus flagrante est que les modèles utilisés en neuroimagerie ignorent complètement l'aspect du traitement de l'information cérébrale : elles ne tiennent pas compte du lien existant entre l'activation mesurée et les mécanismes de traitement de l'information cérébrale sous-jacents à la fonction cognitive. De ce fait, il n'y a bien sûr pas de volonté de représenter dans ces modèles les caractéristiques de l'information cérébrale au niveau intégré (représentation duale, imprécision, etc.)

2. APPROCHE COGNITIVE

Les premiers modèles cognitifs, directement issus de *l'intelligence artificielle* (IA) sont des modèles symboliques purement fonctionnels [Minsky '75; Quillian '67]. Ils ont été à la base de nombreuses interprétations en neuropsychologie. Il existe également une approche beaucoup plus récente fondée sur une représentation probabiliste, dans laquelle des réseaux bayésiens sont utilisés pour modéliser principalement des mécanismes moteurs [Ghahramani & Wolpert '97].

2.1. Modèles symboliques

Les premiers modèles cognitifs du fonctionnement cérébral sont les modèles symboliques, datant des années 50. Ce sont des travaux inspirés de la machine de Von Neumann. Les modèles sont purement fonctionnels, c'est-à-dire que l'on considère les fonctions cognitives comme complètement déconnectées du substrat neuronal qui les implémente. Ce type de modélisation ne peut donc pas être validé par des observations anatomiques ou des données de neuroimagerie. Pour cela, on utilise des données comportementales : le modèle doit être consistant avec les caractéristiques de l'homme, aussi bien en ce qui concerne les qualités que les défauts [Laplane '94; Reeke & Sporns '93].

Globalement, ces modèles symboliques sont basés sur des systèmes experts, comme le formalisme *MOP* (*M*emory *O*rganizer *P*ackets) de Schank [Schank & Abelson '77; Schank & Farrell '88], qui définit un modèle de représentation de la connaissance , ou bien sur des systèmes à base de règles plus spécifiques, comme la théorie *ACT* (*A*daptive *C*ontrol of *T*hough) de Anderson [Anderson '88, '89; Reeke & Sporns '93], qui définit une architecture cognitive. Une architecture cognitive est une théorie décrivant la structure computationnelle immuable de la cognition [Lewis '01] (processus primitifs, structures de mémoire, de contrôle, etc.). Les modèles développés sur ces formalismes sont dédiés à une fonction cognitive de haut niveau en particulier [Anderson '88; Minsky '75; Quillian '67]. Par exemple, le modèle décrit dans [Schank & Abelson '77; Schank & Farrell '88] est dédié à la représentation sémantique de phrases dans la mémoire. La théorie ACT concerne également la mémoire, mais il s'agit cette fois plutôt de la modélisation des relations entre différents types de mémoire (déclarative, procédurale, de travail). Les modèles visent essentiellement l'apprentissage, la mémoire, et la compréhension du langage [Meyer & Kieras '97].

Les travaux plus récents s'orientent vers une réunion des modèles dédiés à des fonctions en particulier, afin de définir une théorie computationnelle unifiée de la cognition [Meyer & Kieras '97]. Le formalisme *Soar* [Laird *et al.* '86; Lewis '01] inclut des éléments supplémentaires, comme l'influence des mécanismes moteurs, perceptuels et attentionnels sur les processus cognitifs. Au niveau formel, les systèmes à base de règles sont toujours utilisés pour représenter l'information au plus bas niveau, mais d'autres mécanismes viennent les compléter à un plus haut niveau, comme par exemple un traitement parallèle des règles de production ou un algorithme de résolution de problème [Meyer & Kieras '97]. Dans le formalisme *EPIC* (*E*xecutive *P*rocess-*I*nteractive *C*ontrol)), les aspects moteurs et sensoriels sont encore plus renforcés, à travers des modules parallèles dédiés

à des traitements spécifiques : modules cognitif, moteur, perceptuel, pouvant être eux-mêmes divisés en sous-modules (auditif, visuel, etc.) [Kieras & Meyer '95; Kieras & Meyer '96].

Actuellement, différents travaux portent sur l'identification des modules composant ces derniers modèles cognitifs en termes de régions cérébrales [Anderson *et al.* '02a]. En d'autres termes, il s'agit d'introduire une certaine plausibilité biologique dans ces modèles. *ACT-R 5.0* [Anderson *et al.* '02a] est une extension de la théorie ACT, utilisée pour tester la validité de ces identifications. ACT-R 5.0 est proche de EPIC en ce qui concerne la décomposition modulaire utilisée. Elle permet de générer des valeurs reflétant l'activité de certains modules. Ces valeurs sont comparées à la réponse BOLD mesurée par IRMf [Anderson *et al.* '02b].

2.2. Réseaux bayésiens

Un réseau bayésien est un modèle graphique représentant des dépendances conditionnelles dans un ensemble de variables aléatoires. Il s'agit d'un graphe acyclique orienté, dans lequel les nœuds représentent les variables aléatoires, et les arcs symbolisent les relations existant entre ces variables. Une distribution de probabilités conditionnelles est associée à chaque relation [Pearl '88].

Les filtres de Kalman constituent une classe de réseaux bayésiens dédiés à la modélisation de systèmes numériques continus dynamiques à relations linéaires, appelés *modèles d'espace d'états* (c.f. chapitre III.1.3.2.c pour une description formelle). En l'occurrence, ils sont utilisés pour développer des modèles explicites des processus d'intégration sensorimoteurs [Ghahramani '97; Wolpert *et al.* '95], notamment la transformation d'une information d'origine visuelle (par exemple, un objet à saisir) en une information utilisable pour les centres qui commandent le mouvement (par exemple, la position que doit occuper le bras pour saisir l'objet), et l'apprentissage de cette action de transformation.

Dans [Ghahramani & Wolpert '97], c'est un modèle à mélange d'experts qui est utilisé pour le même type de modélisation. Un modèle à *mélange d'experts* [Jordan & Jacobs '94], ou modèle d'espace d'états *à bascule* (switching state-space model) [Ghahramani & Hinton '98], permet de modéliser certains types de relations non-linéaires, sous la forme d'une combinaison non-linéaire de résultats provenant de plusieurs modèles linéaires. Ce type de modèle repose sur l'hypothèse qu'il est possible d'effectuer une décomposition modulaire de la fonction modélisée. Par exemple, dans

le cas des mécanismes moteurs, un mouvement complexe est supposé décomposable en un ensemble de primitives motrices [Ghahramani '00b; Wolpert & Ghahramani '00].

Ces modèles reposent sur l'hypothèse que l'homme possède un modèle interne pouvant prédire les retours sensoriels correspondants à la réalisation d'un mouvement donné. Le fait d'apprendre un mouvement donné consiste donc à réduire l'écart entre la prédiction interne et ce qui est effectivement observé. Or, le principe du filtre de Kalman repose précisément sur une succession de prédictions/révisions. La valeur d'une variable décrivant l'état d'un modèle est prédite, puis une révision bayésienne lui est appliquée en fonction de données observées du système (c.f. paragraphe III.1.3.1). Le filtre de Kalman constitue donc un formalisme particulièrement bien adapté à ce type de modélisation.

Mais toutes les fonctions cérébrales ne peuvent pas être modélisées de cette façon. En particulier, celles qui sont d'un plus haut niveau cognitif ne sont pas nécessairement dirigées par un retour sensoriel. De plus, bien que des tentatives soient faites, *a posteriori*, pour rapprocher ces fonctions de certaines structures cérébrales [Ghahramani '97], cette approche est néanmoins largement cognitive : elle permet de donner des explications sur la façon dont certains mécanismes moteurs fonctionnent, mais aucun aspect biologique n'est explicitement intégré dans les modèles. Enfin, le formalisme utilisé contraint la nature (linéaire) du traitement implémenté par les modules. Or, rien ne permet d'affirmer que ce qui est observé dans le cas moteur est applicable pour d'autres fonctions cérébrales de haut niveau.

2.3. Bilan

L'approche cognitive se caractérise par des modèles purement fonctionnels, dans lesquels on considère les fonctions cognitives comme complètement déconnectées du substrat neuronal qui les implémente. Cette absence complète de plausibilité biologique pose un problème, car en l'absence de recoupement avec des données biologiques, rien ne prouve qu'un modèle reproduisant un comportement soit bel et bien un modèle fidèle. De plus, l'absence de plausibilité biologique est incompatible avec l'interprétation d'une fonction en termes d'activité cérébrale mesurable par neuroimagerie.

Toutefois, on peut remarquer que l'orientation actuelle des travaux, qui vise à identifier les éléments de modèles cognitifs en termes de structures cérébrales, montre bien l'intérêt de fonder dès le

départ notre formalisme sur des bases biologiquement plausibles. De plus, à l'instar de certains modèles employés en neuroimagerie (c.f. paragraphe 1.2), les formalismes cognitifs misent sur une approche causale de la modélisation. Ceci est explicite dans les applications à base de systèmes logiques. Dans le cas de l'utilisation des réseaux bayésiens, la causalité de l'approche n'est pas revendiquée, mais le formalisme utilisé (filtre de Kalman) est causal. Un réseau bayésien est dit causal quand ses relations ne sont pas orientées de façon arbitraire, mais bien en fonction du système qui est modélisé [Heckerman & Breese '95]. Pour Judea Pearl, la notion de causalité d'un réseau bayésien n'est pas qu'une question d'interprétation [Pearl '94]. D'après sa définition, une relation causale est avant tout un mécanisme stable et autonome, c'est-à-dire qui peut être modifié sans entraîner de changement dans l'ensemble des autres mécanismes formant le système considéré (le changement est local à la relation). Cela implique que le mécanisme sous-jacent à la relation causale soit explicitement décrit dans le modèle (i.e. sous forme fonctionnelle).

3. NEUROSCIENCES COMPUTATIONNELLES

Le but d'un modèle en neurosciences computationnelles est l'étude et la compréhension des mécanismes cérébraux qui permettent de réaliser une fonction cognitive ou sensorimotrice. On cherche donc ici à savoir *comment* une telle fonction est implémentée, en y incluant plus ou moins de plausibilité physiologique. La plupart des travaux actuel sont basés sur une approche *connexionniste*, c'est-à-dire utilisant des *réseaux de neurones formels* (RNF) [Arbib '95]. Ces modèles explorent différents niveaux de plausibilité biologique, et différentes échelles anatomiques.

Tout récemment, plusieurs travaux ont réalisé la jonction entre l'approche utilisée en neuroimagerie et celle des neurosciences computationnelles. En effet, d'un côté, les chercheurs en neurosciences computationnelles se sont longtemps focalisés sur la modélisation de petites structures cérébrales, et ne se sont intéressés que depuis peu à la modélisation à grande échelle. Or les données d'imagerie reflètent l'activité de ces réseaux-là. D'un autre côté, les chercheurs en neuroimagerie n'ont pas eu conscience de l'évolution des neurosciences computationnelles, qui leur ouvraient une nouvelle voie pour l'interprétation des données d'activation [Horwitz *et al.* '00].

L'objectif est d'utiliser des modèles biologiquement plausibles de l'implémentation de fonctions cognitives ou sensorimotrices, pour l'interprétation de données de neuroimagerie. La première approche, celle des *réseaux neuronaux à grande échelle*, repose sur les RNF, le formalisme le plus

répandu dans la modélisation en neurosciences. Elle consiste à convertir l'activité neuronale simulée par un RNF en des valeurs constituant un équivalent artificiel des mesures de variations du débit sanguin cérébral. L'objectif est de comparer ces valeurs simulées à des valeurs mesurées par techniques tomographiques. La seconde approche est originale et nouvelle en modélisation cérébrale, elle est basée sur un formalisme causal, les *réseaux causaux qualitatifs*.

3.1. Réseaux de neurones

Un neurone formel est la représentation, sous forme de fonctions mathématiques, de certaines caractéristiques du neurone biologique (c.f. l'annexe A.4, pour une description formelle), notamment en ce qui concerne le traitement de l'information et l'apprentissage. Les neurones formels sont des éléments de traitement simples, effectuant des opérations élémentaires. Un réseau de neurones formels (RNF) se compose de neurones formels fortement interconnectés. Les propriétés fonctionnelles du réseau dépendent principalement de son architecture et des fonctions qui définissent les neurones qui le composent. Un autre aspect essentiel est l'algorithme d'apprentissage utilisé pour paramétrer le réseau : il existe différents mécanismes plus ou moins biologiquement plausibles, et adaptés à certaines architectures. Un comportement complexe peut ainsi émerger d'un réseau constitué de nombreux neurones identiques.

Les modèles utilisant des RNF en neurosciences computationnelles sont utilisés avec des plausibilités biologiques très variables, pouvant aller d'une approche purement fonctionnelle [Cohen *et al.* '90; Jani & Levine '00; Levine *et al.* '93] à une approche purement physiologique [Tiesinga *et al.* '01; Wang & Buzsaki '96], en passant par différents niveaux intermédiaires [Levine *et al.* '93; Lumer *et al.* '97].

3.1.1. Approches purement fonctionnelles et purement physiologiques

Les RNF purement fonctionnels ont été développés dans le but de répondre aux critiques exercées à l'encontre des premiers modèles symboliques, concernant leur traitement sériel et leur représentation de l'information cérébrale trop rigide. Au contraire, dans un RNF, l'information est distribuée sur le réseau, et le traitement se fait de façon parallèle. C'est en raison de ces propriétés que l'approche connexionniste de la modélisation fonctionnelle, apparue dans les années 80, a été qualifiée de *PDP* (*P*arallel *D*istributed *P*rocessing) [McClelland & Rumelhart '81]. Les modèles ciblent une fonction en particulier, par exemple, la perception de lettres dans les mots [McClelland

& Rumelhart '81; McClelland *et al.* '86]. Si les premiers modèles symboliques n'avaient aucune plausibilité biologique, il faut souligner que ce n'est pas totalement le cas avec ces RNF. En effet, en général les modèles prennent en compte certains grand mécanismes biologiques (modèle neuronal de base, mécanismes d'inhibitions latérales, traitement parallèle, etc.). Cependant, leur plausibilité biologique est faible.

Il y a plusieurs causes à cela. Tout d'abord, l'organisation du réseau et les propriétés des neurones ne subissent aucune contrainte de plausibilité biologique *a priori*. Bien sur, il est possible d'interpréter le modèle *a posteriori*, et ainsi de rapprocher certains de ses éléments de structures biologiques réelles [Arbib '85b]. La méthode utilisée pour l'apprentissage est également un facteur important, car certains algorithmes tels que la rétro-propagation (c.f. chapitre III.1.1.2.b), qui impliquent l'existence des fibres bidirectionnelles entre neurones, vont à l'encontre du fonctionnement biologique du neurone (bien que, paradoxalement, ils soient efficaces pour imiter le comportement humain).

Au plus haut niveau de plausibilité biologique, le but est la compréhension de mécanismes physiologiques dans des structures cérébrales restreintes, par exemple l'étude du comportement des cellules pyramidales [Tiesinga *et al.* '01] ou des cellules GABAergiques [Wang & Buzsaki '96] dans l'hippocampe. Le neurone est modélisé de façon très complète, incluant ses propriétés biochimiques et électriques [Mel '93]. Mais le niveau de modélisation est trop bas, physiologiquement parlant, pour permettre d'interpréter le comportement du neurone ou de la petite structure modélisée en termes de traitement de l'information.

3.1.2. Approches intermédiaires

Dans les approches intermédiaires, la plausibilité biologique est variable, suivant que l'approche choisie est plutôt physiologique ou fonctionnelle. Souvent, la perte de plausibilité biologique dans les mécanismes cellulaires (les neurones utilisés sont relativement simples) est compensée par la prise en compte des propriétés structurelles ou architecturales caractéristiques des réseaux plus importants. Par exemple, certains réseaux modulateurs tels que les réseaux GABAergiques [Wang & Buzsaki '96] ou dopaminergiques [Onla-or & Winstein '01; Suri *et al.* '01], ou bien les relations entre certaines structures neuronales [Mitchell *et al.* '91].

La modélisation de fonctions cognitives de haut niveau passe par l'utilisation d'une échelle anatomique suffisamment grande, celle des réseaux d'aires cérébrales. Toutefois, il n'est pas

envisageable, à cette échelle, de manipuler un RNF dans lequel chaque neurone formel représente un neurone biologique. Tout d'abord, ce modèle ne serait pas calculatoirement traitable, en raison du trop grand nombre de neurones et surtout de connexions. De plus, les données anatomiques et physiologiques seraient insuffisantes pour spécifier si précisément un tel modèle [Mallot & Giannakopoulos '96]. Pour ces raisons, lors de la modélisation de grandes populations, les neurones formels sont l'abstraction fonctionnelle de groupes de neurones biologiques aux propriétés communes (ex. : ensemble de neurones GABA) , ou fonctionnellement cohérents (ex. : colonne corticales), tandis que leurs connexions représentent fonctionnellement des liaisons anatomiques [Grossberg *et al.* '02]. Par exemple, dans [Guigon *et al.* '94], un neurone formel représente une colonne corticale.

Avec ce type de modèles, l'objectif est en général de modéliser une fonction bien précise. De ce fait, un modèle est restreint à la représentation de certaines populations, bien spécifiques au traitement de la tâche ciblée. Il peut s'agir d'un mécanisme cérébral général tel que la réorganisation somato-sensorielle [Joublin *et al.* '96], l'habituation/déshabituation [Wang & Arbib '92], l'apprentissage ou la mémorisation [Carpenter & Grossberg '93; Guigon *et al.* '94; McClelland & Goddard '96], etc. Il peut également s'agir de processus cognitifs ou sensorimoteurs, tels que la coordination visuo-motrice [Burnod *et al.* '92] ou la perception tactile [Blakemore *et al.* '99].

Pour permettre une meilleure interprétation de l'activité d'un RNF en termes cognitifs, une approche intégrative, i.e. capable de tirer parti d'informations provenant de sources hétérogènes telles que la neurophysiologie, la neuroanatomie, les décompositions fonctionnelles de processus cognitifs tirées des données comportementales, etc., est nécessaire. Or, dans un RNF, la fonction et l'information sont réparties sur l'ensemble du réseau, ce qui entraîne une transparence sémantique relativement faible. En d'autres termes, il est difficile de définir *a priori* l'architecture d'un RNF de façon à implémenter une fonction précise, et, à plus forte raison, d'implémenter explicitement une quelconque décomposition fonctionnelle. Pour palier à ce défaut, Michael Arbib et coauteurs ont introduit, dans leurs travaux [Arbib '85b; Arbib *et al.* '98], une décomposition fonctionnelle au moyen de schémas [Minsky '75; Schank & Abelson '77], conjointement à une modélisation plus structurelle par RNF. Un *schéma* est ici un modèle fonctionnel générique. Une fonction cognitive donnée est implémentée par un réseau d'instances de schémas. Une instance de schéma correspond à une aire cérébrale particulière, et est implémentée par un RNF. Au final, le RCGE implémentant une fonction cognitive donnée est donc modélisé par un réseau de RNF.

3.1.3. Réseaux neuronaux à grande échelle

Les *réseaux neuronaux à grande échelle* (RNGE) constituent une tentative récente [Arbib *et al.* '95] d'utilisation des RNF pour faire le lien entre activité neuronale et résultats issus de la neuroimagerie. Il s'agit en fait d'une extension des RNF basés sur des schémas, la décomposition étant ici biologique, et non plus fonctionnelle. L'idée est d'utiliser un RNF pour représenter une aire cérébrale. Chaque neurone de ce réseau correspond à un module (par exemple une colonne corticale) de l'aire en question. Le RNGE est lui-même un réseau d'aires modélisées de cette façon, il s'agit donc d'un réseau de réseaux de neurones. En général, les neurones formels sont les mêmes dans toutes les aires, ils représentent une colonne corticale qualifiée de canonique [Horwitz *et al.* '99]. Ce sont les natures des interconnections qui différencient les aires composant le RNGE.

Il existe deux approches pour faire le lien entre ces modèles et la neuroimagerie. La première est essentiellement physiologique, et délaisse l'aspect cognitif du fonctionnement cérébral [Taylor *et al.* '00]. Ici, le but est de déterminer comment interpréter en termes d'activité neuronale les coefficients sur les liens issus de la modélisation de données de neuroimagerie par MES. Dans un premier temps, un RNGE est défini, qui permet d'exprimer des contraintes physiologiques, essentiellement sur la structure d'interconnexion des populations de neurones. Il est utilisé pour générer différents MES possibles. Les MES sont complétés par des équations d'observation, décrivant le rapport entre le système modélisé et des variables observables. Ces observations correspondent aux mesures effectuées par neuroimagerie. L'équation d'observation est spécifique à la technique de neuroimagerie utilisée (IRMf, TEP, EEG ou MEG). Par la suite, ces mesures d'activation simulées sont comparées à des mesures réelles.

Dans la seconde approche, un RNGE est utilisé pour construire un modèle correspondant à l'implémentation d'une fonction cérébrale, puis des données de neuroimagerie sont utilisées pour la validation [Arbib *et al.* '95; Tagamets & Horwitz '00]. Le travail effectué par Michael Arbib et coauteurs [Arbib *et al.* '95] illustre bien les possibilités de ce formalisme en ce qui concerne l'utilisation de données provenant de sources hétérogènes (neuroimagerie, neurophysiologie, neuroanatomie, etc.) pour la définition et la validation des modèles. Dans un premier temps, un modèle des mécanismes neuronaux de coordination visuo-motrice est construit grâce à des données physiologiques issues d'expériences portant sur le fonctionnement cérébral au niveau cellulaire chez le singe [Arbib *et al.* '95]. Puis, ce modèle est utilisé pour générer des mesures de TEP artificielles lors de la simulation de certaines tâches. Les mesures de TEP synthétique sont obtenues, hors-modèle, en intégrant l'activité synaptique totale (excitations et inhibitions) des neurones composant

une aire. Par la suite, ces mesures sont comparées à des valeurs de TEP réelles mesurées sur des humains effectuant des tâches équivalentes.

L'intérêt est que cette approche a permis d'intégrer dans un seul modèle des données issues de deux champs différents (neurophysiologie et neuroimagerie), mais aussi de deux espèces différentes (singe et humain). De plus, le RNGE a permis de relier les mesures de variations hémodynamiques issues de la TEP à l'activité neuronale. Il existe également des modèles générant des données d'IRMf synthétique, comme par exemple le modèle moteur décrit dans [Arbib *et al.* '00].

3.2. BioCaEn

BioCaEn (pour *Bio*logical *Ca*usal *En*gine) est le nom d'un formalisme de modélisation cérébrale [Pastor *et al.* '97], basé sur les modèles qualitatifs causaux et le calcul des intervalles (c.f. chapitre III.1.2 pour une description formelle). Il a été développé dans notre groupe de recherche, essentiellement par Josette Pastor et Marc Lafon [Lacotte '96; Lafon *et al.* '97; Pastor *et al.* '97; Pastor *et al.* '00], à partir d'un simulateur qualitatif dédié aux processus physiques [Travé-Massuyès *et al.* '93]. Le travail décrit dans ce mémoire s'inscrit dans le même projet de recherche, à la suite de BioCaEn, et en reprend certains éléments.

De même que dans les MES utilisés en neuroimagerie ou les RNF à grande échelle décrits précédemment, il s'agit de modéliser un réseau cérébral à grande échelle (un réseau d'aires cérébrales). A l'instar des MES, chaque aire est modélisée par une fonction mathématique, mais il s'agit ici d'une fonction de transfert qualitative linéaire, à la place de la régression linéaire utilisée dans les MES. Dans les RNF à grande échelle, chaque aire se décompose en un RNF. Dans BioCaEn, une aire peut elle-même se décomposer en un sous-réseau qualitatif causal de primitives fonctionnelles [Pastor *et al.* '00].

La grande originalité de BioCaEn est le principe de connectivité causale [Pastor *et al.* '00]. La connectivité fonctionnelle ne prend en compte que les coactivations entre des aires cérébrales. La connectivité effective consiste à contraindre l'étude de la connectivité fonctionnelle par des données ou des hypothèses anatomiques. La relation étudiée est mathématiquement réversible, ce qui va à l'encontre des observations anatomiques : les liaisons axonales sont orientées. La connectivité causale permet de prendre en compte non seulement ces liens fonctionnels et

anatomiques, mais aussi le sens de la connexion étudiée, en la modélisant sous la forme d'une relation causale.

La définition de la causalité utilisée dans BioCaEn est très répandue en physique qualitative, qui est le domaine d'origine de la modélisation qualitative. Initialement, cette définition concerne donc des phénomènes physiques se propageant dans une structure [Travé-Massuyès *et al.* '97]. Il s'agit d'une définition en trois points inspirée de de Kleer [de Kleer & Brown '86] : ordre temporel, localité, et nécessité. Le premier point signifie que la cause ne doit pas précéder l'effet, le second que la cause ne peut agir que sur ses voisins directs, et le troisième que si la cause survient, alors l'effet doit apparaître aussi, toutes choses restant égales par ailleurs. Elle se traduit par l'association d'un graphe d'influences au système d'équations qui définit le modèle. Il s'agit d'un graphe orienté qui permet de déterminer le sens (de la cause vers l'effet) de chaque relation. A noter que lorsqu'une variable (effet) a plusieurs parents (causes), les influences des parents sont traitées indépendamment, avant d'être toutes combinées pour obtenir l'influence globale.

Un autre aspect intéressant dans BioCaEn est la volonté de représenter de façon spécifique l'information transitant dans les réseaux cérébraux à grande échelle [Lafon *et al.* '97]. Comme il a été spécifié dans la contrainte portant sur la représentation de l'information (c.f. chapitre I.3), on distingue ici deux composantes : une partie correspondant à l'intensité de l'activation d'une population cérébrale, et une autre représentant la configuration des neurones activés dans la population considérée.

Ces propriétés rendent les modèles de ce type à la fois suffisamment explicites et biologiquement plausibles pour permettre d'étudier les liens entre l'activité neuronale issue de l'exécution d'une tâche cognitive, et l'activité cérébrale mesurée par des techniques de neuroimagerie. Dans [Pastor *et al.* '00], la modélisation d'une partie du cortex permet ainsi d'implémenter diverses hypothèses expliquant les variations du débit sanguin cérébral dans certaines régions, en fonction de la fréquence de présentation de stimuli visuels. Toutefois, le formalisme est relativement limitatif en ce qui concerne l'expression des mécanismes cérébraux (linéarité), ce qui amoindrit la plausibilité biologique des modèles. De plus, l'utilisation du calcul des intervalles entraîne une très grande divergence dans les résultats obtenus par simulation [Lafon *et al.* '99; Pastor *et al.* '00]. Enfin, l'absence de mécanismes d'apprentissage limite l'utilisation de ce formalisme à la modélisation des automatismes cognitifs.

3.3. Bilan

Nous allons distinguer deux catégories de travaux issus des neurosciences computationnelles : d'un côté les RNF destinés à modéliser uniquement la fonction, et de l'autre les RNGE et BioCaEn, dont la vocation est l'interprétation de données de neuroimagerie. Les premiers, qui constituent l'approche classique en neurosciences computationnelles, peuvent s'appréhender en considérant l'orientation (structurelle et/ou fonctionnelle) qui les caractérise : modèles purement fonctionnels, modèles purement structurels, et approches intermédiaires. Pour les premiers, on peut faire les mêmes reproches que ceux déjà adressés aux approches cognitives, c'est-à-dire, principalement, que l'absence de plausibilité biologique est incompatible avec l'interprétation d'une fonction en termes d'activité cérébrale mesurable par neuroimagerie. L'objectif des modèles purement physiologiques est la compréhension de mécanismes biologiques basiques, concernant une structure neuronale limitée. Ces modèles se situent en général à une définition spatiale bien inférieure à celle des techniques de neuroimagerie, ce qui rend impossible l'utilisation de telles mesures d'activation pour la création ou la validation de modèles. C'est également la petitesse de cette échelle qui empêche d'interpréter ces mécanismes physiologiques en termes de traitement de l'information.

En ce qui concerne les approches intermédiaires, l'objectif commun est de répondre à la question négligée par la modélisation en neuroimagerie, c'est-à-dire savoir *comment* les fonctions cognitives émergent des propriétés structurelles et fonctionnelles de neurones ou de populations neuronales. Ces modèles ont une plausibilité biologique au moins partielle, ce qui permet de les construire et de les valider en partie sur des résultats expérimentaux. Par exemple, dans [Monchi *et al.* '00], un RNF modélisant la mémoire de travail chez l'humain sain est lésé pour simuler un patient parkinsonien. La relation avec l'activation mesurée par neuroimagerie est très indirecte : ces modèles peuvent servir à exprimer des hypothèses qui vont être testées ensuite par des études en activation [Monchi *et al.* '00], mais ils ne fournissent pas des données numériques comparables aux données de neuroimagerie. Ces modèles ne sont donc pas concernés par l'interprétation de données de neuroimagerie.

Par rapport aux approches classiques, les travaux visant à faire le lien entre le fonctionnement cérébral et les données d'activations sont très peu nombreux, et relativement récents. Les plus répandus sont ceux basés sur les RNGE. L'approche physiologique [Taylor *et al.* '00] est de peu d'intérêt dans le cadre de ce travail, car l'interprétation de l'activité des réseaux d'aires en termes de traitement de l'information cérébrale n'y est pas envisagée.

Le travail initié par Michael Arbib [Arbib *et al.* '95] nous concerne davantage. Il rejoint certaines de nos contraintes : structure en réseau, plausibilité biologique, représentation explicite du temps, relations non-linéaires et orientées. Toutefois, relativement à nos objectifs, cette approche souffre d'un certain manque d'expressivité. Les interconnections entre les aires cérébrales sont spécifiées, sous la forme de schémas. Mais le comportement d'une aire prend la forme d'un RNF, implémentant une fonction donnée. Or, nous voulons pouvoir spécifier et modifier ces fonctions, et les RNF, qui constituent une sorte de boîte noire (c.f. chapitre III.1.1), ne sont pas un formalisme pratique à utiliser dans cette optique. De plus, la notion d'information cérébrale de haut niveau telle que nous l'avons définie dans les contraintes n'est pas considérée. Dans cette approche, on reste à un niveau de représentation assez bas : une population de neurones biologiques est représentée par un RNF. Celui-ci compte moins d'unité, certes, ce qui signifie qu'un neurone formel représente plusieurs neurones biologiques. Néanmoins, il n'y a pas d'intégration de l'information, la nature de la composante qualitative de l'information ne change pas, puisqu'elle reste modélisée par une configuration de neurones activés. En fait, dans le modèle, la transmission d'aire en aire de l'information n'est pas la propagation de l'abstraction d'une information intégrée sur une population entière, mais un ensemble de transmissions neurone par neurone, conditionné par un système de masques [Dominey & Arbib '92]. Ceci rend, en outre, plus difficile l'interprétation cognitive du traitement et de la propagation de l'information cérébrale.

BioCaEn est bien évidemment le formalisme le plus proche de nos contraintes, puisque notre travail prend place dans le même projet global de modélisation cérébrale pour l'interprétation en neuroimagerie. A ce titre, BioCaEn respecte un grand nombre des contraintes décrites dans le chapitre I.3 : architecture en réseau, représentation explicite du temps, relations causales, manipulation d'information adaptée à un haut niveau de modélisation. Toutefois, certaines différences subsistent, puisque le but de notre travail consiste avant tout à apporter des améliorations à ce qui a été réalisé auparavant. En particulier, le formalisme utilisé, les réseaux causaux qualitatifs, souffre de limitations importantes en ce qui concerne l'expressivité des modèles, se qui se traduit par une moindre plausibilité biologique. On notera ainsi l'absence de propriétés d'apprentissage, et le fait que la forme des relations est fortement contrainte, notamment s'il s'agit de relations non-linéaires. De plus, l'imprécision et l'incertitude ne sont pas explicitement représentées. Ces limitations propres au formalisme sont décrites plus en détail dans le chapitre III portant sur l'état de l'art des formalismes d'IA.

4. FORMALISATION DES CONTRAINTES

Nous avons vu que notre approche était originale, dans le sens où aucun travail existant n'a les mêmes objectifs que nous, et ne se conforme donc complètement à nos contraintes. Le but de notre travail étant de mettre au point un outil de modélisation, il est nécessaire de choisir le formalisme de base le plus adapté à nos contraintes, puis de le modifier éventuellement s'il n'y répond pas complètement. Pour faciliter ce choix, nous allons à présent ré-exprimer les contraintes du chapitre I.3 de façon plus formelle.

4.1. Causalité

La description des travaux existant en modélisation cérébrale a bien montré l'intérêt de suivre une approche causale. Néanmoins, plusieurs définitions de la causalité sont utilisées, et il nous semble nécessaire, à présent, de déterminer quelle définition nous allons utiliser. Le débat sur la définition de la causalité est vaste, il touche de nombreux champs de recherche. Dans le cas présent, il est bien entendu que nous nous plaçons dans un cadre de modélisation, et que nous nous intéressons en particulier à des systèmes numériques ou partiellement numériques.

On s'entend généralement à dire que l'un des traits caractéristiques d'une relation causale est sa régularité. En effet, quand on observe plusieurs fois que quand un évènement A survient, un autre événement B survient aussi, on s'attend, dans le futur, à observer B quand A arrive. La notion de régularité exprime cette propriété : selon Hume, dans les mêmes circonstances, on doit avoir les mêmes effets [Hume '00]. Si cette propriété tient en théorie, elle est critiquable quand on veut l'utiliser dans un cadre pratique, notamment dans un environnement bruité [Pearl & Verma '91]. En effet, elle implique que le système étudié soit très bien défini, dans le sens où la liste exhaustive des causes de chaque effet doit être connue. C'est le cas pour la définition de la causalité utilisée dans BioCaEn : chaque cause est considérée séparément des autres, puis l'influence globale est calculée. Mais considérons un effet dont certaines causes sont cachées ou ignorées. Il est possible que l'on observe plusieurs fois les mêmes causes connues, mais que les causes inconnues soient différentes, aboutissant à un effet différent. On viole alors la propriété de régularité. Or c'est précisément ce qui arrive dans l'étude du cerveau : la complexité des mécanismes mis en jeu lors du traitement de l'information empêche cette connaissance presque parfaite des causes. Il s'agit d'ailleurs de l'une des raisons pour lesquelles nous considérons les relations entre aires cérébrales comme étant de nature aléatoire. La solution consiste à considérer la régularité d'un point de vue non pas déterministe,

mais plutôt probabiliste [Pearl & Verma '91; Suppes '70]. Cette *régularité probabiliste* stipule que, la plupart du temps (et non plus toujours), l'observation des causes s'accompagne de celle de l'effet.

Pour autant, la propriété de régularité, qu'elle soit déterministe ou statistique, ne suffit pas. Ainsi, nous avons déjà vu (c.f. la causalité au sens de Granger, paragraphe 1.2) que ce n'est pas parce que deux variables sont corrélées (i.e. il y a régularité probabiliste) qu'il existe un rapport de cause à effet, la corrélation peut très bien être due à une troisième variable qui serait la cause des deux autres. La causalité est donc une relation plus forte que la simple dépendance probabiliste [Pearl '96], qui ne constitue qu'une condition nécessaire à la présence de causalité [Harthong '96]. Pour que la régularité traduise une relation causale, encore faut il déjà qu'il y ait bel et bien une relation entre les causes et les effets supposés. Hume exprime l'existence d'une relation à travers la notion de *contiguïté* [Hume '00]. Deux éléments sont contigus s'ils sont en contact ou proche l'un de l'autre, relativement à l'échelle du système, et ce aussi bien au niveau spatial que temporel. On retrouve ici la notion de localité du mécanisme contenue dans la définition de Pearl [Pearl '94]. Pour qu'une relation soit causale, on doit retrouver à la fois la propriété de régularité et celle de contiguïté. Or, l'analyse statistique de données ne permet pas à elle seule d'affirmer l'existence de relations causales, car cette analyse repose uniquement sur l'observation d'une régularité [Pearl & Verma '91].

Enfin, une relation causale est également caractérisée par son orientation. On s'entend généralement à dire que la causalité est une relation entre deux éléments, orientée de la cause vers l'effet [Hume '00; Kant '00]. Cette orientation nécessite la prise en compte d'un nouvel aspect d'une relation : sa dimension temporelle. Dans la plupart des définitions de la causalité, il y a ordonnancement temporel : la cause doit précéder l'effet [Pearl & Verma '91; Shoham '88; Suppes '70]. Ceci permet d'orienter une relation entre deux évènements. Ce trait est aussi nécessaire que les autres pour définir une relation causale. Considérons de nouveau l'étude de la corrélation entre deux variables. Même si, de manière certaine (par exemple, par observation) on a déterminé un lien physique entre les objets représentés par les variables (la condition de contiguïté est donc vérifiée), une mesure de corrélation ne permet pas d'orienter la relation. A noter que cette contrainte de *précédence temporelle* de la cause sur l'effet prend une autre forme dans certains formalismes graphiques. En effet, on y impose parfois l'utilisation de graphes sans cycles orientés [Pearl & Verma '91]. Un cycle orienté signifierait que deux éléments seraient cause l'un de l'autre, et il n'y aurait alors pas de précédence temporelle.

Dans le cadre des réseaux probabilistes (c.f. chapitre III.1.3.1.a) Judea Pearl exprime la causalité grâce au concept de *d-séparation* [Pearl '00], qui repose sur la notion de dépendance conditionnelle

probabiliste (c.f. annexe A.1.1.2). Soient trois ensembles de nœuds X, Y, et Z dans un graphe acyclique orienté, X et Y étant tous les deux connectés à Z. On dit que Z d-sépare X et Y si ces derniers sont indépendants conditionnellement à Z. Cela signifie qu'une fois qu'on connaît l'état de Z, la prise en compte d'informations concernant X n'apporte rien de nouveau par rapport à ce que l'on sait déjà de Y, et réciproquement (de nouvelles informations concernant Y n'apportent rien à X). La d-séparation traduit le fait qu'un élément donné ne peut être modulé que par l'ensemble de ses causes. En effet, une fois les causes connues (Z), l'élément considéré (par exemple, Y) est indépendant des autres éléments qui constituent le système (X), à l'exception de ses propres effets (qui dépendent de lui-même).

En résumé, nous allons utiliser une définition de la causalité en trois points : régularité probabiliste, contiguïté, et précédence temporelle. Considérons maintenant un réseau cérébral à grande échelle. Pour que l'information cérébrale soit propagée (directement) d'une aire à l'autre, il doit exister un lien anatomique entre elles : il y a contiguïté spatiale. De plus, la présence de ce lien anatomique fait que, la plupart du temps, une activation suffisante de l'aire située en amont entraîne une activation de celle en aval, et la régularité probabiliste est donc respectée. Enfin, l'information met un certain temps à parcourir la liaison anatomique, ce qui implique que l'activation de la région en aval est toujours ultérieur à celui de la région en amont, et permet de respecter le critère de précédence temporelle. Selon notre définition, le cerveau est donc bien un système causal. On considérera que l'activation de l'une des aires cérébrales décrites plus haut cause l'activation de l'autre si l'intervalle entre les deux activations est suffisamment petit à l'échelle des évènements cérébraux : on aura alors la contiguïté temporelle.

4.2. Autres contraintes

4.2.1. Formalisme graphique

Le formalisme doit prendre la forme d'un réseau orienté pouvant contenir des nœuds fonctionnellement différenciés.

4.2.2. Temps discrétisé

Une représentation explicite du temps peut se faire de façon continue ou discrète. Puisqu'il n'est pas possible d'obtenir de mesure de l'activité cérébrale en continu, il est donc inutile de prévoir une

telle représentation du temps, qui a de plus pour défaut de rendre un modèle plus complexe à traiter au niveau calculatoire. Le formalisme doit permettre de représenter le temps de façon discrète, et l'utilisation des données de neuroimagerie nécessite un découpage du temps sous la forme de périodes de taille *constante*. Le choix de ce pas d'échantillonnage doit être fait en fonction de la plus petite échelle temporelle manipulée dans le modèle. Dans le cas contraire, certaines relations dureraient moins longtemps que le pas, et passeraient pour instantanées, à cause de l'échantillonnage. Elles ne respecteraient pas, de ce fait, notre définition de la causalité (précédence temporelle).

4.2.3. Fonctions non-linéaires

La contrainte de non-linéarité décrite précédemment laisse apparaître deux types de non-linéarité. Le premier, de type RNF, revient à appliquer une fonction non-linéaire à une valeur. Le second, correspondant à la présence de processus de contrôle, signifie que toutes les entrées n'ont pas le même rôle : certaines correspondent à l'information à traiter, et d'autres auront un rôle de contrôle sur ce traitement. Formellement, cela revient à combiner de façon non-linéaire (par exemple sous forme de produit) plusieurs influences différentes. Le formalisme doit offrir suffisamment de liberté dans la définition des relations entre les nœuds pour implémenter ces deux types de relations non-linéaires.

4.2.4. Incertitude & imprécision

Formellement, la contrainte portant sur l'incertitude et l'imprécision se traduit par la nécessité de représenter des mécanismes incertains et une information imprécise. Une représentation *explicite*, par exemple sous la forme de mesures d'incertitude et/ou d'imprécision, permettrait, en outre de mieux contrôler cet aspect lors de la manipulation de l'information.

4.2.5. Information mixte

Le formalisme devra permettre de manipuler une information ayant une composante numérique correspondant au niveau d'activation, et une composante qualitative ou symbolique, représentant la configuration de neurones activés.

4.2.6. Plasticité

La contrainte de plasticité nécessite du formalisme la possibilité de modifier, au cours de la simulation, la fonction implémentée par un nœud suivant les entrées qu'ils reçoit. Cette modification peut consister à modifier la fonction, ou simplement à changer ses paramètres si les relations sont modélisées de façon paramétrique. Cela peut être effectué par un algorithme d'apprentissage propre au formalisme, mais dans ce cas, il doit être suffisamment biologiquement plausible. Cela signifie, en particulier, qu'il ne peut pas s'agir d'apprentissage supervisé, puisque seules les entrées du nœuds déterminent son évolution fonctionnelle. De plus, l'algorithme doit être adapté à l'information duale manipulée dans le modèle.

Chapitre III

MODELISATION CAUSALE

1. CARACTERISTIQUES DES FORMALISMES EXISTANTS

Dans le premier chapitre, nous avons défini un ensemble de contraintes à partir des propriétés du système à modéliser, le cerveau, des moyens d'observation existants, les techniques de neuroimagerie, et de nos objectifs de modélisation. Dans le deuxième chapitre, ces contraintes ont été reformulées en termes formels. Un certain nombre de formalismes existant dans le champ de l'intelligence artificielle sont *a priori* susceptible de les respecter. Il s'agit des réseaux de neurones, des modèles qualitatifs, et des formalismes graphiques dédiés à la modélisation de systèmes incertains ou de l'imprécis (i.e. modèles probabilistes et possibilistes). En fait, nous avons vu dans le chapitre précédent que chacune de ces familles de formalismes était déjà employée en modélisation cérébrale (notamment les réseaux de neurones), mais avec des objectifs, et donc des contraintes, différents des nôtres. Par conséquent, il nous est nécessaire d'effectuer une critique de chacun des types de formalismes, relativement à nos critères. Ceci nous permettra de sélectionner le plus adapté, afin de l'utiliser comme base de notre formalisme.

Le but de ce chapitre n'est pas de réaliser une description exhaustive des formalismes, mais de présenter les notions qui en constituent les bases, et de se concentrer ensuite sur les évolutions susceptibles de respecter les contraintes et les objectifs définis précédemment. Nous nous focalisons en particulier sur le pouvoir expressif des formalismes, ainsi que sur leurs mécanismes de propagation et d'apprentissage. Il faut souligner que dans le contexte de ce chapitre, le terme d'apprentissage n'a pas la même signification que précédemment. La notion cognitive d'apprentissage désigne le fait, pour le système cérébral, de subir une évolution fonctionnelle. Cette évolution est uniquement guidée par les entrées (c.f. chapitre I.1.3.2). Certains mécanismes d'apprentissage formels ont également pour but la modification fonctionnelle d'un modèle, le plus souvent par estimation de paramètres [Buntime '94]. Il peut s'agir, ici aussi, d'apprentissage guidé par les entrées, on parle alors d'apprentissage non-supervisé. Mais la plupart des mécanismes

d'apprentissage implémentent un apprentissage supervisé, c'est-à-dire utilisant un retour sur les sorties du modèle. Le terme d'apprentissage, employé dans un contexte formel, peut également désigner le fait de déterminer la structure d'un modèle [Ghahramani '00a].

1.1. Réseaux de neurones formels

Les *réseaux de neurones formels* (RNF) sont un formalisme apparu dans les années 40, grâce aux travaux de McCulloch et Pitts [McCulloch & Pitts '43] portant sur la mise au point d'un modèle booléen du neurone humain. Aujourd'hui, il existe un très grand nombre de types de réseaux différents, appliqués essentiellement en traitement du signal, classification, approximation de fonctions, etc., et bien sûr, en modélisation. Le principe du neurone formel (NF) est donné en annexe (annexe A.4). Nous allons décrire ici les différentes façons de combiner des NF pour former un RNF, avant d'aborder les principes d'apprentissage. La combinaison de ces deux aspects, i.e. le type de réseau et le type d'apprentissage, permet de distinguer différents types de RNF, qui seront décrits dans une troisième partie.

1.1.1. Structure d'un réseau de neurones formels

Un RNF est caractérisé par le type de NF employé et par la topologie des connexions entre les NF. En effet, la fonction globale implémentée par le RNF est déterminée par les fonctions élémentaires (i.e. par les NF), et par la façon dont ces fonctions élémentaires sont composées (i.e. par la structure du RNF).

Un réseau de neurones est habituellement décomposé *en couches* de neurones. Ces couches correspondent à des groupes de neurones successivement parcourus par l'information. Le premier groupe de NF soumis à l'information extérieure est appelé *couche d'entrée*, et le groupe des NF contenant les résultats du traitement est la *couche de sortie*. Les couches intermédiaires sont qualifiées de couches cachées. Le RNF obtenu est dit *multicouches*. La propagation de l'information dans un RNF se fait simplement en calculant les valeurs d'émission des neurones de chaque couche, et en les utilisant pour calculer celles des neurones qui forment la couche suivante. Lorsque le graphe représentant le RNF ne contient pas de cycle orienté, on parle de RNF *feed-forward* (Figure III.1.1). Quand le réseau contient de telles *boucles de rétroaction*, il est qualifié de *récurrent*, et on ne peut plus vraiment parler de couches, puisque l'information peut en quelque sorte revenir en arrière (Figure III.1.2).

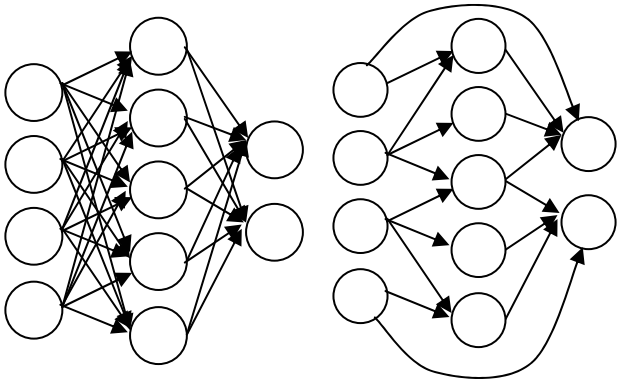

Figure III.1.1 : exemples de RNF feed-forward.

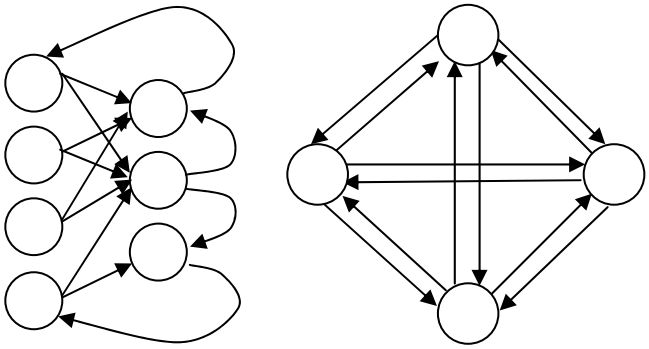

Figure III.1.2 : exemples de RNF récurrents.

Un RNF peut être *statique*, c'est-à-dire que le temps n'y est pas explicitement représenté. Ce type de réseau est utilisé notamment en classification, ou pour réaliser des approximations de fonctions non-linéaires. L'introduction du temps se fait en associant des délais aux relations existant entre les NF. Les équations modélisant les NF deviennent alors des équations différentielles. On parle de RNF *dynamiques*, et en général le temps est manipulé de façon discrète. Ce type de RNF est essentiellement utilisé en modélisation. Lors de la simulation, un RNF dynamique est converti en un RNF statique représentant le traitement réalisé par le RNF à un instant donné. Ce traitement permet de briser les cycles orientés éventuellement présents, grâce aux délais associés aux relations. On parle de représentation canonique du RNF dynamique [Dreyfus & Idan '98], et elle est résumée de façon très générale par les équations suivantes :

$$x_{k+1} = \varphi\left(x_k, u_k\right) \qquad \text{(eq. III.1.1)}$$

$$y_k = \psi\left(x_k, u_k\right) \qquad \text{(eq. III.1.2)}$$

où x_k, y_k, u_k, sont respectivement le vecteur des états des neurones de la couche cachée, le vecteur des neurones de la couche de sortie, et le vecteur des neurones de la couche d'entrée, pris à l'instant k. ψ et φ sont les fonctions implémentées respectivement par le RNF en entier et par la couche cachée uniquement, à un instant donné [Dreyfus '98].

La fonction implémentée par un RNF et l'information qu'il contient sont *réparties* sur le réseau. C'est-à-dire que même si la fonction implémentée par le RNF est analytiquement décomposable en plusieurs fonctions, une telle décomposition ne sera pas forcément possible au niveau du réseau. Ceci s'oppose à d'autres formalismes tels que les réseaux probabilistes ou qualitatifs, dans lesquels un nœud du réseau joue un rôle bien défini. Dans un RNF, un nœud tout seul n'a que peu d'importance. S'il est supprimé, le RNF continue de fonctionner (avec une perte plus ou moins importante de performance, tout de même). On dit que ce formalisme est résistant aux pannes.

1.1.2. Apprentissage

Un RNF est caractérisé par la fonction définissant chaque neurone et par la topologie du réseau. Mais en général, la description complète d'un RNF nécessite également de préciser l'algorithme utilisé pour son apprentissage.

a. Principes

L'apprentissage se fait par modifications des paramètres des NF, notamment les poids. En général, l'apprentissage constitue une phase préliminaire à l'utilisation proprement dite du RNF. Toutefois, certains types d'algorithmes qualifiés d'adaptatifs sont caractérisés par un apprentissage permanent [Marcos *et al.* '93]. En fait, ce type d'algorithme revient à alterner de très courtes phases d'apprentissage et d'utilisation sans apprentissage. A noter que l'apprentissage utilisé dans les NF est une simplification de l'apprentissage des neurones biologiques, puisqu'il ne tient compte que du mécanisme correspondant à une modification de l'efficacité de la transmission synaptique (c.f. chapitre I.1.3.2).

Durant la phase d'apprentissage, le RNF est soumis à un échantillon de données, et les poids sont sujets à modifications. L'apprentissage est terminé quand la fonction globale implémentée par le RNF dans son ensemble correspond à la fonction recherchée. Puis, dans la phase d'utilisation, les valeurs résultant de l'apprentissage sont bloquées, et le RNF est testé sur l'échantillon d'apprentissage, et sur un autre échantillon différent. Lorsque l'erreur faite par le RNF sur

l'échantillon d'apprentissage est trop importante, on dit qu'il y a *sous-apprentissage*. Lorsque cette erreur est faible, mais que le RNF ne réalise pas de bonnes performances sur l'échantillon de test, on dit qu'il y a *sur-apprentissage* : le RNF s'est trop ajusté aux données d'apprentissage, et n'est pas assez général.

Le principe fondateur de l'apprentissage est que la force de la connexion entre deux neurones augmente quand ces deux neurones émettent en même temps [Touzet '92]. L'application de ce principe a donné la *règle de Hebb* [Hebb '49] :

$$w_{ij}' = w_{ij} - \Delta w_{ij} \qquad \text{(eq. III.1.3)}$$

$$\Delta w_{ij} = \alpha x_i x_j \qquad \text{(eq. III.1.4)}$$

où w_{ij} et w_{ij}' sont respectivement l'ancien et le nouveau le poids de la relation entre les neurones x_i et x_j, Δw_{ij} est la variation de poids, et α est un facteur d'apprentissage.

b. Types d'apprentissage

On distingue les grandes familles de méthodes d'apprentissage suivant que l'algorithme nécessite ou pas d'être supervisé par l'utilisateur.

L'apprentissage *supervisé* nécessite de définir au préalable un ensemble d'*exemples*. Chaque exemple se compose de deux ensembles de données : le premier décrit un problème, et le second décrit la solution de ce problème. Lors de la phase d'apprentissage, le problème est posé au RNF, c'est-à-dire qu'on lui présente, en entrée, les données décrivant le problème. Le RNF va générer une réponse à ce problème. Cette réponse est comparée à la solution du problème, permettant de mesurer l'erreur faite par le RNF. C'est cette erreur qui va être utilisée pour réviser les poids des NF, le but étant de minimiser l'erreur faite par le RNF.

Formellement, il est donc nécessaire de définir une fonction de coût mesurant l'écart entre les sorties du réseau et la solution, et un algorithme de minimisation de la fonction de coût par rapport aux paramètres. Souvent, c'est une fonction de l'erreur quadratique qui est utilisée comme fonction de coût.

L'algorithme de minimisation fonctionne la plupart du temps par *rétro-propagation*, c'est-à-dire que les valeurs des paramètres des NF vont être révisées de façon récursive, en parcourant le RNF à l'envers : depuis les sorties vers les entrées. Ce mécanisme n'a rien de biologiquement plausible (les axones biologiques sont orientés), et s'éloigne donc de l'idée du NF vu comme un modèle du neurone biologique (NB). Pour cette raison, cette méthode a été très critiquée [Rumelhart *et al.* '86]. Paradoxalement, la rétro-propagation s'est révélée très efficace, dans de nombreuses applications [Crick '89], et elle est de plus relativement facile à employer, ce qui explique son succès.

Parfois, il n'est pas possible de fournir au RNF la solution au problème, mais seulement une indication quant au résultat à fournir par le RNF. Par exemple, on ne connaît pas la valeur exacte à attendre en sortie ; mais on sait déterminer si le résultat donné par le RNF est vrai ou faux. On parle alors d'apprentissage *semi-supervisé* (aussi appelé apprentissage par *renforcement* ou par *récompense/pénalité*). Les paramètres du RNF sont révisés selon la méthode du gradient, lequel est calculé en fonction de la sortie du réseau et d'un coefficient fixé avant l'apprentissage.

Un algorithme effectuant un apprentissage non-supervisé ne nécessite pas de connaître à l'avance les sorties du RNF. Le principe consiste à présenter des entrées au RNF jusqu'à ce que ses paramètres se stabilisent. L'apprentissage non-supervisé est essentiellement utilisé pour réaliser de la classification sans connaître les classes à priori. Il est également utilisé pour modéliser des réseaux de neurones biologiques, puisqu'il est beaucoup plus biologiquement plausible que les méthodes supervisées.

Sous certaines conditions, les RNF ont la propriété d'être des *approximateurs universels parcimonieux* [Dreyfus '98]. Le fait d'être un approximateur universel signifie qu'un RNF peut approcher, avec une précision arbitraire, toute fonction bornée suffisamment régulière. Le fait d'être parcimonieux signifie que, à précision égale, un RNF a besoin de moins (ou d'autant) de paramètres qu'un autre approximateur universel.

1.1.3. Types de RNF

Il existe un grand nombre de formalismes basés sur les RNF. En général, la structure du réseau et le type de neurones utilisés (en termes de fonction) sont contraints, de façon à obtenir des propriétés intéressantes au niveau de la fonction implémentée par le réseau, et/ou au niveau des possibilités d'apprentissage du réseau, ce qui peut se traduire par des algorithmes dédiés à ce type bien précis de réseau.

a. RNF feed-forward à apprentissage supervisé

Le *Perceptron* [Rosenblatt '59] est un RNF feed-forward tel que chaque NF d'une couche est relié aux neurones de la couche suivante. La fonction d'activation des NF est la simple somme pondérée, et la fonction de transfert est en général non-linéaire : une fonction de Heavyside ou une sigmoïde. Lorsque le perceptron ne possède qu'une seule couche, son apprentissage se ramène à une régression statistique. Lorsqu'il est multicouche, il faut employer des algorithmes plus élaborés tels que la rétro-propagation du gradient de la fonction de coût. Ce type de réseau est statique et est utilisé comme estimateur de fonctions non-linéaires, ainsi que pour effectuer de la classification.

Les RNF à *fonction de base radiale* [Hush & Horne '93] sont un cas particulier du Perceptron, composé de 3 couches. La couche d'entrée propage les entrées sans distorsion. La couche de sortie est composée d'un seul neurone non-linéaire. La couche cachée se compose de neurones dont la fonction de sortie est une fonction à noyau (i.e. à résultats dans un espace limité), par exemple une gaussienne (c.f. annexe A.1.1.5) centrée sur un point de l'espace d'entrée. Ce type de RNF est utilisé comme le Perceptron, ses performances pouvant être supérieures suivant la nature du problème.

b. RNF récurrents à apprentissage supervisé

Les RNF récurrents à apprentissage supervisé sont utilisés essentiellement pour la modélisation de systèmes dynamiques partiellement observables, ils constituent donc le type de RNF le plus intéressant dans le cadre de notre travail. Le temps peut être représenté de manière continue [Lu '00] ou discrète [Pearlmutter '95]. Le principe général est de réinjecter dans le RNF les valeurs de certains neurones calculées à des instants précédents. Lorsqu'il s'agit des neurones de sortie, on parle de *réseau de Jordan*, et lorsqu'il s'agit de neurones d'une couche cachée, on a un *réseau d'Elman*. Les algorithmes d'apprentissage utilisés sont des variantes de ceux déjà utilisés pour les RNF feed-forward [Nerrand *et al.* '93].

La plupart de ces modèles sont des modèles de type boîte noire [Sarle '94], c'est-à-dire que le but n'est pas de comprendre le fonctionnement du système, mais seulement de le simuler, de façon par exemple à effectuer des prédictions de pannes, etc. Toutefois, il est possible d'introduire la connaissance que l'on du système dans un RNF, et d'obtenir ainsi un *modèle neuronal de connaissances* [Dreyfus '98], ou modèle de type *boîte grise* [Oussar & Dreyfus '01]. Le principe consiste à introduire, dans le modèle incomplètement spécifié, des neurones ou des réseaux de neurones qui vont permettre de déterminer les éléments inconnus du modèle. Souvent, il s'agit de modéliser des bases de règles en logique des propositions [Opitz & Shavlik '93], mais il existe également des boîtes

grises dédiées à des systèmes physiques [Oussar & Dreyfus '01]. Le modèle résultant prend la forme d'un ensemble d'équations analytiques ou de réseaux de neurones aux paramètres fixes pour représenter les éléments connus du système, et d'un ensemble de réseaux de neurones plus classiques pour les éléments qu'on ne peut ou ne sait pas spécifier, le tout étant interconnecté.

c. RNF à apprentissage non-supervisé

Ce type de RNF est très utilisé pour effectuer de la classification de données sans connaître les classes *a priori*. On trouve notamment les cartes auto-organisatrices de *Kohonen*. Il s'agit d'un réseau composé de deux couches : une couche d'entrée et une couche de sortie. La couche de sortie est généralement organisée sous la forme d'une grille à deux dimensions (d'où le terme de *carte*), et chacun de ses neurones est relié à toutes les entrées. Les NF de sortie sont généralement connectés uniquement à leurs voisins, spatialement parlant, par des relations inhibitrices. Quand un stimulus est présenté, on détermine le neurone vainqueur, c'est-à-dire le neurone de la carte qui, par ses poids synaptiques, est le plus proche de cette entrée. En phase d'utilisation, c'est ce seul neurone qui décharge. En phase d'apprentissage, le processus consiste à renforcer les poids de ce neurone, et, dans une moindre mesure, de renforcer également ceux de ses voisins proches (en fonction de la distance spatiale qui les sépare du neurone vainqueur). Ce mécanisme d'apprentissage permet de faire apparaître une organisation topologique. L'intérêt réside dans la possibilité de visualiser la carte obtenue à la fin de l'apprentissage. Pour cela, on utilise les poids synaptiques pour calculer les distances entre les neurones de la couche de sortie. Les neurones sont alors représentés par des points disposés suivant ces distances. Cette représentation prend la forme d'amas de points, représentant chacun une classe déterminée lors de l'apprentissage.

Les RNF à apprentissage non-supervisé sont ceux qui se prêtent le mieux à la modélisation cérébrale, car on se place alors dans le même cadre d'apprentissage. Dans cette optique, on trouve notamment les réseaux de Hopfield et les réseaux ART. Un réseau de *Hopfield* est un RNF dont tous les NF sont des neurones binaires interconnectés, et utilisant la règle de Hebb pour l'apprentissage [Hopfield '82]. Un tel réseau implémente une mémoire associative, c'est-à-dire adressable par son contenu. La phase d'apprentissage consiste à présenter une information au réseau, dont les poids vont évoluer puis se stabiliser. Une fois que le réseau a appris l'information, il est capable de la restituer en entier si on lui en présente une version partielle ou bruitée.

Les réseaux *ART*, pour théorie de la résonance adaptative (*A*daptive *R*esonance *T*heory), ont été introduits dans les années 70 par Stephen Grossberg [Grossberg '76a, b] et Gail Carpenter

[Carpenter & Grossberg '87a; Carpenter & Grossberg '91]. Ils réalisent une classification non-supervisé et adaptative. Ils sont caractérisés par un algorithme d'apprentissage basé sur un mécanisme de résonance (une boucle) entre la couche d'entrée et la couche de sortie, et utilisant deux mémoires distinctes : une mémoire à *court* terme (MCT) et une mémoire à *long* terme (MLT). A travers les connexions neurales, la MCT peut influencer la MLT (apprentissage), et la MLT peut influencer la MCT (remémoration). Les réseaux ART représentent un type d'apprentissage très proche de celui que nous voulons inclure dans notre formalisme : il est non-supervisé, permet d'effectuer de catégoriser une information en entrée, et est adaptatif.

1.1.4. Bilan

Au premier abord, les RNF, qui sont déjà largement utilisés en modélisation cérébrale, semblent constituer le formalisme le plus biologiquement plausible, et donc le plus proche de nos contraintes. Nombre des propriétés de fonctionnement des neurones se retrouvent au niveau de la population neuronale. Tout d'abord, le traitement de l'information en deux étapes : une intégration spatio-temporelle des différentes entrées (activation) et la transmission d'une information en sortie, sous certaines conditions (émission). De plus, les fonctions utilisées lors de ce traitement peuvent être de nature non-linéaire.

Dans notre cas, les RNF récurrents dynamiques paraissent être les RNF les plus adaptés. Ils incluent une représentation discrète du temps compatible avec notre contrainte de temporalité, et leur topologie n'est pas contrainte. De plus, une liaison entre deux neurones est orientée, le délai qui lui est associé garantit la précédence temporelle, et si un neurone émet dans un contexte donné (i.e. pour un certain état des neurones en amont), il émettra quand il se retrouvera dans le même contexte. On peut donc considérer que ce type de réseau respecte notre définition de la causalité. Nous avons vu qu'il existe deux façon d'aborder ce type de RNF : une approche boîte noire et une approche boîte grise.

Le principe de l'approche boîte noire est d'utiliser les facultés d'apprentissage supervisé des RNF pour ajuster le réseau de neurones à une série temporelle de données, à l'instar d'un modèle d'espace d'états. Le RNF n'est alors pas autre chose qu'un outil statistique efficace utilisé pour approcher une fonction, qui est en général non-linéaire. En réalité, les mécanismes qui régissent l'apprentissage de la plupart des types de RNF décrits plus haut, à l'exception de certains réseaux à apprentissage non-supervisé (les réseaux ART, notamment) peuvent se ramener à des techniques de régression statistique [Sarle '94]. La grande différence qui sépare les deux techniques est que

l'approche statistique est analytique : il est nécessaire de spécifier partiellement un modèle, d'avoir une idée du type de fonction que l'on veut approcher. L'approche connexionniste est une approche boîte noire, qui permet de déterminer une approximation de la fonction sans savoir *a priori*, et qui en contrepartie utilise des méthodes qui ne sont pas forcément optimales [Sarle '94].

Dans l'approche boîte noire, tous les neurones doivent être similaires (ils doivent être définis par les mêmes fonctions) afin de pouvoir utiliser un algorithme d'apprentissage, et de conserver au réseau la propriété d'approximateur parcimonieux universel [Dreyfus '98]. Ceci constitue une limite très importante, puisque l'expressivité du formalisme s'en trouve fortement réduite : on ne peut pas spécifier exactement les fonctions désirées. Il est difficile de traduire les hypothèses ou les connaissances disponibles en un tel réseau. De plus, un RNF de ce type est caractérisé par son opacité sémantique, ce qui signifie qu'une fois l'apprentissage terminé, il est difficile de l'interpréter, de déterminer le rôle d'un neurone ou même d'un groupe de neurones.

L'approche boîte grise est plus adaptée à nos objectifs de modélisation, puisqu'elle consiste à construire un modèle basé sur des connaissances (et également, dans notre cas, des hypothèses). Les connaissances figurent sous forme d'équations, et les parties qu'on ne peut pas ou ne sait pas spécifier sont remplacées par des réseaux de neurones. Le problème d'opacité sémantique refait son apparition à ce niveau, puisqu'il est malaisé d'effectuer une interprétation analytique du rôle du RNF qui a appris. De plus, l'apprentissage des différents RNF composant le modèle ne se fait pas globalement, mais bien RNF par RNF. Ce type de modélisation nécessite donc d'avoir accès à la valeur de sortie de chaque RNF, ce qui n'est pas notre cas.

Considérons maintenant la contrainte portant sur la représentation de l'incertitude et de l'imprécision. Certains formalismes permettent d'utiliser un RNF en combinaison avec l'une des deux grandes théories de représentation de l'incertain et de l'imprécis : la théorie des probabilités et celle des nombres flous (c.f. annexes A.1 et A.2). La dénomination de *RNF flous* laisse penser que ce formalisme permet de manipuler une information imprécise, mais il s'agit en réalité d'utiliser des RNF classiques comme outils annexes pour le développement de bases de règles floues [Wang *et al.* '01], un neurone codant un opérateur flou ou une fonction d'appartenance. En ce qui concerne les RNF *probabilistes*, il en existe deux types différents. Tout d'abord, ceux dans lequel le comportement du neurone est aléatoire, et suit une distribution de probabilités [Woodburn *et al.* '00] : pour une même valeur, il peut émettre ou ne pas émettre. Ainsi, la *machine de Boltzman* est une extension probabiliste du réseau de Hopfield destinée à palier à certains défauts de ces derniers, notamment leur sensibilité aux minima locaux [Anderson & Rosenfeld '88]. Les RNF probabilistes du deuxième

type propagent une probabilité d'émettre, au lieu d'une valeur d'émission. Par exemple, dans [Hintz-Madsen *et al.* '99], un RNF est utilisé pour une tâche de classification. Chaque sortie du RNF correspond à la probabilité que l'entrée appartienne à une certaine classe. Dans les trois cas, le formalisme ne permet pas de respecter notre contrainte.

Une différence importante existe entre un neurone biologique simple et une population de NB, en ce qui concerne la nature de l'information manipulée (c.f. chapitre I.1.2). L'information manipulée dans un RNF est soit une valeur binaire, soit une valeur réelle (comprise entre 0 et 1 la plupart du temps). Il n'y a donc pas de possibilité de *propager* de l'information de nature symbolique. Pourtant, les RNF ont déjà été utilisés pour *manipuler* de l'information symbolique, en associant certains neurones à des symboles [Wermter & Sun '00]. Par exemple, dans leur modèle TRACE, James McClelland et David Rumelhart ont utilisé un RNF dans lequel ils associent certains neurones à des syllabes ou à des groupes de syllabes [McClelland *et al.* '86]. Cette méthode pose des problèmes dans le cadre d'un modèle nécessitant de manipuler un grand nombre de symboles, ou dans un modèle dynamique, dans lequel de nouveaux symboles peuvent apparaître ou disparaître au cours de la simulation. Dans notre cas, cet aspect des RNF rend difficile le respect de la contrainte concernant la représentation de l'information cérébrale de haut niveau.

En ce qui concerne la contrainte d'apprentissage, nous pouvons d'ores et déjà écarter les méthodes supervisées ou semi-supervisées, qui ne sont pas biologiquement plausibles. Il faut de plus remarquer que certains algorithmes d'apprentissage dans les RNF, tels que la rétro-propagation, imposent des relations réciproques simultanées, ce qui casse la causalité. En revanche, les algorithmes non-supervisés sont intéressants dans notre cas, car ils correspondent à notre définition de l'apprentissage : le modèle s'adapte à l'information reçue en entrée, sans intervention extérieure. De plus, ils sont en général d'inspiration biologique. Les réseaux ART, en particulier, modélisent le processus de catégorisation au niveau d'une population de neurones. Bien sûr, le formalisme lui-même ne peut pas être retenu tel quel, car il est très contraint, spécifiquement dédié à la modélisation d'une tâche bien précise, et ne respecte pas la plupart de nos contraintes. Toutefois, les mécanismes implémentés peuvent être une source d'inspiration pour notre propre processus d'apprentissage non-supervisé.

1.2. Simulation qualitative

Les réseaux qualitatifs sont une méthode de modélisation issue de la physique qualitative de la fin des années 70 [Hayes '85], puis adaptée de façon plus générale aux systèmes du monde réel. Le but initial était de pouvoir simuler ou prévoir, grâce au calcul qualitatif, le comportement de systèmes trop complexes ou trop peu formalisés pour être modélisés par des méthodes numériques [Travé-Massuyès *et al.* '97]. Les algèbre qualitatives sont présentées en annexe (annexe A.3.1). Par la suite, le calcul qualitatif a été étendu à l'utilisation des intervalles réels, donnant naissance aux algèbres semi-qualitatives (annexe A.3.2), permettant de réaliser des modèles à la fois qualitatifs et numériques. En raison de la nature en partie numérique de l'information que nous voulons modéliser, nous allons nous intéresser tout particulièrement aux modèles utilisant cette algèbre. En dehors de l'algèbre utilisée pour représenter et manipuler l'information, un modèle qualitatif est caractérisé par deux propriétés : la méthode qu'il utilise pour modéliser le temps, et la présence explicite ou pas de la causalité.

1.2.1. Temps

La gestion du temps peut se faire de façon statique, c'est à dire qu'on ne va prendre en compte que les états d'équilibre du système dans la modélisation. Dans ce cas, les relations entre les variables sont fonctionnelles, et le modèle qualitatif se présente sous la forme d'un système d'équations linéaires qualitatives ou d'équations algébriques, suivant l'algèbre utilisée. Ce type de modèle est adapté aux systèmes dans lesquels, d'une part, les états d'équilibre durent relativement longtemps, et, d'autre part, les temps de réponses des variables aux influences sont si faibles et leurs vitesses de variations si grandes que l'on peut considérer les transitions d'un état d'équilibre à un autre comme étant instantanées.

Si, au contraire, il est nécessaire de représenter tous les états transitoires entre deux états d'équilibre, le modèle utilisé est de type dynamique. Il est représenté par un système d'équations différentielles, en raison des origines du formalisme (sciences physiques). On peut effectuer le découpage du temps de façon évènementielle, chaque instant correspondant à un changement d'état du système. Le principe est de calculer les états suivants à partir de l'état courant et des contraintes spécifiées dans le modèle, et de réitérer ce processus à chacun de ces successeurs : on obtient un arbre de tous les états possibles, qui peut comporter un très grand nombre de nœuds. Il est également possible d'utiliser un découpage du temps avec un pas constant, en incluant une horloge unique dans le

modèle, qui permettra de mesurer chaque variable à intervalles réguliers. Bien que nécessitant plus de ressources, ce procédé permet de contrôler la précision temporelle des calculs.

1.2.2. Causalité

A l'instar des autres formalismes décrits précédemment, la causalité peut être représentée explicitement dans un modèle qualitatif. Pour cela, on utilise des diagrammes de causalité analogues aux graphes acycliques orientés employés dans les formalismes probabilistes et possibilistes. Cette fois encore, on considère qu'il n'y pas de boucle instantanée. Ces diagrammes vont compléter les équations déjà présentes dans le modèle, en leur imposant des contraintes. En effet, ces équations sont réversibles, il est possible de calculer le membre de droite si on connaît celui de gauche, et réciproquement. Le diagramme causal associé à l'équation permet d'orienter la relation, et d'empêcher cette réversibilité. Par exemple, considérons la loi d'Ohm [Travé-Massuyès *et al.* '97] :

$$U = RI \qquad \text{(eq. III.1.5)}$$

qui permet de calculer la tension U du courant traversant un conducteur en fonction de la résistance R du conducteur et de l'intensité I du courant. Dans un modèle qualitatif non-causal, il est également possible de calculer:

$$I = \frac{U}{R} \text{ et } R = \frac{U}{I} \qquad \text{(eq. III.1.6)}$$

Mais si on se place dans un modèle causal, conforme au graphe de la Figure III.1.3, alors il est possible de calculer U par propagation des valeurs de I et R (qui sont ses causes), mais par contre il est impossible de calculer R à partir de U et I ou bien I à partir de U et R, comme c'était le cas dans le modèle non-causal.

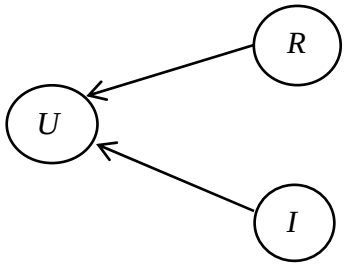

Figure III.1.3 : diagramme causal d'un réseau qualitatif simple.

1.2.3. Modèles semi-qualitatifs

Dans un modèle semi-qualitatif [Travé-Massuyès *et al.* '97], les relations sont définies formellement par des fonctions appelées *influences*. Suivant les propriétés (temps, causalité, nature de l'information, méthode de propagation) du formalisme utilisé, les influences doivent prendre des formes plus ou moins contraintes. Nous allons laisser de côté les formalismes semi-qualitatifs dynamiques non-causaux tels que *QPE* (*Q*ualitative *P*rocess *E*ngine) [Forbus '84], ou bien *Qsim* (*Q*ualitative *sim*ulation) [Kuipers '86] et ses variantes [Berleant & Kuipers '92; Kuipers '01], pour nous concentrer sur les modèles causaux, et plus particulièrement ceux qui manipulent le temps sous la forme de pas réguliers (et non pas de façon évènementielle). Dans ce cas-là, les relations sont de nature linéaire.

Ca-En [Travé-Massuyès *et al.* '97] fait partie de cette catégorie de formalismes. Une influence entre deux variables y est représentée par une fonction différentielle linéaire du premier ordre :

$$\frac{\alpha_1 dy_t}{dt} + \alpha_2 y_t = \beta x_{t-d} \qquad \text{(eq. III.1.7)}$$

où x est la cause et y l'effet, les deux étant des intervalles. α_1, α_2, β et d sont des intervalles ou des réels. Le formalisme est destiné aux systèmes physiques, ce qui explique la forme sous laquelle la relation est exprimée. Cette équation permet, dans le cas du temps échantillonné, d'exprimer la relation sous une forme plus comparable aux autres familles de formalismes causaux [Travé-Massuyès *et al.* '97] :

$$y_{t+1} = ay_t + bx_{t-d} \qquad \text{(eq. III.1.8)}$$

Trois grandeurs physiques sont prises en compte : un gain (paramètre multiplicatif), un délai (d) et un temps de réponse. Les paramètres a et b dépendent du temps de réponse, tandis que b dépend également du gain. De plus, une condition pèse sur b : il s'agit d'une expression logique déterminant l'état de l'influence (actif ou inactif) à un instant donné. Le *temps de réponse* désigne le temps mis par y pour se stabiliser après une variation de x.

Dans le cas où plusieurs variables $x^{(1)},\ldots,x^{(n)}$ en influencent une autre, le calcul de leurs influences se fait séparément, c'est-à-dire que l'on calcule des influences marginales $y^{(1)},\ldots,y^{(n)}$, de la même façon que dans (eq. III.1.8) :

$$y_{t+1}^{(i)} = a^{(i)} y_t^{(i)} + b^{(i)} x_{t-d^{(i)}}^{(i)} \qquad \text{(eq. III.1.9)}$$

Puis, ces influences marginales sont combinées au moyen d'une somme, pondérée par des coefficients $w^{(i)}$:

$$y_{t+1} = \sum_{i=1}^{n} w^{(i)} y_{t+1}^{(i)} \qquad \text{(eq. III.1.10)}$$

Dans BioCaEn, c'est une loi de combinaison $\circ$ qui est utilisée, par exemple la somme ou la multiplication :

$$y_{t+1} = y_{t+1}^{(1)} \circ \ldots \circ y_{t+1}^{(n)} \qquad \text{(eq. III.1.11)}$$

Cette loi de combinaison doit traiter chaque influence marginale de la même façon, indépendamment de sa position dans l'expression.

A l'instar des autres formalismes présentés dans ce chapitre, il existe des algorithmes d'apprentissage (dans le cadre de la modélisation qualitative, on parle également d'identification) pour les modèles semi-qualitatifs, dont le but est d'estimer automatiquement les paramètres. Toutefois, ce champ est beaucoup moins développé qu'il ne l'est dans les autres formalismes, (comme les réseaux probabilistes, par exemple). En effet, comme les réseaux qualitatifs sont dédiés à la modélisation de systèmes physiques observables ou pour le moins relativement bien connus, les paramètres du modèle sont souvent issus de mesures ou d'estimations réalisées par des experts. En fait, des algorithmes d'apprentissage existent pour les modèles non-causaux tels que Qsim [Kay *et al.* '00], mais ces algorithmes utilisent des mécanismes statistiques incompatibles avec le cas causal.

1.2.4. Modifications

Le problème du calcul semi-qualitatif est, d'une part, qu'il provoque une rapide expansion des intervalles manipulés (c.f. annexe A.3.2.2). D'autre part, les résultats de certains calculs ne sont pas définis dans l'algèbre des intervalles, alors que leurs équivalents dans $\mathbb{R}$ le sont, par exemple, pour $a \neq 0$:

$$\left[-a, a\right]^{-1} = \left]-\infty, -a^{-1}\right] \cup \left[a^{-1}, +\infty\right[\qquad \text{(eq. III.1.12)}$$

Ici, le résultat n'est pas un intervalle, mais une union d'intervalles. C'est également le cas pour le domaine de variations de la fonction inverse d'une fonction non-monotone [Travé-Massuyès *et al.* '97].

Pour résoudre ce problème, il est possible d'utiliser l'intervalle le plus petit encadrant l'union d'intervalles [Travé-Massuyès *et al.* '97] :

$$intervalle\left(\bigcup_i [a_i, b_i]\right) = \left[\min_i a_i, \max_i b_i\right] \qquad \text{(eq. III.1.13)}$$

Pour cela, il est nécessaire d'utiliser l'ensemble des réels étendus, qui permettent de manipuler $-\infty$ et $+\infty$ comme des réels, et donc de représenter (eq. III.1.12) par $]-\infty,+\infty[$ [Lafon '00]. Mais cela renforce encore plus le phénomène d'expansion des intervalles. Une autre solution consiste à propager un ensemble d'intervalles au lieu d'un seul intervalle, pour représenter une valeur donnée, au risque de voir le nombre d'intervalles augmenter au cours de la simulations [Hyvönen '92].

Enfin, il est également possible de remplacer les intervalles par des ensembles flous à fonction d'appartenance trapézoïdale ou triangulaire [Bonarini & Bontempi '94; Travé-Massuyès *et al.* '97] (c.f. annexe A.2.1). L'utilisation des opérateurs flous à la place des opérateurs définis sur l'algèbre des intervalles permet alors de limiter l'accroissement de l'imprécision au cours de la simulation. L'utilisation d'intervalles flous revient en fait à propager deux intervalles pour représenter une information : une évaluation optimiste et une évaluation pessimiste, correspondant respectivement au noyau et au support de l'ensemble flou (c.f. annexe A.2.1) [Dubois & Prade '94].

1.2.5. Bilan

Les réseaux qualitatifs représentent de bons candidats à la modélisation cérébrale telle qu'elle a été précédemment définie. En effet, ils sont particulièrement adaptés aux cas pour lesquels on n'a besoin que de résultats approchés, ou bien au contraire pour lesquels on n'a pas de données assez précises pour réaliser une simulation numérique. La topologie du réseau n'est pas contrainte, et certaines versions dynamiques des modèles semi-qualitatifs supportent une représentation discrète du temps respectant la contrainte de temporalité. De plus, il existe des réseaux dynamiques semi-qualitatifs dans lesquels les relations, définies sous une forme fonctionnelle, sont présentées comme causales. Par rapport à notre définition de la causalité, on peut dire que l'aspect de contiguïté est respecté, au même titre que pour les réseaux probabilistes et possibilistes. De même pour la notion

de précédence temporelle, grâce à l'orientation des relations et à l'affectation d'un délai aux relations. Enfin, la définition déterministe des relations assure la régularité.

En ce qui concerne l'information manipulée, les modèles qualitatifs sont, par définition, capables de représenter une information symbolique. Il est également possible de représenter une information numérique sous forme d'intervalles ou d'ensembles flous particuliers, dans des modèles semi-qualitatifs. La combinaison de ces deux aspects dans un seul modèle nécessite donc seulement de définir des variables sur des domaines différents, ce qui permet de remplir la contrainte sur la représentation de l'information cérébrale à un niveau intégré. Cet aspect a notamment été traité par Marc Lafon [Lafon *et al.* '99], avec le formalisme BioCaEn, qui avait été précédemment développé dans notre groupe de recherche.

L'utilisation de simples intervalles pour la partie numérique ne permet pas d'avoir une mesure de l'imprécision ou de l'incertitude. Si des ensembles flous trapézoïdaux ou triangulaires sont utilisés à la place, la fonction d'appartenance permet de quantifier l'imprécision des données. Toutefois, dans les deux cas, se pose le problème de l'accroissement excessif de l'imprécision, inhérent à l'utilisation de l'algèbre des intervalles [Travé-Massuyès *et al.* '97]. Le fait de manipuler des ensembles flous déplace le problème au niveau du noyau de l'ensemble flou, qui est un intervalle, mais ne résout pas le problème. De plus, le fait d'employer des relations non-linéaires accroît encore l'expansion des intervalle (et donc l'imprécision).

Dans un réseau causal dynamique semi-qualitatif de type Ca-En, les fonctions employées sont linéaires. En fait, un certain type de non-linéarité peut être inclus dans la relation grâce aux conditions qui entrent dans la définition d'une influence. Considérons maintenant un nœud soumis à plusieurs influences simultanées. L'influence de chaque nœud est calculée séparément des autres, puis toutes les influences sont combinées sans distinction, afin de calculer l'influence globale. Or, rien n'indique qu'il s'agit du processus qui se déroule lors du traitement de l'information cérébrale à un niveau intégré, ce qui fait de cette contrainte un élément limitatif relativement important.

Enfin, il n'existe pas, dans les modèles qualitatifs, de processus d'apprentissage respectant la contrainte que nous avons définie, c'est-à-dire réalisant une adaptation du réseau à l'information qu'il reçoit en entrée. A noter qu'il n'existe pas non plus d'algorithme de paramétrage automatique pour les modèles causaux dynamiques semi-qualitatifs. Ce sont principalement pour ces raisons que l'approche de la simulation qualitative avait été abandonnée, à la suite du développement de BioCaEn dans notre groupe de recherche.

1.3. Formalismes de l'incertain et de l'imprécis

On peut distinguer deux familles de formalismes graphiques causaux destinés à la modélisation de systèmes avec prise en compte explicite de l'incertitude ou de l'imprécision : les réseaux *probabilistes* et les réseaux *possibilistes*. Par le terme de réseaux probabilistes, nous regroupons ici un ensemble très large de formalismes issus (et appliqués dans) de nombreux champs de recherches différents, et à la nomenclature profuse et, parfois, confuse : réseaux bayésiens [Pearl '99], réseaux causaux [Jensen *et al.* '90], réseaux de croyances [Druzdzel & Simon '93], modèles graphiques [Buntime '94], modèles d'équations structurelles [Pearl '98], classificateurs bayésiens [Friedman *et al.* '98], modèles linéaires généralisés [Breslow '96], modèles gaussiens [Roweis & Ghahramani '99], graphes ou modèles (d'espaces) d'états [Fahrmeir '99], filtres de Kalman [Anderson & Moore '92], modèles à mélange d'experts [Jordan & Jacobs '94], modèles de Markov cachés [Minka '99], réseaux bayésiens dynamiques [Kjaerulff '93]… Ils ont tous en commun l'utilisation plus ou moins explicite de la conception bayésienne des probabilités pour représenter les relations unissant les divers éléments composant un modèle. Les réseaux possibilistes sont les pendants des réseaux probabilistes, utilisant la théorie des possibilités au lieu de la théorie des probabilités. Ils sont nettement moins développés que leurs homologues : la théorie des possibilités étant beaucoup plus jeune que celle des probabilités, elle est également, à l'heure actuelle, moins utilisée.

Dans la description que nous faisons, nous supposons que le lecteur est familier des deux théories précédemment citées. Dans le cas contraire, des rappels sur les notions de probabilités et de possibilités sont disponibles en annexe A. Nous allons dans un premier temps nous attacher à dépeindre les réseaux bayésiens et possibilistes classiques, avant de passer à leurs extensions dynamiques, qui nous intéressent plus particulièrement dans le cadre de ce travail.

1.3.1. Principes

a. Réseaux bayésiens

Les réseaux bayésiens (RB) sont un outil permettant de modéliser un système du monde réel. Il sont utilisés dans de nombreux domaines, et pour de nombreuses tâches différentes : modélisation de processus (physiques, biologiques, économiques, etc.), planification, systèmes experts, diagnostic de pannes, prise de décision, etc. Un RB est composé, d'une part, d'un ensemble de variables aléatoires (v.a.) $\{X_1,\ldots,X_n\}$, dont chacune représente un des éléments constituant le système, et, d'autre part, d'un ensemble de dépendances entre les variables, qui symbolisent les relations existant

entre les éléments du système réel. Ces dépendances prennent la forme d'une distribution de probabilités conjointes $P(X_1,\ldots,X_n)$ incluant toutes les variables composant le modèle [Pearl '00].

Considérons dans un premier temps un RB de v.a. discrètes. D'après la définition de la probabilité conditionnelle (c.f. annexe A.1.1.2), on peut ordonner arbitrairement les variables de façon à décomposer $P(X_1,\ldots,X_n)$ en un produit de distributions de probabilités conditionnelles :

$$P(X_{1:n}) = P(X_1)\prod_{j=2}^{n} P\left(X_j \middle| X_{1:j-1}\right) \text{ (\textit{règle d'enchaînement})} \qquad \text{(eq. III.1.14)}$$

où $X_{1:n}$ désigne $X_1,\ldots,X_n$. Pour considérer que le RB est causal, il est nécessaire que la décomposition ne soit pas arbitraire, mais soit au contraire réalisée suivant une analyse des mécanismes sous-jacents.

Maintenant, supposons qu'une variable X_j ne dépende pas de $\{X_1,\ldots,X_{j-1}\}$, mais seulement d'un sous-ensemble de $\{X_1,\ldots,X_{j-1}\}$. On note PA_j ce sous-ensemble, les nœuds qui le composent sont appelés *parents* de X_j, et X_j est un *descendant* de chacun de ses parents. On a alors :

$$P\left(X_j \middle| X_{1:j-1}\right) = P\left(X_j \middle| PA_j\right) \qquad \text{(eq. III.1.15)}$$

Le produit (eq. III.1.14) est grandement simplifié en appliquant ce raisonnement à toutes les v.a. qui composent le modèle. De cette façon, à la distribution de probabilités conjointes, très complexe à représenter, on peut substituer un ensemble de distributions de probabilités conjointes plus faciles à manipuler. Pour des raisons de lisibilité, on utile cette factorisation pour représenter le modèle sous la forme d'un *graphe orienté*, dans lequel les nœuds correspondent aux v.a. et les arcs aux dépendances, une variable X_j étant reliée à ses parents. Les relations sont orientées des parents vers la variable. Un arc entre deux variables signifie donc une dépendance. Une variable X_j a la propriété d'être indépendante de ses non-descendants $\{X_1,\ldots,X_{j-1}\} \setminus PA_j$ conditionnellement à ses parents PA_j. De par les propriétés de la distribution conjointe sous sa forme factorisée (eq. III.1.15), le graphe orienté obtenu est *acyclique* (dans le sens où il ne contient pas de cycle *orienté*).

Prenons par exemple un modèle contenant trois variables $\{X, Y, Z\}$, avec la distribution conjointe $P(X, Y, Z) = P(X)P(Y|X)P(Z|X)$. La Figure III.1.4 est la représentation graphique de cette distribution. Elle illustre bien le fait que Y et Z sont indépendantes conditionnellement à X.

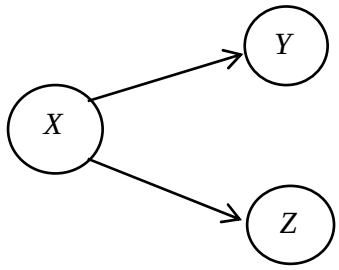

Figure III.1.4 : réseau bayésien simple représenté sous la forme d'un graphe orienté acyclique.

En raison de nos contraintes de modélisation, nous allons nous intéresser plus particulièrement aux variables numériques continues. L'intérêt de manipuler des variables numériques est de pouvoir, dans certains cas, représenter leurs distributions en utilisant un nombre réduit de paramètres (on parle de représentation paramétrique), alors qu'une quantité importante de données est nécessaire pour représenter la distribution d'une v.a. qualitative, dépendant de la taille de son domaine de définition. De plus, la nature numérique des v.a. permet d'exprimer de façon fonctionnelle les distributions de probabilités conditionnelles d'un réseau de v.a., et donc les relations définissant le RB. Soient deux v.a. dépendantes X_1 et X_2, la distribution de probabilités conditionnelles $p(X_2|X_1)$ peut être définie en utilisant la fonction f quelconque :

$$X_2 = f(X_1, W) \qquad \text{(eq. III.1.16)}$$

où W est une v.a. indépendante de toute autre v.a. modélisée. Elle représente un bruit ou une influence provenant d'un élément non modélisé [Pearl '96]. On dit que W est *exogène*, i.e. elle représente quelque chose d'externe au modèle, par opposition à X_1 et X_2 qui sont *endogènes*. La distribution de probabilités conditionnelles est ici décomposée en une partie déterministe, le calcul représenté par f, et une partie probabiliste, la v.a. W. Dans le cas de v.a. continues, W est en général une v.a. normale (annexe A.1.1.5), qui offre de gros avantages calculatoires. En particulier, une combinaison linéaire de variables normales indépendantes est également une variable normale.

Pour reprendre l'exemple décrit en (eq. III.1.16), si X_1 et X_2 sont linéairement dépendantes, il est possible de définir la distribution de probabilités conditionnelles $p\left(X_2 \middle| X_1\right)$ par :

$$X_2 = f\left(X_1\right) + W \qquad \text{(eq. III.1.17)}$$

où f est une fonction linéaire et W une v.a. normale de moyenne nulle : $W \sim N\left(0, \sigma^2\right)$ [Ghahramani '98]. La distribution de probabilités conditionnelles suit par conséquent la loi suivante :

$$p\left(X_2 \middle| X_1 = x_1\right) \sim N\left(f\left(x_1\right), \sigma^2\right) \qquad \text{(eq. III.1.18)}$$

A noter qu'il est possible de manipuler des v.a. continues et discrètes dans le même réseau [Lauritzen & Jensen '99; Olesen '93].

L'*inférence* dans un RB consiste à utiliser la règle de Bayes comme décrit dans l'annexe A.1.1.6, de façon à calculer la distribution de probabilités conjointes des variables cachées, conditionnellement aux variables observées. Par exemple, si Z est observée, on pourra calculer $P\left(X, Y \middle| Z = z\right)$, et à partir de là, en déduire des distributions de probabilités marginales telles que $P\left(X \middle| Z = z\right)$ ou $P\left(Y \middle| Z = z\right)$. Le calcul exact de toutes les distributions de probabilités conditionnelles d'un RB (par exemple pour une mise à jour) constitue une tâche *NP-difficile*, ce qui signifie que le coût calculatoire peut-être relativement important. Dans ce cas, on s'oriente en général vers un traitement approché. La rigidité du formalisme peut constituer un autre défaut : l'introduction d'une nouvelle variable peut nécessiter la révision de l'ensemble des distributions de probabilités conditionnelles.

L'*apprentissage* dans un RB désigne deux actions différentes. D'une part, ce terme peut désigner l'ajustement des distributions de probabilités conditionnelles du réseau à un certain nombre d'observations [Binder *et al.* '97; Lauritzen '95]. Les algorithmes sont basés sur diverses méthodes issues des statistiques telles que la méthode du maximum de vraisemblance (c.f. annexe A.1.2). D'autre part, l'apprentissage peut consister à déterminer la structure d'un réseau bayésien (i.e. des interdépendances des variables) à partir d'observations [Peña *et al.* '00].

b. Réseaux possibilistes

Actuellement, les réseaux possibilistes (RP) sont utilisés essentiellement pour la modélisation de systèmes logiques (à travers la logique possibiliste) [Benferhat *et al.* '01] ou de systèmes qualitatifs [Borgelt *et al.* '98], pour la classification [Borgelt & Gebhardt '99] et pour les problèmes de prise de décision [Dubois & Prade '97]. Les RP reposent sur les mêmes principes que leurs homologues probabilistes. Ils prennent la forme d'un graphe acyclique orienté, la différence résidant dans l'expression numérique des relations incertaines, qui prennent la forme de distributions de possibilités. En utilisant une des différentes définitions des possibilités conditionnelles existant (c.f. annexe A.2.2.2), la distribution conjointe peut être décomposée. Pour un réseau de n variables $X_1,\ldots,X_n$, avec une interprétation quantitative des possibilités conditionnelles, on se rapproche de la règle d'enchaînement utilisée dans les réseaux bayésiens :

$$\Pi\left(X_{1:n}\right)=\Pi\left(X_1\right)\prod_{i=2}^{n}\Pi\left(X_i\middle|X_{1:i-1}\right) \qquad \text{(eq. III.1.19)}$$

Dans une RP, on ne manipule pas des v.a. mais des variables floues (c.f. annexe A.2.1.3). Considérons une relation entre deux variables floues x et y de domaines U et V, décrite par l'équation suivante :

$$y=f\left(x\right) \qquad \text{(eq. III.1.20)}$$

Les fait d'inférer revient ici à propager une distribution de possibilités, c'est-à-dire à calculer π_y à partir de π_x et de f. Ce calcul s'effectue en utilisant le principe d'extension [Dubois & Prade '94] déjà utilisé par Zadeh pour définir ses opérateurs sur les variables floues (c.f. annexe A.2.1.3) :

$$\pi_{f(x)}\left(v\right)=\begin{cases}\sup\limits_{u\in U:\,f(u)=v}\left(\pi_x\left(u\right)\right) & \text{si } \exists u:\,f\left(u\right)=v\\ 0 & \text{sinon}\end{cases} \quad \textit{(principe d'extension)} \qquad \text{(eq. III.1.21)}$$

La possibilité que y prenne la valeur v est égale à la possibilité maximale que x prenne comme valeur une valeur u dont l'image par f est v. Cela se traduit en termes de mesure de possibilité par :

$$\Pi\left(y=v\right)=\Pi\left(x=f^{-1}\left(v\right)\right) \qquad \text{(eq. III.1.22)}$$

Dans le cas de fonctions de plusieurs variables non-interactives (c.f. chapitre VI.2.1.3), (eq. III.1.21) devient :

$$\pi_{f(x_{1:n})}(v) = \begin{cases} \sup\left(\min\left(\pi_{x_1}(u_1),\ldots,\pi_{x_n}(u_n)\right)\right) & \text{si } \exists u_{1:n} : f(u_{1:n}) = v \\ 0 & \text{sinon} \end{cases} \qquad \text{(eq. III.1.23)}$$

Le principe d'extension s'applique également à des variables interactives, mais la propagation implique des calculs plus complexes [Dubois & Prade '94].

Au niveau du réseau, l'inférence se fait en combinant le principe d'extension à des algorithmes dédiés aux réseaux bayésiens [Borgelt *et al.* '98]. Des algorithmes d'apprentissage existent également, que ce soit pour déterminer la structure du réseau [Sangüesa & Cortés '97] ou les distributions de possibilités correspondant aux relations [Borgelt *et al.* '98]. Ils ont l'avantage, sur leurs équivalents probabilistes, d'être plus adapté aux données imprécises. Toutefois, il faut remarquer qu'en raison de leurs domaines d'application (modèles de systèmes logiques, notamment), les RP sont utilisés la plupart du temps avec des variables discrètes, et qui plus est binaires ou symboliques.

1.3.2. Réseaux dynamiques

a. Généralités

Les RB sont adaptés au traitement de systèmes statiques, c'est-à-dire de systèmes dans lesquels une variable est observée une fois pour toutes, et ne changera plus de valeur. Cependant, ce n'est pas le cas pour un système évoluant rapidement, et pour lequel de nouvelles observations de la même variable sont disponibles régulièrement. Pour palier ce défaut, les réseaux bayésiens dynamiques (RBD) allient les principes décrits pour les RB à une représentation explicite du temps, dans le but de modéliser ces séries de données temporelles. Bien qu'il existe également des formalismes que l'on peut qualifier de RBD dans lesquels le temps est représenté de façon continue [Anderson & Moore '92], nous allons nous limiter à la description de ceux dans lesquels le temps est discrétisé, dans l'optique de déterminer le formalisme le plus adapté aux contraintes définies précédemment.

La différence fondamentale entre un RBD et un RB est que, dans un RBD, on associe une durée temporelle aux relations. C'est-à-dire que l'on considère qu'un nœud met un certain temps pour influencer un descendant. Les relations sont bien sûr orientées dans le sens chronologique, et on ne peut pas les retourner, ce qui rend ce type de réseau causal [Ghahramani '01]. En effet, on fait

l'hypothèse qu'un évènement peut en causer un autre qui lui est ultérieur, mais que le contraire n'est pas possible. De plus, à la différence des RB statiques, on ne manipule plus une v.a. X représentant un élément du système, mais une v.a. X_t représentant un élément du système *pris à un instant donné* t. Par conséquent, pour un modèle s'étendant sur n instants, tout élément du système sera représenté par n variables $X_1,\ldots,X_n$ dans le RBD. Pour faciliter leur représentation, les RBD sont, en général, décomposés en tranches temporelles, regroupant toutes les variables qui décrivent le système à un instant donné.

Reprenons l'exemple de RB illustré par la Figure III.1.4, en définissant cette fois la relation $X \to Y$ comme instantanée, ou durant moins d'un instant, et la relation $X \to Z$ comme durant un instant. La Figure III.1.5 est la représentation graphique de ce RBD :

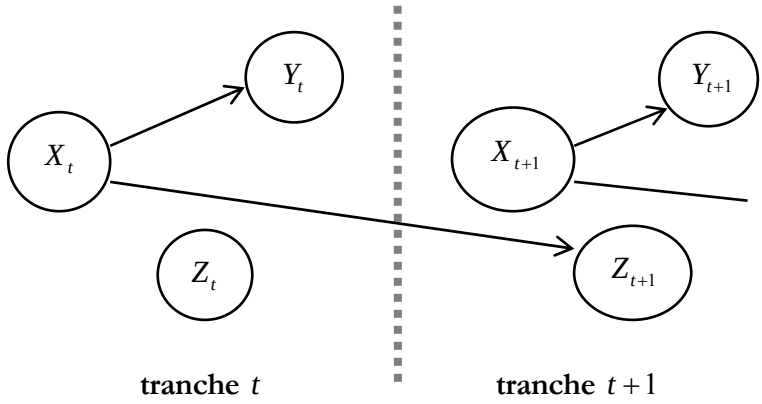

Figure III.1.5 : deux tranches temporelles d'un réseau bayésien dynamique simple.

Malgré l'introduction du temps, la décomposition de la distribution de probabilités conjointes utilisée dans les RB reste valable, et donc l'inférence dans un RBD est réalisée de la même façon que dans un RB. Avec le modèle qui nous sert d'exemple, si nous considérons n tranches temporelles, nous avons :

$$P\left(X_{1:n},Y_{1:n},Z_{1:n}\right) = P\left(X_1\right)P\left(Y_1\middle|X_1\right)P\left(Z_1\right) \times\prod_{i=2}^{n} P\left(X_i\right)P\left(Y_i\middle|X_i\right)P\left(Z_i\middle|X_{i-1}\right) \qquad \text{(eq. III.1.24)}$$

Cette distribution de probabilités conjointes respecte bien la formule énoncée en (eq. III.1.15).

Souvent, un élément d'un système dépend non seulement des états d'autres éléments, mais également de ses propres états passés. Formellement, cela revient à manipuler $P\left(X_t \middle| X_{1:t-1}, Y\right)$, où X_t est une v.a. dépendant de toutes ses valeurs passées, notées $X_{1:t-1}$ et d'un ensemble d'autres v.a. noté Y. On s'aperçoit que la complexité de traitement de ce type de réseau va rapidement devenir énorme, en raison du grand nombre de v.a., et donc de distributions de probabilités conditionnelles à calculer. Quant il est possible d'utiliser des v.a. numériques, la représentation fonctionnelle des relations permet de diminuer cette complexité. De plus, certains processus aléatoires, décrits ci-après, possèdent une organisation temporelle qui permet de simplifier leur traitement.

b. Processus de Markov

Si l'état futur d'un système ne dépend que de l'état du système à l'instant présent, on dit que le système suit un processus markovien. Soit une variable X modélisant un processus markovien. On a alors :

$$P\left(X_{t+1} \middle| X_{1:t}\right) = P\left(X_{t+1} \middle| X_t\right) \qquad \text{(eq. III.1.25)}$$

Autrement dit, le futur de X est indépendant de son passé conditionnellement à son présent. On dit que la séquence des X_t forme une chaîne de Markov (Figure III.1.6).

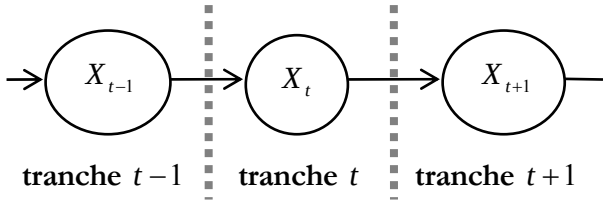

Figure III.1.6 : une chaîne de Markov d'ordre 1.

Il est possible d'étendre cette notion en définissant un processus markovien d'ordre n par le fait que le futur du système ne dépende que des n derniers états du système, c'est-à-dire :

$$P\left(X_{t+1} \middle| X_{1:t}\right) = P\left(X_{t+1} \middle| X_{t-(n-1):t}\right) \qquad \text{(eq. III.1.26)}$$

Cette propriété permet de grandement simplifier les calculs, puisque pour une chaîne de Markov d'ordre 1, on a [Ghahramani '98] :

$$P(X_{1:t}) = P(X_1)\prod_{k=2}^{t} P(X_k | X_{1:k-1}) = P(X_1)\prod_{k=2}^{t} P(X_k | X_{k-1}) \qquad \text{(eq. III.1.27)}$$

Le même concept peut être utilisé dans le cadre des réseaux possibilistes, et certains travaux récents décrivent des chaînes de Markov possibilistes [Dubois *et al.* '94; Janssen *et al.* '96]. Toutefois, ils sont dédiés à la modélisation de l'évolution de systèmes symboliques, et les données manipulées ne sont pas numériques.

c. Modèles de Markov cachés & modèles d'espaces d'états

Dans un grand nombre de systèmes, les éléments observables dépendent des éléments non-observables, et ces derniers suivent un processus de Markov. En représentant les éléments observables par un vecteur de v.a. noté Y et les éléments cachés par un vecteur de v.a. noté X, on obtient la représentation graphique donnée dans la Figure III.1.7.

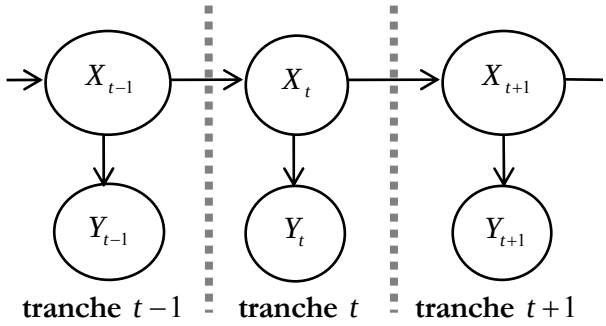

Figure III.1.7 : exemple de modèle de Markov caché/modèle d'espace d'états.

On appelle X la variable d'état et Y la variable observable. Si X est discrète, cette structure porte le nom de *modèle de Markov caché* (MMC), alors que si elle est continue, on parle de *modèle d'espace d'états* (MEE). Il existe de nombreuses variantes des MEE, la plus répandue étant les *modèles dynamiques linéaires gaussiens*, dans lesquels X et Y suivent toutes les deux des lois normales. On parle parfois de *filtre de Kalman*, du nom de l'algorithme utilisé pour l'inférence.

La distribution de probabilités conjointes d'un tel modèle (que ce soit un MMC ou un MEE) est donc de la forme [Ghahramani '98]:

$$P(X_{1:t}, Y_{1:t}) = P(X_1)P(Y_1 | X_1)\prod_{k=2}^{t} P(X_k | X_{k-1})P(Y_k | X_k) \qquad \text{(eq. III.1.28)}$$

Dans un MMC, les distributions de probabilités conditionnelles de transition $P\left(X_k \middle| X_{k-1}\right)$ et d'observation $P\left(Y_1 \middle| X_1\right)$ sont représentées par des matrices. Dans un MEE, elles sont représentées de façon fonctionnelle, à l'instar des RB continus :

$$X_k = f\left(X_{k-1}, W_{k-1}\right) \qquad \text{(eq. III.1.29)}$$

$$Y_k = h\left(X_k, V_k\right) \qquad \text{(eq. III.1.30)}$$

Ces distributions de probabilités conditionnelles définissent respectivement les mécanismes qui font la dynamique du système et les processus d'observation de l'état du système. Des algorithmes optimisés pour ce type de RBD permettent, en utilisant les même principes que pour les RB, d'y réaliser de l'inférence et de l'apprentissage [Roweis & Ghahramani '99] (c.f. paragraphe 1.3.1.a).

A noter que parfois, le système modélisé dispose d'une entrée qui permet en général de le commander. Dans ce cas, une variable U_k est rajoutée au modèle, il s'agit d'une variable déterministe qui influence X_k (Figure III.1.8).

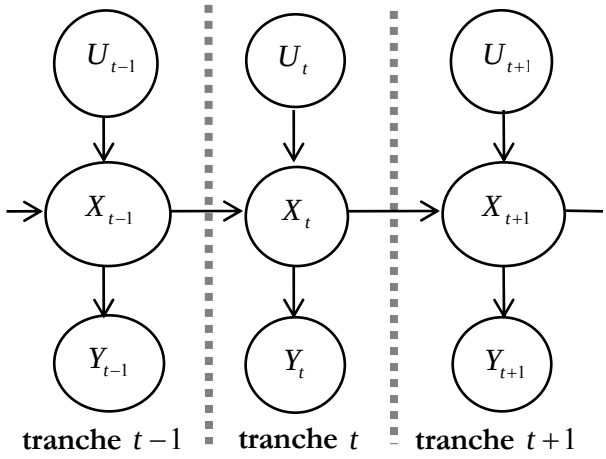

Figure III.1.8 : modèle de Markov caché/modèle d'espace d'états avec une entrée.

On trouve également des MMC possibilistes [Benferhat *et al.* '00]. Toutefois, là encore, ces modèles sont dédiés à la manipulation d'information symbolique, à travers l'étude de l'évolution de systèmes logiques.

1.3.3. Bilan

Parmi les réseaux probabilistes, ce sont les modèles d'espaces d'états qui se rapprochent le plus du formalisme idéal défini par nos contraintes. Ils ont la propriété de respecter la définition de la causalité donnée par Pearl (c.f. chapitre II.2.2), qui est compatible avec la nôtre (c.f. chapitre II.4.1). Les relations sont représentées sous la forme de fonctions, qui peuvent être linéaires ou non-linéaires. La contrainte temporelle est respectée dans les MEE, puisque le temps est explicitement représenté, sous une forme discrétisée.

Comme tous les autres formalismes présentés ici, les MEE ont une structure de réseau, sur laquelle pèsent certaines contraintes. Les variables non-observables doivent constituer une chaîne de Markov, et les variables observables doivent dépendre des ces variables d'états. Le cas de la modélisation cérébrale telle qu'elle est définie dans ce travail s'adapte très bien à ces contraintes. En effet, le cerveau est vu comme un système dynamique constitué d'un réseau d'aires cérébrales dont on ne peut pas observer le mécanisme directement (nos variables d'état). Par contre, il est possible d'obtenir des mesures indirectes de l'activité cérébrale (nos variables observables), qui dépendent du fonctionnement des réseaux d'aires cérébrales.

A ce niveau, on note d'importantes différences entre réseaux probabilistes et réseaux possibilistes, en ce qui concerne le domaine d'application. Si les réseaux probabilistes sont largement répandus, notamment pour la modélisation via des variables numériques, en revanche l'usage actuel des réseaux possibilistes est plutôt restreint à des modèles de nature qualitative, dans le sens où ils sont composés de variables discrètes. De plus, ce n'est que récemment que des modèles possibilistes dynamiques sont apparus. A ce stade, il est intéressant de remarquer que les réseaux qualitatifs propageant des intervalles flous à la place d'intervalles nets (c.f. paragraphe 1.2.4) se rapprochent beaucoup de ce que pourrait être un réseau possibiliste dynamique de variables continues.

Cette absence d'application dans le domaine numérique peut peut-être s'expliquer par la relative jeunesse de la théorie des possibilités (par rapport à celle des probabilités). En effet, les outils possibilistes sont beaucoup moins développés que leurs homologues probabilistes et statistiques, et plusieurs définitions s'affrontent parfois pour le même concept, comme par exemple pour les notions de moyenne et de variance d'une variable floue. Ceci rend plus complexe l'interprétation des variables floues numériques, et leur comparaison avec des données statistiques telles que les séries temporelles issues de la neuroimagerie, qui nous intéressent tout particulièrement. Or, ces mesures d'activation se présentent en général sous la forme couples moyennes/écart-types.

L'utilisation de réseaux probabilistes permet une meilleure comparaison des résultats obtenus par simulation avec les valeurs issues d'expériences réelles, puisque ce sont les mêmes outils mathématiques qui sont utilisés. L'évaluation qualitative des modèles s'en trouve facilitée.

Les modèles probabilistes statiques permettent d'inclure dans le même modèle des variables discrètes et des variables statiques. En ce qui concerne les modèles dynamiques basés sur des chaînes de Markov, on distingue le cas discret (MMC) du cas continu (MEE). Toutefois, on répertorie dans la littérature des modèles comptant plusieurs chaînes de v.a. d'états [Ghahramani & Jordan '97], qui interagissent, et d'autres modèles constitués de plusieurs MEE ou MMC [Wan & Nelson '96]. Il est donc possible d'utiliser une chaîne de v.a. continues ou un MEE pour représenter la partie quantitative de l'information cérébrale, et une chaîne de v.a. discrètes ou un MMC pour la partie qualitative, et ainsi de respecter la contrainte portant sur la représentation de l'information cérébrale à un niveau intégré.

De nombreux algorithmes existent pour inférer dans les MEE. Certains (e.g. filtre de Kalman [Kalman '60; Kalman & Bucy '61]) sont limités à des v.a. d'un type donné (des v.a. normales en général), représentées à travers leurs paramètres. D'autres propagent les distributions entières, sous la forme d'approximations non-paramétriques (Méthodes de Monte Carlo [Fahrmeir & Knorr-Held '00]). Des algorithmes d'apprentissage supervisé existent également, qui permettent de paramétrer le modèle. L'apprentissage tel qu'il est défini dans le cadre des réseaux probabilistes ne correspond pas à la notion décrite dans les contraintes, à savoir une adaptation du modèle en fonction des données qu'il reçoit en entrée. En effet, les algorithmes d'apprentissage des RB ont pour but de déterminer les paramètres du modèle en fonction de ses sorties, en procédant par comparaison avec des mesures effectuées sur le système réel.

1.4. Conclusion

Dans cette partie, nous avons décrit le principe de fonctionnement et les propriétés de chacun des formalismes utilisables dans le cadre de ce travail, et nous avons analysé leurs caractéristiques en considérant les contraintes précédemment définies. Comparativement, ce sont les réseaux probabilistes, et plus précisément les MEE qui se révèlent être les plus adaptés [Labatut '00; Labatut & Pastor '01].

Même s'ils possèdent de nombreux points communs avec les réseaux possibilistes, les réseaux probabilistes ont sur ces derniers l'avantage d'être beaucoup plus répandus dans le cadre de la modélisation de systèmes numériques et dynamiques. Ceci implique l'existence de nombreux algorithmes d'inférence et d'apprentissage. De plus, la théorie des probabilités semble beaucoup plus adaptée à nos objectifs, elle rend plus aisée la comparaison des données simulées avec les données réelles issues de la neuroimagerie (statistiques).

Le principal motif de rejet des réseaux qualitatifs est qu'ils présentent l'inconvénient de contraindre fortement la définition des relations entre les nœuds du modèle. Or, la contrainte portant sur les relations des MEE pèse avant tout sur la topologie du réseau et n'est pas limitative dans notre cas. De plus, à la différence des MEE, les réseaux qualitatifs ne permettent pas de gérer l'incertitude ou l'imprécision de façon explicite, et peuvent conduire à des résultats extrêmement imprécis, dans le cas (qui est le nôtre) de variables numériques et d'un modèle complètement non-observable.

En ce qui concerne les RNF, ils présentent l'avantage sur les MEE de disposer de mécanismes d'apprentissage plus proches de notre contrainte. De plus, si on considère la forme canonique d'un RNF dynamique, on s'aperçoit que l'aspect de ses équations ((eq. III.1.1) et (eq. III.1.2)) est très proche de celui des équations définissant un MEE ((eq. III.1.29) et (eq. III.1.30)), en considérant que celui-ci dispose d'une variable d'entrée (c.f. paragraphe 1.3.2.c). Toutefois, l'approche boîte noire qui caractérise ces RNF constitue un énorme handicap vis-à-vis des MEE. Ces derniers, au contraire, sont un outil de modélisation analytique, et sont, de ce fait, plus adaptés à l'objectif explicatif qui est le nôtre. De plus, les RNF ne remplissent pas la contrainte portant sur la possibilité de propager une information symbolique, et ne permettent pas de représenter explicitement l'imprécision ou l'incertitude. Or, ces deux contraintes sont respectées par les MEE.

Mais si les MEE sont le formalisme le plus adapté à nos critères de modélisation, ils ne respectent pas pour autant toutes nos contraintes. Les algorithmes d'apprentissage sont dédiés au paramétrage automatique des modèles, mais n'ont rien à voir avec l'apprentissage tel que nous l'avons défini (chapitres I.3.7 et II.4.2.6), et qui est propre au système que nous voulons modéliser. Un autre aspect caractéristique du cerveau porte sur l'information manipulée à un niveau intégré. Les notions d'apprentissage et de manipulation d'information sont ici intrinsèquement liées, et leur prise en compte dans notre formalisme va nécessiter la définition de mécanismes bien spécifiques. A ce titre, le formalisme ART, à base de RNF, présente pour nous un intérêt certain. En effet, il est d'inspiration biologique et inclut un mécanisme d'apprentissage adaptatif non-supervisé lui permettant de catégoriser de l'information.

2. ALGORITHMES POUR L'INFERENCE ET L'APPRENTISSAGE

Après avoir déterminé quel formalisme était le plus adapté à nos contraintes de modélisation, nous allons, dans la seconde partie de ce chapitre, nous intéresser à la façon de le faire fonctionner, c'est-à-dire aux aspects de propagation de l'information et d'apprentissage. Ceci implique de passer en revue les différents algorithmes d'inférence dans les MEE non-linéaires. Nous allons également présenter plus en détail l'algorithme d'apprentissage utilisé dans le formalisme ART.

2.1. Inférence dans les modèles d'espace d'états non-linéaires

2.1.1. Généralités

Dans les réseaux bayésiens en général, on appelle inférence le fait de calculer les distributions conditionnelles des variables constituant le modèle. Il s'agit de distributions *a posteriori*, c'est-à-dire que les algorithmes dédiés à cette tâche tiennent compte, dans le calcul, de la mesure d'une variable observable (ou de plusieurs variables). Dans un réseau bayésien dynamique, on distingue principalement deux types d'inférence : le filtrage, qui consiste à estimer l'état actuel du modèle en fonction des observations passées et présentes, et le lissage, qui consiste à estimer l'évolution du modèle *a posteriori*, c'est-à-dire que pour un instant donné, on tient compte des observations passées, présentes et futures [Fahrmeir '99; Murphy '98a]. Le premier type est adapté à un traitement en direct (on-line), dans lequel on alimente la simulation avec des données observées sur le système en temps réel [Forbes *et al.* '95; Kanazawa *et al.* '95; Kjaerulff '92, '95; Russell *et al.* '94]. Le terme de *filtre* vient du fait qu'historiquement, ce type d'algorithme était utilisé pour traiter un signal parasité par un bruit, et obtenir leur séparation sous la forme d'une somme [Grewal & Andrews '93]. Le second type s'applique forcément à une simulation en différé (off-line), c'est-à-dire pour laquelle on dispose déjà de toutes les observations. Le lissage se fait souvent en deux passes [Minka '99] : la première passe consiste à filtrer les états dans l'ordre chronologique, puis la deuxième passe permet de réviser les estimations à la lumière des observations qui leur sont postérieures [Arulampalam *et al.* '02]. Le lissage est fondamentalement non-causal [Hurd '96], puisqu'il tient compte des observations futures (c.f. l'importance du temps dans notre définition de la causalité, chapitre II.4.1), et peut engendrer des résultats biaisés [Wan & Nelson '96] en cas d'utilisation dans un système causal. Par conséquent, nous ne discuterons que des filtres.

Pour inférer dans des modèles d'espaces d'états linéaires, il existe un algorithme optimal appelé le filtre de Kalman, et dont le principe est décrit par la suite (paragraphe 2.1.2). Cet algorithme manipule exclusivement des v.a. de distributions normales. La combinaison linéaire de variables normales est également une variable normale, ce qui permet de garder des variables normales tout le long de la simulation. Mais dans notre cas, le système à modéliser est de nature non-linéaire. Le résultat de la combinaison non-linéaire de variables normales n'est pas forcément de distribution normale, ce qui rend généralement intraitable l'inférence exacte dans ce type de modèles. Pour résoudre ce problème de perte de normalité, deux grandes méthodes existent à ce jour. La première consiste à linéariser localement les relations, de façon à conserver des variables normales. C'est ce que font les variantes non-linéaires du filtre de Kalman. La seconde méthode, utilisée dans les algorithmes à échantillonnage, consiste à propager des approximations de ces distributions non-normales.

Ces méthodes à échantillonnage regroupent des algorithmes aux dénominations aussi diverses que *méthodes de Monte Carlo* [Fahrmeir & Knorr-Held '00], *filtres à particules*, algorithme *Condensation* (*Con*ditional *dens*ity propag*ation*) [Arulampalam *et al.* '02], *bootstrap filters* [Fahrmeir '98], etc. Les densités de probabilités y sont représentées de façon non-paramétrique, au moyen d'échantillons contenant suffisamment de points (appelés particules) pour les caractériser. Le principe de propagation se déroule en trois temps. Tout d'abord, l'échantillon est généré de manière séquentielle, par exemple en utilisant le méthode *MCMC* (*M*arkov *C*hain *M*onte *C*arlo) [Carter & Kohn '96]. Puis a lieu la phase de pondération, qui consiste à calculer un poids pour chaque particule, en tenant compte de l'équation d'évolution propre au modèle et des éventuelles observations. Enfin, éventuellement, les particules sont rééchantillonnées pour éviter une certaine déperdition d'information, appelée dégénérescence. Les algorithmes diffèrent essentiellement sur les méthodes employées pour générer les particules, et sur la façon dont la dégénérescence est traitée. A partir d'un de ces échantillons, il est possible d'estimer les valeurs caractéristiques telles que la médiane, la moyenne, le(s) mode(s) et autres. Grâce à cette propriété, ce type d'algorithmes offre l'avantage de pouvoir propager des distributions non-normales, et ne nécessite donc aucune hypothèse concernant les distributions des v.a. qui sont manipulées. Cette propriété présente également des défauts qui apparaissent au cours de la propagation des distributions. Le plus important est la dégénérescence des particules, qui correspond à une dégradation progressive de l'information propagée, et peut faire aboutir à des résultats erronés [van der Merwe & Wan '03a], d'où la nécessité de la phase de rééchantillonnage. De plus, ce type d'approche demande énormément de ressources lors de la simulation [Fahrmeir & Knorr-Held '00], en raison du grand

nombre de points nécessaires pour obtenir des approximations efficaces des distributions propagées : de quelques milliers pour une variable d'état de dimension 2 à plusieurs millions quand la dimension dépasse 5 [Julier & Uhlmann '02]. Pour ces raisons, bien que ces approches soient prometteuses, nous avons choisi de ne pas les retenir.

Dans la famille des versions non-linéaires du filtre de Kalman, on distingue principalement le filtre de Kalman étendu [Welch & Bishop '01], les filtres à sigma-points [van der Merwe & Wan '03a], encore appelés filtres de Kalman à régression linéaire [Lefebvre *et al.* '01] : filtre unscented [Julier & Uhlmann '97], filtre à différences centrales [Schei '97] et filtre à différences divisées [Nørgaard *et al.* '00] ; et enfin les filtres à mélanges ou à bascule [Murphy '98b]. Dans ces derniers, on utilise des combinaisons (le plus souvent, il s'agit de sommes) d'un très grand nombre de variables normales pour approximer les distributions non-gaussiennes résultant des relations non-linéaires. En pratique, cela revient à avoir plusieurs sous-modèles linéaires en parallèle et à construire une estimation globale de la variable d'état à partir de combinaisons linéaires des variables d'états des sous-modèles [Ghahramani & Hinton '98; Murphy '98b]. En fait, il s'agit d'une application particulière des modèles d'espaces d'états à bascule (switching state-space models), qui sont plus généralement utilisés pour modéliser des systèmes pouvant être décomposés en plusieurs modules agissant en parallèle (chaque module étant représenté par un sous-modèle). Une application en neurosciences [Ghahramani & Wolpert '97] est mentionnée dans le chapitre II.2.2. Le problème est l'explosion combinatoire engendrée : si on utilise n variables normales pour représenter une variable non-normale à l'instant t, et que chacune d'entre elle devient également non-normale à l'instant suivant, on se retrouve avec n^2 variables à l'instant $t+1$, etc. Différentes méthodes existent pour réduire cette augmentation exponentielle de la complexité, mais elles impliquent d'introduire plus d'erreur à chaque instant de la simulation, et peuvent aboutir à des résultats erronés. Pour ces raisons, seuls les filtres à sigma-points seront présentés plus en détail.

Le principe général du filtre de Kalman étendu et des filtres à sigma-points consiste à linéariser les équations non-linéaires qui constituent le modèle, et à appliquer un filtre de Kalman classique. Les bruits sont supposés suivre des distributions normales de moyenne nulle. Seule l'espérance et la covariance des v.a. sont propagées. Par rapport aux filtres à particules, on voit donc que la forme des distributions est contrainte. En contre-partie, cela permet une représentation paramétrique et un traitement plus aisé. Les versions diffèrent sur la façon dont la linéarisation est effectuée : développement de Taylor pour le filtre de Kalman étendu, qui est le plus ancien, et développement limité utilisant des interpolations à la place des dérivées, pour les trois autres algorithmes. Toutefois,

ils présentent tous la même complexité calculatoire en $O\left(n_x^3\right)$, n_x étant la dimension de la variable d'état [van der Merwe & Wan '01]. Le filtre de Kalman étendu offre, notamment au niveau de la précision, des performances se situant bien en dessous de celles des trois autres filtres [Nørgaard *et al.* '00; Welch & Bishop '01]. Au contraire, les performances de ceux-ci sont très proches car ils reposent sur les mêmes principes mathématiques [Lefebvre *et al.* '01]. En fait, le calcul de la moyenne est le même dans les trois algorithmes, et seule les méthodes de propagation des matrices de covariances diffèrent entre le filtre unscented et les filtres à différences finies [van der Merwe & Wan '01]. Les filtres à différence finies offrant une meilleure précision [Nørgaard *et al.* '00], nous allons maintenant présenter plus en détails le filtre à différences divisées.

2.1.2. Filtre de Kalman classique

Le filtre de Kalman, qui est initialement destiné aux systèmes linéaire, doit son nom à son inventeur, R.E. Kalman [Kalman '60; Kalman & Bucy '61]. Il s'agit d'un algorithme itératif fonctionnant par prédictions/corrections. Dans un modèle d'espaces d'états donné, il permet de réaliser une estimation au sens des moindres carrés de la variable d'état. Cet algorithme est optimal pour un système dans lequel les relations sont linéaires, ce qui signifie que la variance de l'estimation est minimale. Des versions et des preuves détaillées de l'algorithme présenté ici sont disponibles dans la littérature [Anderson & Moore '92; Grewal & Andrews '93; Harvey '89; McNamee '02; Welch & Bishop '01].

Soit le modèle linéaire défini par le couple d'équations suivant :

$$x_k = A_k x_{k-1} + B_k u_k + w_k \qquad \text{(eq. III.2.1)}$$

$$y_k = H_k x_k + v_k \qquad \text{(eq. III.2.2)}$$

Dans ces équations, $x_k \in \mathbb{R}^{n_x}$ est un vecteur décrivant l'état du système à l'instant k. On l'appelle variable d'état ou variable cachée, puisqu'on ne connaît pas sa valeur (c'est elle que l'on veut estimer). $u_k \in \mathbb{R}^{n_u}$ est un vecteur correspondant aux entrées du système à l'instant k, c'est la variable de contrôle. $y_k \in \mathbb{R}^{n_y}$ est un vecteur représentant ce que l'on peut observer ou mesurer indirectement du système à l'instant k. On l'appelle la variable observable.

A_k , B_k et H_k sont des matrices connues représentant respectivement le processus de transition, le processus de contrôle et le processus d'observation du système. Leurs dimensions respectives sont $(n_x \times n_x)$, $(n_x \times n_u)$ et $(n_y \times n_x)$.

$w_k \in \mathbb{R}^{n_x}$ et $v_k \in \mathbb{R}^{n_y}$ sont des vecteurs aléatoires représentant respectivement le bruit interne au modèle et le bruit venant altérer les observations. Ces bruits sont supposés indépendants les uns des autres à chaque instant, de moyenne nulle et suivant une distribution normale, avec des matrices de covariance respectives Q_k et R_k :

$$w_k \sim N(0, Q_k) \qquad \text{(eq. III.2.3)}$$

$$v_k \sim N(0, R_k) \qquad \text{(eq. III.2.4)}$$

Soient $\tilde{x}_k$ l'estimation de x_k *a priori* et $\hat{x}_k$ son estimation *a posteriori*. Les matrices de covariance des erreurs d'estimation *a priori* $\tilde{e}_k$ et *a posteriori* $\hat{e}_k$ sont respectivement :

$$\tilde{P}_k = E\left[\tilde{e}_k \tilde{e}_k^T\right] = E\left[\left(x_k - \tilde{x}_k\right)\left(x_k - \tilde{x}_k\right)^T\right] \qquad \text{(eq. III.2.5)}$$

$$\hat{P}_k = E\left[\hat{e}_k \hat{e}_k^T\right] = E\left[\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T\right] \qquad \text{(eq. III.2.6)}$$

où $E[\ ]$ dénote l'espérance mathématique d'une variable aléatoire et X^T la transposée d'une matrice X .

L'étape de *prédiction* consiste, en partant des estimations d'un état initial $\hat{x}_{k-1}$ et d'une matrice de covariance initiale $\hat{P}_{k-1}$, à utiliser l'équation (eq. III.2.1), appelée équation de mise à jour temporelle, pour calculer les estimations *a priori* de l'état et de la matrice de covariance à l'instant suivant k :

$$\tilde{x}_k = A_k \hat{x}_{k-1} + B_k u_k \qquad \text{(eq. III.2.7)}$$

$$\tilde{P}_k = A_k \hat{P}_{k-1} A_k^T + Q_k \qquad \text{(eq. III.2.8)}$$

On parle d'estimation *a priori* car on réalise une prédiction concernant l'état suivant, avant de connaître la mesure de la variable observée.

Dans l'étape de *correction*, on utilise cette mesure pour effectuer une révision des estimations et aboutir ainsi à des estimations *a posteriori* (i.e. après la mesure). Pour cela, on a besoin d'un coefficient appelé gain de Kalman. Celui-ci est calculé de façon à minimiser l'erreur d'estimation *a posteriori* $\hat{P}_k$:

$$K_k = \tilde{P}_k H_k^T \left(H_k \tilde{P}_k H_k^T + R_k \right)^{-1} \quad \text{(eq. III.2.9)}$$

On utilise ce gain pour pondérer l'erreur d'estimation de la variable mesurée. Cette erreur correspond à la différence entre la valeur effectivement mesurée y_k et son estimation réalisée en appliquant l'équation (eq. III.2.2) à l'estimation de l'état *a priori* $\tilde{x}_{k+1}$. Cette valeur pondérée par K_k permet de mettre à jour $\tilde{x}_k$ pour obtenir $\hat{x}_k$:

$$\hat{x}_k = \tilde{x}_k + K_k \left(y_k - H\tilde{x}_k \right) \quad \text{(eq. III.2.10)}$$

Et l'erreur devient :

$$\hat{P}_k = \left(I - K_k H_k \right) \tilde{P}_k \quad \text{(eq. III.2.11)}$$

où I dénote la matrice identité. Ces estimations *a posteriori* sont ensuite utilisées dans l'itération suivante (comme dans (eq. III.2.7) et (eq. III.2.8)).

2.1.3. Filtre de Kalman à différences divisées d'ordre 2

Cet algorithme est dû à M. Nørgaard, la preuve complète figure dans [Nørgaard *et al.* '00]. A la différence du filtre de Kalman classique, les équations du modèle peuvent être non-linéaires :

$$x_k = f\left(x_{k-1}, u_k, w_{k-1} \right) \quad \text{(eq. III.2.12)}$$

$$y_k = h\left(x_k, v_k \right) \quad \text{(eq. III.2.13)}$$

Les bruits sont des variables normales indépendantes w_k et v_k, de dimensions respectives n_w et n_v (au contraire du filtre de Kalman linéaire, leurs dimensions peuvent différer de celles de x_k et y_k, en raison de la nature non-linéaire des relations employées ici). Leur moyenne n'est pas forcément nulle :

$$w_k \sim N\left(\overline{w}_k, Q_k\right) \quad \text{(eq. III.2.14)}$$

$$v_k \sim N\left(\overline{v}_k, R_k\right) \quad \text{(eq. III.2.15)}$$

L'algorithme inclut une étape supplémentaire dédiée à la linéarisation des fonctions. Au lieu d'utiliser un développement de Taylor comme c'est le cas dans le filtre de Kalman étendu [Welch & Bishop '01], l'auteur utilise un développement limité basé sur des interpolations de Stirling. Cela permet d'éviter de manipuler des dérivées, et ainsi de simplifier à la fois l'implémentation et le traitement des modèles.

Pour un développement d'ordre 2 de $f\left(x\right)$ au voisinage de $\overline{x}$, x étant une variable vectorielle, la formule d'interpolation de Stirling (également appelée interpolation ST ou de Newton-Stirling) prend la forme suivante :

$$f\left(x\right) = f\left(\overline{x} + \Delta x\right) \approx f\left(\overline{x}\right) + \tilde{D}_{\Delta x} f + \frac{1}{2!}\tilde{D}_{\Delta x}^2 f \quad \text{(eq. III.2.16)}$$

où $\tilde{D}_{\Delta x}$ et $\tilde{D}_{\Delta x}^2$ sont des opérateurs définis par :

$$\tilde{D}_{\Delta x} f = \frac{1}{b}\left(\sum_{p=1}^{n} \Delta x_p \mu_p \delta_p\right) f\left(\overline{x}\right) \quad \text{(eq. III.2.17)}$$

$$\tilde{D}_{\Delta x}^2 f = \frac{1}{b^2}\left(\sum_{p=1}^{n} \left(\Delta x_p\right)^2 \delta_p^2 + \sum_{p=1}^{n}\sum_{q=1, q\neq p}^{n} \Delta x_p \Delta x_q \left(\mu_p \delta_p\right)\left(\mu_q \delta_q\right)\right) f\left(\overline{x}\right) \quad \text{(eq. III.2.18)}$$

et δ_p et μ_p, deux opérateurs basés sur les différences centrées :

$$\delta_p f\left(\overline{x}\right) = f\left(\overline{x} + \frac{b}{2} e_p\right) - f\left(\overline{x} - \frac{b}{2} e_p\right) \quad \text{(eq. III.2.19)}$$

$$\mu_p f(\bar{x}) = \frac{1}{2}\left(f\left(\bar{x} + \frac{b}{2} e_p\right) + f\left(\bar{x} - \frac{b}{2} e_p\right)\right) \quad \text{(eq. III.2.20)}$$

Dans ces équations, b désigne un intervalle fixé, qui est un paramètre de l'interpolation, et e_p est la $p^{ème}$ colonne de la matrice identité.

Pour des raisons calculatoires, les matrices de covariances des erreurs d'estimation font l'objet d'une décomposition en produits de facteurs de Choleski. La décomposition de Choleski d'une matrice symétrique définie positive M prend la forme suivante :

$$M = SS^T \quad \text{(eq. III.2.21)}$$

La décomposition de Choleski est l'équivalent matriciel d'une racine carrée. Le facteur S est une matrice triangulaire inférieure. Les matrices de covariance étant symétriques semi-définies positives, les décompositions de $\tilde{P}_k$, $\hat{P}_k$, Q_k et R_k donnent donc :

$$\tilde{P}_k = \tilde{S}_{x_k} \tilde{S}_{x_k}^T \quad \text{(eq. III.2.22)}$$

$$\hat{P}_k = \hat{S}_{x_k} \hat{S}_{x_k}^T \quad \text{(eq. III.2.23)}$$

$$Q_k = S_{w_k} S_{w_k}^T \quad \text{(eq. III.2.24)}$$

$$R_k = S_{v_k} S_{v_k}^T \quad \text{(eq. III.2.25)}$$

Les développements limités utilisés pour le calcul de ces matrices font eux aussi l'objet d'une telle décomposition. Les matrices correspondant au terme d'ordre 1 du développement limité sont ici notées $S_{x\tilde{x}_k}^{(1)}$, $S_{xw_k}^{(1)}$, $S_{y\tilde{x}_k}^{(1)}$ et $S_{yv_k}^{(1)}$. Celles qui figurent dans le terme d'ordre 2 sont notées $S_{x\tilde{x}_k}^{(2)}$, $S_{xw_k}^{(2)}$, $S_{y\tilde{x}_k}^{(2)}$ et $S_{yv_k}^{(2)}$.

L'estimation de l'état *a priori* est effectuée en appliquant (eq. III.2.16) à (eq. III.2.12) et en développant les opérateurs définis précédemment :

$$\begin{aligned}\tilde{x}_k &= \frac{b^2 - n_x - n_w}{b^2} f\left(\hat{x}_{k-1}, u_k, \overline{w}_{k-1}\right) \\ &+ \frac{1}{2b^2}\sum_{p=1}^{n_x} f\left(\hat{x}_k + b\hat{s}_{x-1,p}, u_k, \overline{w}_{k-1}\right) + f\left(\hat{x}_{k-1} - b\hat{s}_{x,p}, u_k, \overline{w}_{k-1}\right) \\ &+ \frac{1}{2b^2}\sum_{p=1}^{n_w} f\left(\hat{x}_{k-1}, u_k, \overline{w}_{k-1} + bs_{w,p}\right) + f\left(\hat{x}_{k-1}, u_k, \overline{w}_{k-1} - bs_{w,p}\right)\end{aligned} \qquad \text{(eq. III.2.26)}$$

Ici, $\hat{s}_{x,p}$ et $s_{w,p}$ représentent respectivement les $p^{ème}$ colonnes de $\hat{S}_{x_k}$ et de S_{w_k}.

L'erreur d'estimation *a priori* prend la forme d'une matrice composée rectangulaire et non-triangulaire :

$$\tilde{S}_{x_k} = \begin{bmatrix} S^{(1)}_{x\hat{x}_{k-1}} & S^{(1)}_{xw_{k-1}} & S^{(2)}_{x\hat{x}_{k-1}} & S^{(2)}_{xw_{k-1}} \end{bmatrix} \qquad \text{(eq. III.2.27)}$$

Elle est rendue carrée et triangulaire en appliquant la méthode de Householder [Cohn '02], afin de pouvoir être utilisée dans la suite des calculs.

Il est nécessaire de calculer l'estimation de la variable observée et de sa covariance pour obtenir le gain de Kalman. On procède comme pour l'estimation *a priori* de la variable d'état, c'est-à-dire qu'on applique (eq. III.2.16) à (eq. III.2.13) pour calculer $\tilde{y}_k$, et que la covariance est obtenue grâce à des facteurs de Cholesky :

$$\begin{aligned}\tilde{y}_{k+1} &= \frac{h^2 - n_x - n_v}{h^2} g\left(\tilde{x}_k, \overline{v}_k\right) \\ &+ \frac{1}{2h^2}\sum_{p=1}^{n_x} g\left(\tilde{x}_k + h\tilde{s}_{x,p}, \overline{v}_k\right) + g\left(\tilde{x}_k - h\tilde{s}_{x,p}, \overline{v}_k\right) \\ &+ \frac{1}{2h^2}\sum_{p=1}^{n_v} g\left(\tilde{x}_k, \overline{v}_k + hs_{v,p}\right) + g\left(\tilde{x}_k, \overline{v}_k - hs_{v,p}\right)\end{aligned} \qquad \text{(eq. III.2.28)}$$

où n_v est la taille du vecteur de bruit de mesure et $\tilde{s}_{x,p}$ et $s_{v,p}$ représentent respectivement les $p^{ème}$ colonnes de $\tilde{S}_{x_k}$ et de S_{v_k}.

Là encore, l'erreur d'estimation est représentée par une matrice composée rectangulaire, qui doit être rendue carrée afin d'être transformée en un facteur de Cholesky utilisable pour la suite du traitement :

$$\tilde{S}_{y_k} = \begin{bmatrix} S^{(1)}_{y\tilde{x}_k} & S^{(1)}_{xv_k} & S^{(2)}_{y\tilde{x}_k} & S^{(2)}_{yv_k} \end{bmatrix} \quad \text{(eq. III.2.29)}$$

On peut alors calculer le gain de Kalman avec la formule suivante :

$$K_k = \tilde{S}_{x_k} \left(S^{(1)}_{y\tilde{x}_k} \right)^T \left(S_{y_k} S^T_{y_k} \right)^{-1} \quad \text{(eq. III.2.30)}$$

Ce gain permet de calculer l'estimation *a posteriori* de la variable d'état :

$$\hat{x}_k = \tilde{x}_k + K_k \left(y_k - \overline{y}_k \right) \quad \text{(eq. III.2.31)}$$

Et l'erreur d'estimation *a posteriori*, qui nécessite également une transformation pour devenir un facteur de Choleski :

$$\hat{S}_{x_k} = \begin{bmatrix} \tilde{S}_{x_k} - K_k S^{(1)}_{y\tilde{x}_k} & K_k S^{(1)}_{yv_k} & K_k S^{(2)}_{y\tilde{x}_k} & K_k S^{(2)}_{yv_k} \end{bmatrix} \quad \text{(eq. III.2.32)}$$

2.2. Apprentissage dans les réseaux ART

Nous avons vu (paragraphe 1.1.3.c) que les réseaux *ART* réalisaient une classification non-supervisée et adaptative, en implémentant un système de résonance entre entrée et sortie du réseau, utilisant une mémoire à court terme (MCT) et une mémoire à long terme (MLT). A noter qu'il existe de nombreuses extensions de ART, en plus de la version initiale, ART1, qui accepte uniquement des entrées binaires. On trouve, entre autres, ART2 [Carpenter & Grossberg '87b] et ART3 [Carpenter & Grossberg '90], qui acceptent des entrées réelles, Fuzzy-Art [Carpenter *et al.* '91b], qui est une extension de ART1 acceptant des valeurs réelles comprises entre 0 et 1, et diverses versions supervisées : (ARTMAP [Carpenter *et al.* '91a], fuzzy-ARTMAP [Carpenter *et al.* '92] , etc…).

Le système est composé de deux couches notées *F1* et *F2* (Figure III.2.1), et appelées respectivement *couche de comparaison* et *couche de reconnaissance*. Les neurones composant les couches peuvent être binaires (ART1) ou réels (ART2). Dans le cadre de cette description, nous allons considérer que l'information manipulée est binaire (ART1). La MCT correspond à la configuration des neurones qui émettent (dans F1 et dans F2). Le nombre de neurones de F1 correspond au

nombre de caractéristiques définissant l'information traitée par le réseau. La couche de comparaison a la particularité de posséder deux modes de fonctionnement : saisie ou comparaison.

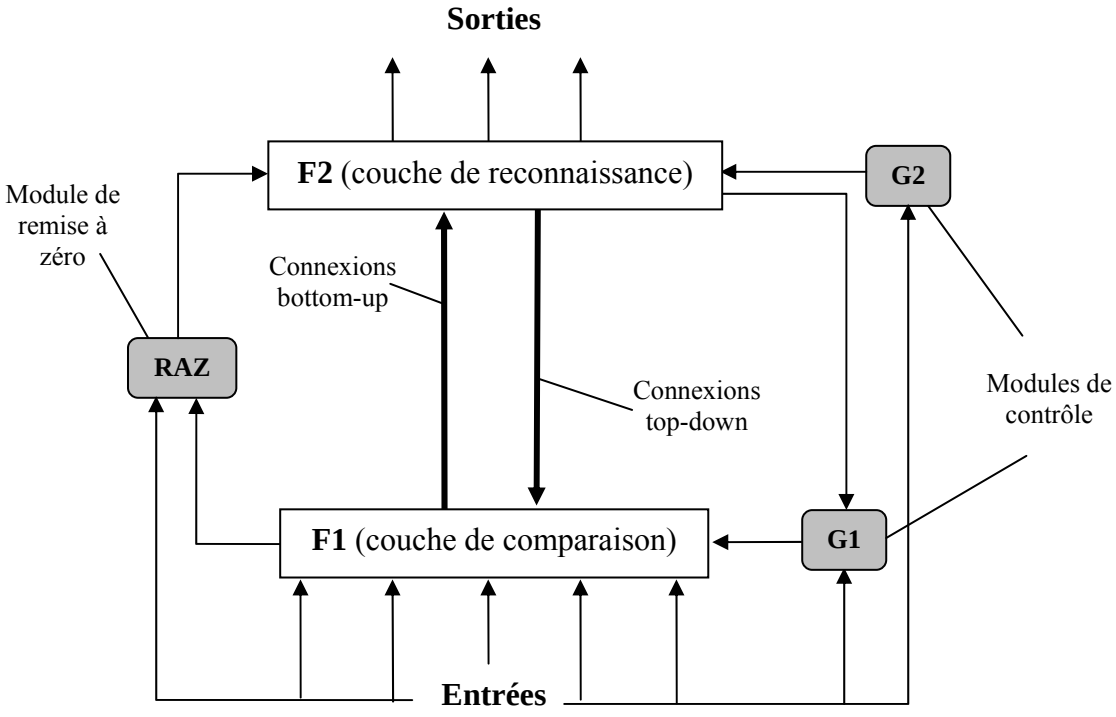

Figure III.2.1 : structure d'un modèle ART1.

Chaque neurone de la couche de sortie, appelé *attracteur*, correspond à une classe : le réseau peut donc différencier autant de classes qu'il possède de neurones dans la couche F2. L'allocation de neurones est réalisée de façon dynamique, car on ne sait pas *a priori* combien de classes vont être nécessaires. Chaque neurone de F1 est relié à tous les neurones de F2 (connexions *bottom-up*), et inversement (connexions *top-down*). Les poids qui définissent ces connexions représentent la MLT. En fait, ces poids définissent une sorte « d'entrée idéale » pour la classe représentée par l'attracteur : pour cet idéal, la réponse de l'attracteur est maximale. De plus, chaque neurone de F2 est relié à tous autres neurones de F2 par des connexions inhibitrices et à lui-même par une connexion excitatrice. On a là un réseau à compétition, ou de type winner-takes-all : pour une classe donnée, un seul neurone émet.

Deux modules de contrôle *G1* et *G2* viennent compléter le modèle. Chacun d'eux est associé à une couche (respectivement F1 et F2), en étant connecté à chacun des neurones qui la composent. Ces contrôleurs sont chargés d'inhiber le fonctionnement des couches en l'absence de signal d'entrée. Ils sont influencés par les entrées. G1 a en plus la charge de commuter le mode de fonctionnement

de la couche de comparaison, et il est influencé par F2 (en plus des entrées). Enfin, un dernier module nommé RAZ (remise à zéro) contrôle la boucle de résonance. Il est connecté en entrée aux entrées du réseau et à la couche de comparaison. En sortie, il influence la couche de reconnaissance.

Considérons le fonctionnement des différents éléments. G1 tente de faire basculer F1 en mode comparaison si un des attracteurs est actif ou de l'inhiber s'il n'y a aucune entrée disponible. En mode saisie, la couche de comparaison F1 est un simple tampon (les entrées sont recopiées telles quelles). Dans le mode comparaison, chaque neurone de F1 réagit en fonction de l'attracteur actif. Les poids correspondant aux connexions top-down entre l'attracteur actif et F1 forment un vecteur binaire représentant l'entrée idéale pour cet attracteur. Un poids égal à 1 dans ce vecteur signifie que la caractéristique correspondante dans l'information en entrée doit également être égale à 1 pour que cet attracteur reconnaisse l'information. En résumé, en mode comparaison, chaque neurone de F1 compare une caractéristique de l'information en entrée avec la caractéristique correspondante de l'idéal propre à l'attracteur actif, et le neurone ne prend la valeur 1 que si la caractéristique d'entrée et la caractéristique idéale sont toutes les deux égales à 1.

Le module de contrôle G2 inhibe F2 quand il n'y a pas d'information en entrée du réseau. Dans ce cas-là, même si F1 émet en direction de F2, cette dernière ne réagit pas. En ce qui concerne le module RAZ, il est inactif lors de la phase de saisie. En fait, les connexions entre l'information en entrée et RAZ sont excitatrices, alors que celles entre F1 et RAZ sont inhibitrices. Or lors de la saisie, F1 est égale aux entrées, donc les deux influences s'annulent. En revanche, lors de la phase de comparaison (résonance), l'état de F1 est modifié (par F2) et devient différent des entrées. Si cet état devient trop dissemblable à l'information en entrée, RAZ devient actif, et inhibe l'attracteur actif (ce n'est pas un bon candidat). Dans le cas où aucun attracteur n'est suffisamment adapté, une nouvelle classe est créée pour l'information, soit en affectant la classe à un attracteur vacant, soit en créant un nouvel attracteur. Un seuil d'attention permet de définir le niveau de similitude en dessous duquel l'attracteur actif doit être inhibé. Si, au contraire, les deux valeurs sont suffisamment semblables, alors la phase de comparaison est interrompue et l'attracteur actif constitue le résultat final. Indirectement, le seuil d'attention détermine le nombre de classes reconnues par le réseau : plus il est faible, plus le nombre de classes sera important.

En phase de comparaison, chaque neurone de la couche F2 est excité par la couche F1 entière (s'il n'est pas inhibé par RAZ). Son activation dépend donc de la correspondance entre les poids de ses connexions d'entrée et l'information que lui envoie F1. L'attracteur le plus activé inhibe les autres et

s'auto-renforce, il émet et devient l'attracteur actif. L'apprentissage consiste alors à transformer la MCT (état de F1) en MLT (poids des connexions). Pour cela, les poids des connexions top-down de l'attracteur actif prennent pour valeur l'état de F1, alors que celles des connexions bottom-up sont calculées en fonction de ce même état. Puis, le traitement entame un nouveau cycle, en calculant les nouvelles valeurs de F1. C'est ce mécanisme en forme de boucle entre F1 et F2 qui est qualifié de résonance.

Dans le cas de ART2, le principe général est le même, mais la structure du réseau est légèrement différente : la couche F1 a une structure plus complexe, destinée à effectuer un traitement sur les entrées (diminution du bruit, contraste sur certaines configurations d'activation, etc..). De plus, la définition de la distance entre F1 et l'entrée est différente de celle de ART1. Dans ce dernier, on utilise le rapport de la somme des neurones de F1 sur la somme des entrées (ou parfois la différence entre les entrées et F1). Dans ART2, on utilise le cosinus de l'angle formé par les deux vecteurs que constituent les entrées et F1 [Levine '00].

Chapitre IV

CONCEPTS ET DEFINITIONS

L'approche de ce travail est guidée par des objectifs de modélisation et la prise en compte de certaines caractéristiques du système cérébral, qui sont résumés par un ensemble de contraintes décrit dans le chapitre I.3. Nous avons vu dans le chapitre II qu'aucun formalisme dédié à la modélisation cérébrale, que ce soit dans le champ des neurosciences computationnelles ou dans celui de la neuroimagerie, ne respecte complètement ces contraintes. Après avoir passé en revue un certain nombre de formalismes causaux dans le chapitre III, il s'est avéré que les réseaux bayésiens dynamiques, et plus précisément les modèles d'espaces d'états non-linéaires, semblaient constituer le moyen le plus adapté d'atteindre nos objectifs de modélisation. A partir des notions fondamentales et des contraintes du chapitre I, nous proposons un ensemble de concepts et de définitions destinés à servir de base à la définition de notre formalisme final.

1. REPRESENTATION EN RESEAU

La modélisation d'un RCGE passe d'abord par la construction, à partir de données anatomiques, d'un réseau *structurel* qui représente le RCGE. Ce réseau est par la suite modifié pour exprimer certaines informations ou hypothèses fonctionnelles. Le réseau obtenu est appelé réseau *fonctionnel*. La simplicité de cette forme permet d'utiliser les propriétés de décomposition de certains nœuds du réseau, et de profiter du concept de modèle générique, autorisant la réutilisation d'éléments d'un modèle dans un autre. Une fois la spécification fonctionnelle complètement effectuée, on obtient le réseau *statique*. La simulation constitue la deuxième étape. Elle repose sur une version *dynamique* du réseau, obtenue en instanciant dans le temps les nœuds du réseau statique.

1.1. Réseau statique

1.1.1. Réseau structurel

Notre objectif est de modéliser un réseau cérébral à grande échelle (RCGE), c'est-à-dire un réseau de populations neuronales. Les premières informations disponibles sont d'ordre anatomique et concernent essentiellement l'identification des aires concernées et de leurs liaisons. Afin d'exprimer ces données, ou de spécifier des hypothèses d'ordre structurel, nous définissons les concepts de nœud et de réseau *structurels* :

> **Nœud structurel** : un nœud structurel représente une population neuronale fonctionnellement homogène et définie anatomiquement de façon relativement précise (identifiable et localisable) [Lafon '00].

> **Réseau structurel** : Un réseau structurel décrit un réseau d'aires cérébrales. Chaque nœud de ce réseau est un nœud structurel [Lafon '00], et chaque lien est orienté et représente une liaison anatomique.

Dans le cadre d'un réseau structurel, on ne s'intéresse donc pas à la structure anatomique propre d'une population neuronale (sa taille, par exemple), mais aux connexions qui l'unissent aux autres populations du réseau. La Figure IV.1.1 présente un exemple de réseau structurel.

Selon cette définition, l'échelle spatiale est donc très variable pour un nœud structurel : une aire de Brodmann constitue un bon exemple de nœud structurel, mais il peut également s'agir de structures plus importantes comme des gyri, ou bien de taille plus réduite, comme des colonnes corticales. L'échelle utilisée dans un modèle dépendra de la nature de la fonction étudiée, et des données anatomiques permettant de construire le modèle.

Les arcs représentent les liaisons anatomiques orientées (paquets d'axones) qui connectent ces populations neuronales et transmettent l'information entre elles. On ne fait pas ici de distinction entre les différents types de liaisons (distales ou locales), dépendant de la longueur spatiale du lien anatomique. Il n'est pas rare de trouver, entre deux régions, des faisceaux contenant à la fois des fibres orientées dans un sens, et d'autres fibres orientées dans l'autre sens. On représente dans ce cas-là deux relations différentes (chacune orientée dans un sens), créant ainsi un cycle orienté entre deux populations cérébrales. C'est par exemple le cas entre le gyrus fusiforme et cortex cingulaire

postérieur, dans la Figure IV.1.1. De manière plus générale, on peut trouver dans le réseau statique des cycles orientés impliquant plus de deux nœuds : on n'a donc pas ici un réseau causal.

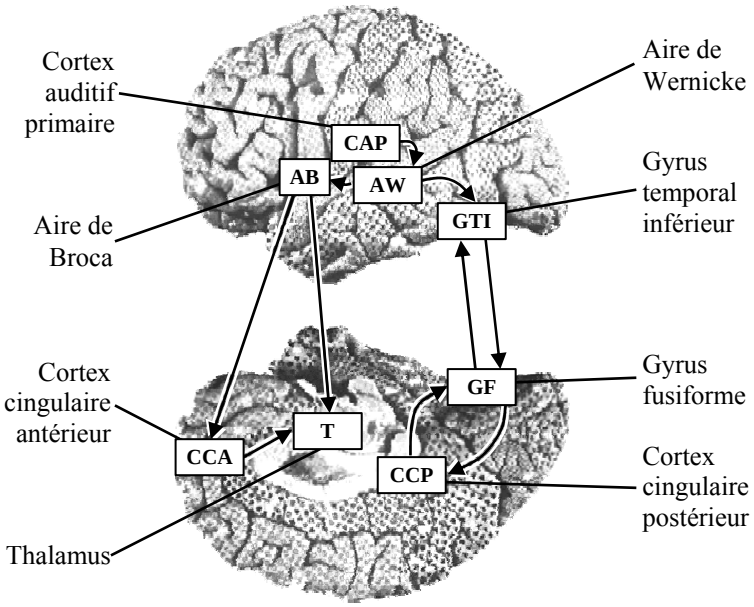

Figure IV.1.1 : exemple de réseau structurel.
Le modèle représente les aires impliquées dans un processus de détection phonémique, d'après [Démonet *et al.* '94].

Parfois, les données anatomiques permettent de décomposer une population neuronale en plusieurs sous-populations (c.f. chapitre I.1.2.1), comme par exemple la décomposition somatotopique du cortex moteur primaire. Au niveau du modèle, cela se traduit par la possibilité de décomposer un nœud structurel en un réseau de nœuds structurels. Chaque nœud de ce réseau représente une petite population neuronale, géographiquement contenue dans la population initiale.

1.1.2. Décomposition fonctionnelle

Après l'aspect physique, la seconde catégorie d'information décrivant les populations neuronales concerne leurs aspects fonctionnels. Afin de les exprimer dans nos modèles, nous introduisons le concept de nœud *fonctionnel*, qui étend la définition de Marc Lafon [Lafon '00] :

> **Nœud fonctionnel** : un nœud fonctionnel représente la fonction implémentée par une population neuronale. Cette dernière n'est pas forcément identifiée ou localisée, mais doit néanmoins être fonctionnellement cohérente, et topographiquement délimitée par l'ensemble des connexions qui l'unissent avec les autres populations.

La définition de nœuds fonctionnels relève de l'exploitation de données physiologiques et d'hypothèses d'ordre fonctionnel. Les seules données nécessaires pour modéliser le rôle d'une population neuronale sont ses caractéristiques fonctionnelles et la nature de ses connexions fonctionnelles avec les autres populations neuronales. En effet, ces liens ont également des propriétés fonctionnelles, qui peuvent être déduites de certaines caractéristiques anatomiques. Par exemple, le temps de propagation dans un lien fonctionnel dépend de la longueur, du type et de la myélinisation des fibres le composent.

Le concept de nœud fonctionnel permet d'introduire la notion de décomposition hiérarchique du réseau structurel. En effet, n'importe quel nœud structurel peut théoriquement être représenté par un nœud fonctionnel, ou par plusieurs d'entre eux, organisés en réseau :

> **Réseau fonctionnel** : Un réseau fonctionnel décrit le traitement de l'information implémenté par un nœud structurel, ou fonctionnel de niveau hiérarchique plus élevé. Il n'y apparaît aucune décomposition anatomique.

Ceci permet d'exprimer de façon plus explicite des hypothèses sur le fonctionnement d'une population neuronale. Cette décomposition en un sous-réseau fonctionnel peut se faire même lorsqu'il n'est pas possible de localiser anatomiquement des sous-populations de façon précise. C'est par exemple le cas des neurones GABAergiques (c.f. chapitre I.1.1.3.b), qui implémentent une fonction d'inhibition. La Figure IV.1.2, d'après [Lafon *et al.* '98], représente une décomposition purement fonctionnelle du nœud structurel représentant le cortex cingulaire postérieur dans la Figure IV.1.1.

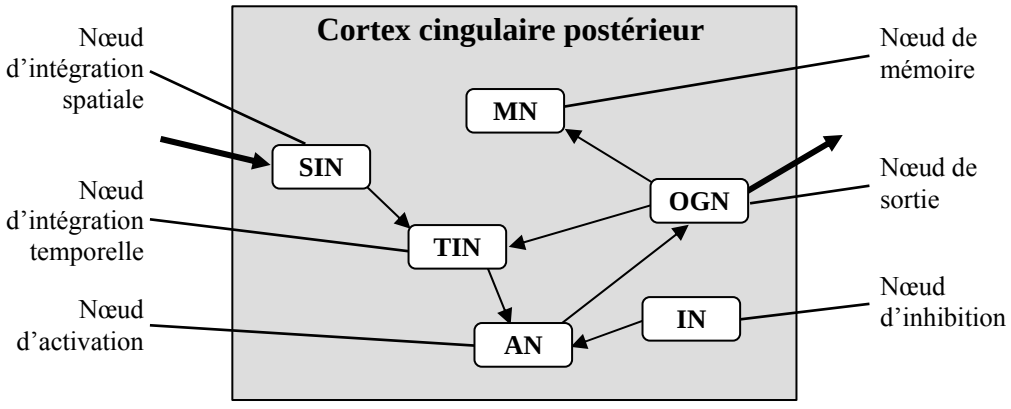

Figure IV.1.2 : décomposition fonctionnelle du nœud structurel représentant le cortex cingulaire postérieur de la Figure IV.1.1.

Un nœud fonctionnel peut lui-même être décomposé en un réseau de nœuds fonctionnels. Au plus bas niveau de décomposition, on dit d'un nœud fonctionnel qu'il implémente une *primitive fonctionnelle*. Ces mécanismes de décomposition hiérarchique permettent d'adapter un modèle à de nouvelles connaissances ou hypothèses, concernant le fonctionnement ou la structure d'une population en particulier. Par exemple, un nœud fonctionnel implémentant une primitive fonctionnelle peut être remplacé par un réseau de nœuds fonctionnels, si on décide de se placer à un niveau plus bas de représentation fonctionnelle (la fonction implémentée par le nœud de départ perdant alors son statut de primitive). A l'inverse, plusieurs primitives fonctionnelles peuvent être réunies en un seul nœud fonctionnel afin de simplifier un modèle.

Finalement, chaque nœud du réseau structurel initial peut être décomposé en un réseau fonctionnel. Lorsque le réseau ne contient plus que des nœuds fonctionnels, on parle de réseau *statique* :

> **Réseau statique** : Un réseau statique décrit le traitement de l'information dans un réseau cérébral à grande échelle. Il apparaît lorsque tous les nœuds d'un réseau structurel ont été décomposés en réseaux fonctionnels.

1.1.3. Modèle générique

Des études sur le plasticité cérébrale (c.f. chapitre I.1.3) ont mis en évidence l'existence de caractéristiques fonctionnelles communes à plusieurs populations neuronales différentes (par exemple : l'organisation topique des cortex primaires). On peut donc réunir ces aires dans des classes fonctionnelles. Considérons la décomposition de ces populations neuronales sous la forme de réseaux fonctionnels. Une de nos hypothèses fondamentales est qu'une caractéristique fonctionnelle commune à plusieurs aires se traduit par des propriétés communes au niveau des réseaux fonctionnels représentant ces aires. Ces propriétés peuvent concerner la nature des nœuds fonctionnels utilisés, ou la façon dont ces nœuds sont interconnectés.

Le concept de modèle générique d'une aire cérébrale permet de tirer parti de cette hypothèse. Un *modèle générique* est un réseau fonctionnel partiellement défini, dans le sens où certains nœuds ou relations sont manquants ou incomplètement définis. On distingue plusieurs niveaux de généricité suivant le degré de spécification. Les seuls éléments définis dans le modèle générique sont ceux caractéristiques de la classe fonctionnelle correspondante, c'est-à-dire ceux communs à toutes les aires qui composent la classe. Chaque aire de la classe correspond à une instanciation différente du modèle générique.

Par exemple, considérons le réseau fonctionnel décrit par la Figure IV.1.3. Il correspond à un modèle générique d'aire corticale [Lafon *et al.* '97]. La Figure IV.1.4.a montre une première instanciation de ce modèle, utilisée pour représenter le cortex visuel [Pastor *et al.* '00]. La Figure IV.1.4.b montre une seconde instanciation, différente de la première, utilisée pour modéliser une aire auditive associative [Labatut *et al.* '03a]. Dans la première instanciation, le nœud de mémoire MN du modèle générique est remplacé par un réseau de trois nœuds, formant un cycle orienté si on inclut l'OGN du cortex. Ce réseau comprend trois nœuds fonctionnels : l'entrée et la sortie d'une structure thalamique, et un nœud modélisant un seuil d'activation.

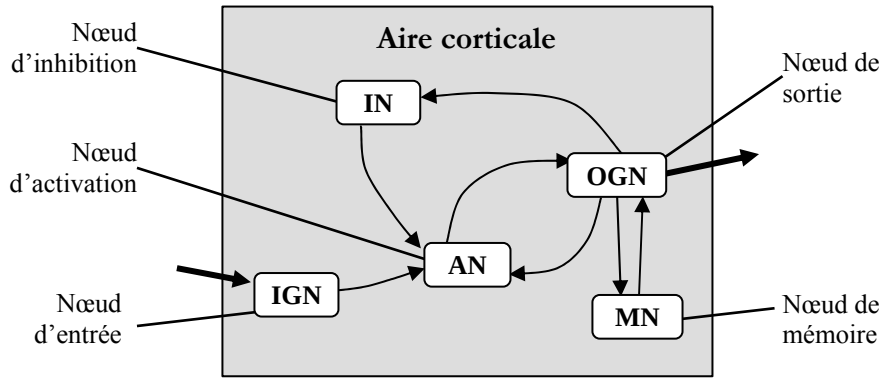

Figure IV.1.3 : exemple de modèle générique d'une aire corticale (adapté de la représentation graphique propre à BioCaEn) [Lafon *et al.* '97].

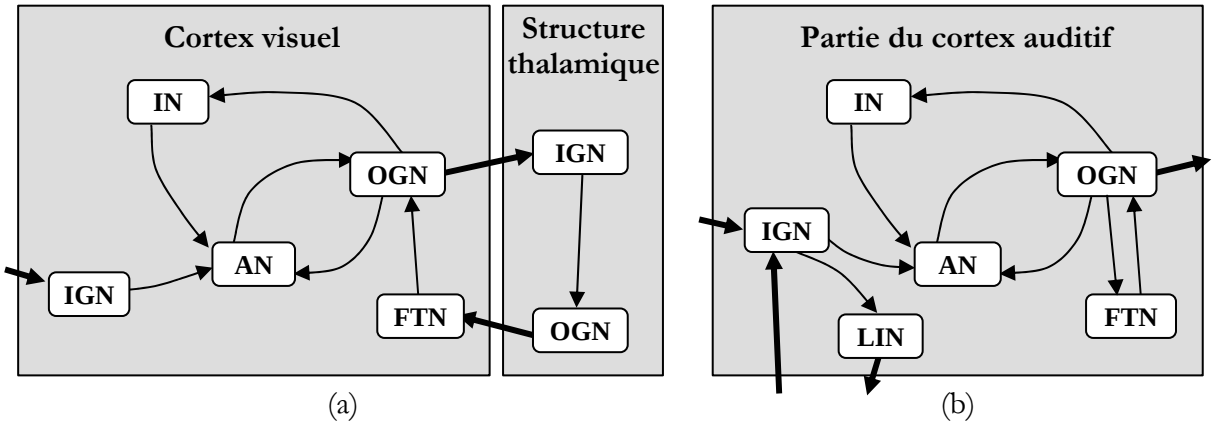

Figure IV.1.4 : deux instanciations du modèle générique décrit dans la Figure IV.1.3.
(a) : modèle de cortex visuel (adapté de la représentation graphique propre à BioCaEn) [Pastor *et al.* '00]. (b) : modèle d'aire auditive associative [Labatut *et al.* '03a].

Dans la seconde instanciation, MN est transformé en un nœud de seuil FTN (sa fonction est modifiée). On remarque aussi l'apparition d'un nouveau nœud LIN, qui vient compléter le modèle générique afin de l'adapter à ce cas précis d'utilisation. Il faut savoir que le réseau présenté dans la

Figure IV.1.4.b ne représente qu'une partie du modèle de cortex auditif. Le réseau global se décompose en fait en plusieurs sous réseaux, dont chacun est construit sur le modèle générique Figure IV.1.3, et prend la forme décrite en Figure IV.1.4.b (ce modèle est présenté au chapitre VI.2).

La notion de généricité peut être étendue au niveau structurel. En effet, on retrouve des organisations anatomiques similaires dans des RCGE différents, ne serait-ce que par symétrie entre les deux hémisphères. Un modèle générique structurel consiste à définir partiellement un réseau faisant intervenir des nœuds structurels.

1.2. Réseau dynamique

Le réseau statique décrit les interconnexions d'un ensemble de populations neuronales, représentées par des nœuds fonctionnels. Ces connexions sont les abstractions de liaisons axonales, caractérisées par des propriétés physiques, et notamment par leur longueur. Le temps que met l'information pour être transmise d'une aire à l'autre dépend de ces caractéristiques. Il faut souligner qu'en l'absence de pathologie, ce délai est constant, il n'évolue pas au cours du temps. La prise en compte du temps dans la représentation graphique du RCGE modélisé permet de dérouler le réseau statique, et ainsi d'en briser les éventuels cycles orientés.

> **Réseau dynamique** : Un réseau dynamique est l'instanciation temporelle d'un réseau statique. Il ne contient pas de cycle orienté.

Dans le réseau dynamique, on s'intéresse non plus à la nature de la population modélisée, mais à son comportement. Le temps est représenté sous forme discrète et régulière, comme spécifié dans la contrainte temporelle (c.f. chapitre II.4.2.2). Cela signifie que le comportement de chaque population neuronale modélisée est représenté par une succession d'états séparés par une durée constante tout le long de la simulation Figure IV.1.5. Un nœud du réseau dynamique correspond donc à un nœud du réseau statique, pris à un instant donné.

De même, un arc du réseau dynamique ne correspond pas à une liaison anatomique, mais représente fonctionnellement ce lien. On parle ici de connectivité causale [Pastor *et al.* '00], car il s'agit d'un lien causal (selon la définition du chapitre II.4.1).

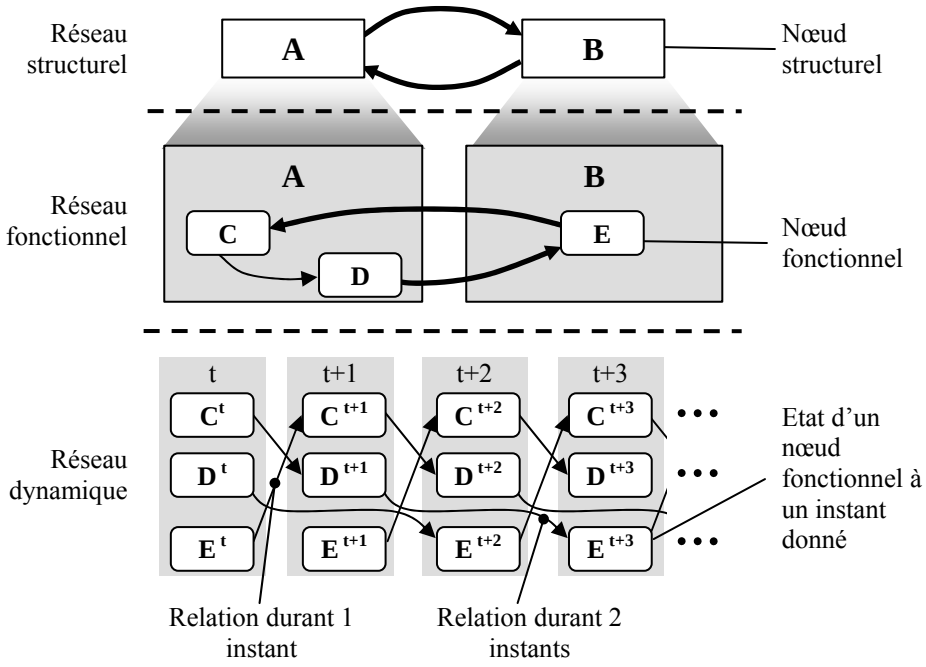

Figure IV.1.5 : résumé du processus menant d'un réseau structurel à un réseau dynamique.

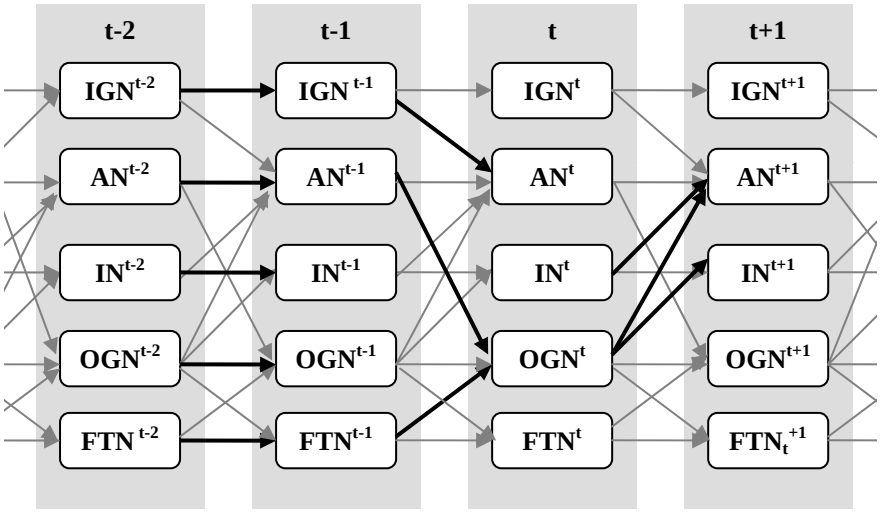

Figure IV.1.6 : exemple de réseau dynamique correspondant au réseau fonctionnel de la Figure IV.1.4.a.
On suppose ici que tous les délais de transmission associés aux relations sont de 1 instant.

La durée associée à chacune des relations se traduit, dans la représentation graphique, par des arcs reliant des nœuds situés à des instants différents. C'est cette propriété en particulier qui permet de dérouler le réseau statique. La Figure IV.1.6 est l'équivalent dynamique du réseau statique de la Figure IV.1.4.a. Le cycle entre les nœuds AN et OGN, observé dans le réseau statique, disparaît dans le réseau dynamique. A noter les connexions impliquant deux nœuds représentant la même

population à deux instants données. Ces liens modélisent le fait qu'en général, l'état d'une population neuronale à un instant donné dépend, entre autres, de son état à l'instant précédent.

L'état d'une population neuronale à un instant donné est décrit, d'une part, par les caractéristiques fonctionnelles de la population à ce moment-là, et, d'autre part, par l'information qui transite dans ce nœud à cet instant. La suite de la description des fondements de notre formalisme passe donc par la présentation de l'information cérébrale et des mécanismes qui permettent de la traiter et de la propager.

2. REPRESENTATION DE L'INFORMATION CEREBRALE

Afin de clarifier nos propos, nous utiliserons maintenant la terminologie suivante pour désigner l'information cérébrale suivant que l'on se place d'un point de vue externe ou interne au modèle :

Stimulus : le stimulus est l'information externe, c'est-à-dire celle qui provient de l'extérieur du système cérébral, et va être traitée dans nos modèles.

Information interne : par opposition au stimulus, l'information interne est celle qui est manipulée à l'intérieur des modèles, et propagée de nœud en nœud.

Le terme d'information (cérébrale) désignera indifféremment le stimulus et l'information interne. Conformément à la contrainte concernant la représentation de l'information dans les RCGE, l'information manipulée dans notre formalisme est définie par deux composantes, appelées le type et la magnitude.

Magnitude : La magnitude est une valeur numérique qui représente l'intensité de l'information transmise.

Type : le type est une valeur symbolique qui représente la sémantique de l'information transmise.

Par conséquent, la description de l'information transitant dans une population neuronale passe par deux valeurs, ce qui correspond à la décomposition décrite par la Figure IV.2.1.

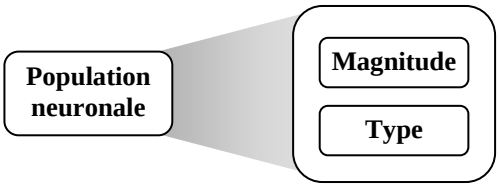

Figure IV.2.1 : décomposition correspondant à la description de l'information transitant par une population de neurones.

2.1. Magnitude

L'information traitée dans le modèle peut être interprétée de deux façons différentes suivant les nœuds considérés. On distingue, d'une part, l'information à l'instant où elle est traitée dans un nœud représentant le comportement d'une population neuronale, ou bien à l'instant où elle transite entre deux nœuds de ce type, et d'autre part l'information au niveau d'un nœud exogène, c'est-à-dire représentant un stimulus externe au système cérébral.

Dans le cas d'une population de neurones, la magnitude reflète l'intensité de son activation globale. En d'autres termes, la magnitude reflète à la fois le nombre de neurones activés dans la population, et leurs niveaux d'activation. Cette interprétation se répercute sur la transmission de l'information : à ce niveau, la magnitude représente à la fois le nombre d'axones activés correspondant aux neurones qui déchargent parmi ceux de l'aire qui émet, et les intensités moyennes (i.e. en tenant compte des fréquences de décharge individuelles) des influx qui transitent dans ces axones.

Quand on se place en dehors du modèle, donc au niveau du stimulus, la magnitude correspond à l'énergie du stimulus, qui se calcule en fonction des caractéristiques physiques du stimulus. Par exemple, pour un stimulus auditif, la magnitude peut correspondre au volume sonore.

2.2. Type

2.2.1. Interprétation externe

La notion de type repose sur les concepts de symbole et de champ catégoriel.

Symbole : d'un point de vue externe, le terme de symbole désigne une information pure, c'est-à-dire qui n'est pas parasitée par un bruit ou par d'autres informations.

Par exemple lorsqu'on manipule une information de nature linguistique, un symbole peut désigner un phonème non-ambigu. Si le stimulus est une note de musique, un symbole peut correspondre à un intervalle de fréquences sonores suffisamment réduit.

> **Champ catégoriel** : d'un point de vue externe, un champ catégoriel est un ensemble de symboles représentant, pour un stimulus donné, une dimension psychophysique particulière.

Par conséquent, tous les symboles doivent se placer sur le même plan sémantique, et être caractérisés par la même granularité. L'expression « même plan sémantique » signifie que les symboles doivent décrire un seul aspect du stimulus. Il est par exemple possible de définir un champ catégoriel dans lequel chaque symbole correspond à un phonème. Mais il n'est pas possible d'inclure dans ce champ-là un symbole représentant une couleur. L'expression « même granularité », veut dire que les symboles doivent décrire des informations de la même complexité. Par exemple, on ne peut pas mélanger des symboles représentant des phonèmes et d'autres correspondant à des syllabes au sein d'un même champ catégoriel. Le but de ces deux contraintes est d'avoir un ensemble de symboles cohérent et homogène. On peut les résumer de façon plus intuitive, en disant que les symboles d'un champ catégoriel doivent être en concurrence pour décrire un aspect du stimulus. On a donc une organisation hiérarchique particulière, basée sur la nature de la dimension psychophysique associée au champ catégoriel. Par exemple, le champ catégoriel des syllabes ne contient pas celui des phonèmes (d'un point de vue ensembliste), mais il lui est néanmoins supérieur en termes de granularité.

En pratique, l'information va être composée de plusieurs symboles, à cause de la présence de bruit, ou simplement parce qu'elle peut être la combinaison de plusieurs informations basiques. Le *type*, qui est défini sur un *domaine de définition* précis, permet de représenter cette combinaison :

> **Type** : d'un point de vue externe, un type décrit une répartition de symboles définissant une information complexe.

> **Domaine de définition** : le domaine de définition d'un type est l'ensemble des symboles dont le type décrit la répartition.

A ce niveau (externe au modèle), le domaine de définition est un champ catégoriel, et on parle de *domaine simple* et de *type simple*.

Domaine simple : domaine de définition correspondant à un champ catégoriel.

Type simple : type défini sur un domaine simple.

Ici, le type décrit donc la répartition des différents symboles composant le champ catégoriel. Par exemple, supposons que le stimulus soit de nature linguistique, un mixage de deux syllabes différentes. Le champ catégoriel le plus adapté est sans doute l'ensemble des symboles représentant chacun une syllabe. Le type décrivant le stimulus sera une distribution des symboles de ce champ catégoriel, privilégiant largement les deux symboles correspondant aux deux syllabes utilisées dans le mixage. A l'instar de la magnitude, on peut calculer le type en fonction des caractéristiques psychophysiques du stimulus d'entrée.

Lorsque l'information contenue dans un stimulus est complexe, dans le sens où elle s'étend sur plusieurs dimensions psychophysiques (par exemple : la forme et la couleur d'un objet), on utilise plusieurs types pour la décrire, chacun possédant un domaine de définition correspondant à une de ces différentes dimensions.

2.2.2. Interprétation interne

A l'intérieur du modèle, l'information est émise par une population neuronale, ce qui modifie la signification du concept de symbole.

Symbole : d'un point de vue interne, un symbole correspond à une sous-population de neurones ayant la propriété d'être sensibles à l'information représentée par ce symbole, dans son interprétation externe.

Reprenons l'exemple d'un symbole représentant un intervalle de fréquences. Si on se place dans le cortex auditif primaire, le symbole désigne la sous-population sensible à (i.e. déchargeant pour) cet intervalle de fréquence. Cette interprétation est consistante avec l'organisation topique des cortex primaires, qui est partiellement conservée au long des circuits corticaux et dont on trouve des réminiscences dans certaines aires associatives (c.f. chapitre I.1.2).

L'interprétation de la notion de champ catégoriel est également modifiée. L'information étant émise par une population neuronale, nous supposons qu'un champ catégoriel est associé à la sortie de chacune d'entre elle. Ce champ catégoriel dépend en fait du processus de catégorisation implémenté par la population, et nous supposons qu'il est unique pour une population donnée (i.e. on n'associe

qu'un seul champ à une population). Nous faisons également l'hypothèse qu'il existe différents champs catégoriels à divers niveaux d'abstraction dans un même réseau cérébral à grande échelle. Le traitement de l'information effectué par les populations neuronales permet de la faire passer successivement d'un champ catégoriel à un autre, et d'implémenter ainsi, par exemple, un mécanisme de raffinement de l'information.

> **Champ catégoriel** : d'un point de vue interne, un champ catégoriel est un ensemble de symboles représentant, pour une information émise par une population neuronale donnée, une dimension sémantique particulière.

Le champ catégoriel regroupe ici, sous forme de symboles, les différentes sensibilités que peuvent potentiellement posséder les sous-populations de neurones qui constituent la couche émettrice de l'aire cérébrale. De façon plus intuitive, le champ catégoriel répertorie les informations que l'aire peut potentiellement émettre. Il faut souligner que les symboles qui définissent ce champ catégoriel ne sont pas forcément tous connus a priori. En effet, les phénomènes d'apprentissage peuvent conduire une population neuronale à émettre des symboles jusqu'alors ignorés (au niveau de cette aire), et appartenant néanmoins au champ catégoriel.

> **Type** : d'un point de vue interne, un type décrit une répartition de symboles représentant la configuration des neurones qui ont déchargé parmi ceux de la population émettrice de l'information.

En raison des hypothèses émises plus haut, les types émis par une aire sont tous simples, c'est-à-dire définis sur un seul et unique champ catégoriel. En temps normal, ce champ catégoriel est invariant, mais des mécanismes de réorganisation fonctionnelle peuvent en modifier la nature. Toutefois, dans le cadre de notre travail, nous nous restreignons pour l'instant à des domaines de définitions invariants. La répartition des symboles décrite par le type traduit le comportement des différentes sous-populations (chacune correspondant à un symbole).

Si les types transitant d'un nœud à l'autre sont toujours des types simples, en revanche certains traitements internes à une population neuronale (détaillés plus loin) peuvent nécessiter de manipuler une information hétérogène.

> **Domaine multiple** : domaine de définition correspondant à l'association de plusieurs champs catégoriels.

> **Type multiple** : type défini sur un domaine multiple.

Les types *multiples*, permettent de regrouper des informations décrivant des aspects différents d'une information, en manipulant plusieurs champs catégoriels en même temps. Cette représentation de l'information n'est que transitoire, elle est toujours associée à un mécanisme de traitement qui la transforme en un type simple destiné à être propagé.

2.3. Complémentarité des deux composantes

Les deux composantes sont extrêmement différentes, et elles ne dépendent pas l'une de l'autre. Elles sont toutefois très complémentaires. Complémentaires tout d'abord dans la description de l'information cérébrale, puisque considérées séparément, la magnitude et le type ne permettent pas de caractériser cette information. Par exemple, il est possible que deux informations possèdent un niveau global d'activation similaire, et soient donc représentés par la même magnitude. Mais cette même activation peut correspondre à de nombreuses configurations possibles dans la population émettrice, et donc à de nombreux types différents, ce qui permet de différencier les deux informations (Figure IV.2.2).

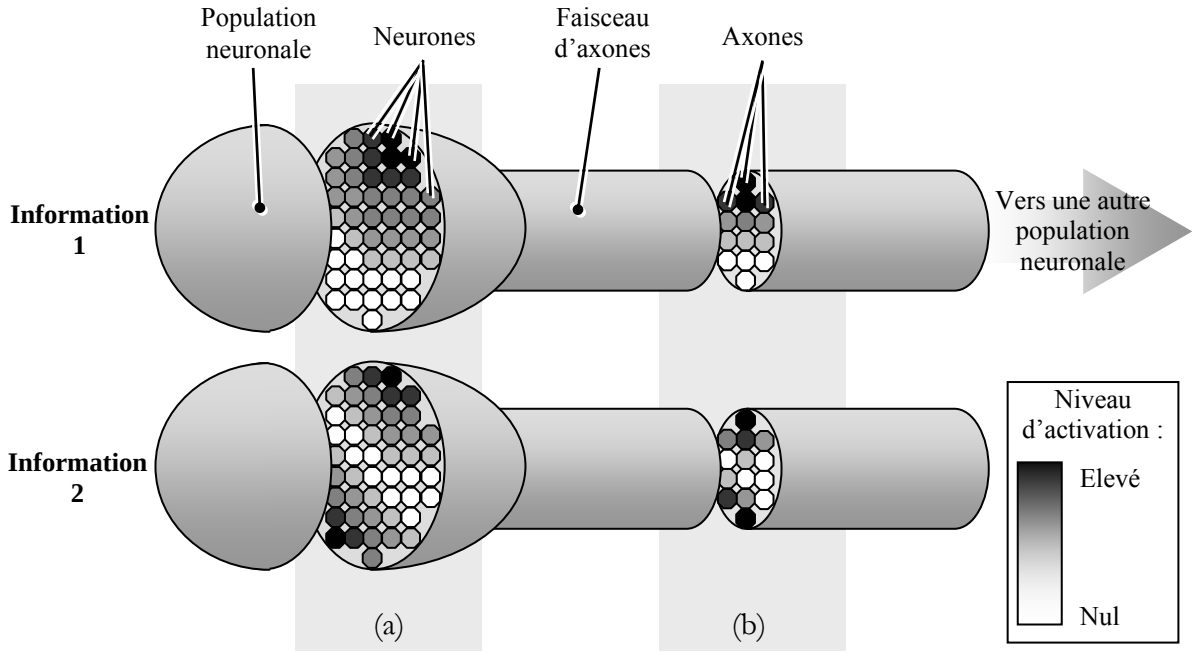

Figure IV.2.2 : illustration des concepts de magnitude et de type.
(a) : au niveau d'une population neuronale. (b) : au niveau d'un faisceau d'axones. Les deux informations possèdent la même magnitude, mais les types sont différents.

A l'inverse, les deux informations peuvent correspondre à la même configuration d'activation, et donc être décrits par le même type. Mais il est possible que les niveaux d'activation soient différents, ce qui implique deux magnitudes différentes (Figure IV.2.3).

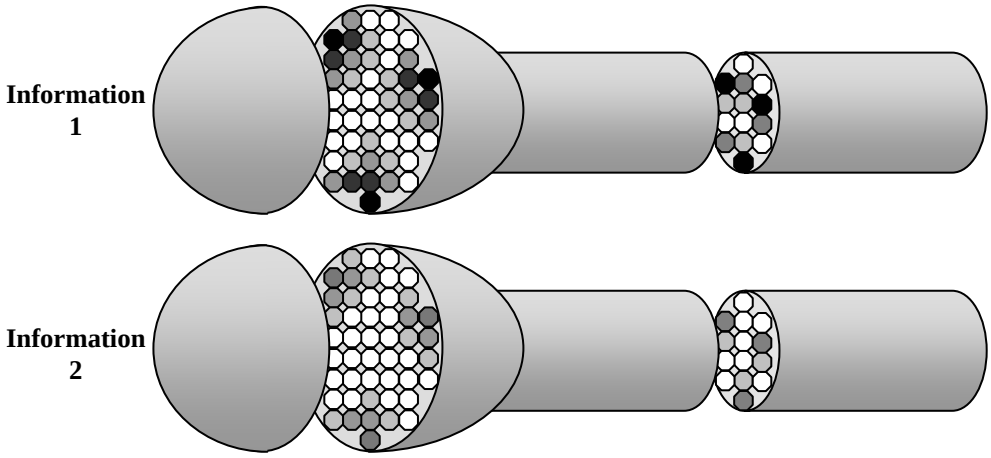

Figure IV.2.3 : illustration des concepts de magnitude et de type.
Cette fois, les deux informations possèdent le même type, mais leurs magnitudes sont différentes.

La complémentarité est également présente au niveau de l'utilisation des deux composantes : la magnitude va s'avérer utile du point de vue de l'interprétation de données de neuroimagerie, alors que le type servira à étudier le traitement cognitif de l'information cérébrale. En effet, la magnitude est assez proche des mesures issues des techniques d'imagerie cérébrale. Elle ne représente pas directement une variation de débit sanguin ou de champ électromagnétique, cependant elle peut être interprétée comme reflétant une grandeur biologique qui influence ce type de mesures. Le type, lui, ne constitue pas une valeur mesurable par neuroimagerie. On n'a donc aucun moyen de l'observer dans le système réel. Il est néanmoins essentiel à la modélisation et au processus de simulation, car la nature de l'information qui transite dans les RCGE détermine en partie le traitement qui est effectué. En plus de son rôle dans la propagation de l'information, le type va également intervenir dans les mécanismes d'apprentissage, décrits plus loin. Enfin, les types permettent de donner une certaine transparence sémantique [Wallace *et al.* '93] à notre formalisme, propriété qui est absente ou, du moins, présente à un niveau très faible, dans les RNF utilisés en modélisation cérébrale.

3. TRAITEMENT ET PROPAGATION DE L'INFORMATION CEREBRALE

3.1. Définitions

Lorsqu'on se place au niveau cellulaire, le traitement et la propagation de l'information cérébrale sont réalisés grâce aux mécanismes d'activation et de décharge que l'on trouve chez tous les neurones. Un neurone soumis à un ensemble de stimuli va intégrer (au niveau du soma) les diverses influences (potentiels post-synaptiques) d'un point de vue aussi bien spatial que temporel, et obtenir ainsi un potentiel d'activation. Puis, si ce potentiel dépasse un certain seuil d'émission, il est traité (au niveau du cône d'emboîtement) pour obtenir le potentiel d'action, qui va être transmis aux neurones situés en aval (c.f. chapitre I.1.1.2).

Il est plus difficile de définir des mécanismes généraux d'intégration des entrées, de traitement et de propagation de l'information au niveau de la population cérébrale. Néanmoins, en s'appuyant sur les connaissances et hypothèses décrites dans le chapitre I, il est possible d'obtenir une décomposition utilisable dans le cadre de ce travail de modélisation.

Lorsque l'on considère une liaison entre deux populations neuronales A et B, les connexions anatomiques ne concernent qu'une partie de chaque populations : tous les neurones de la population A ne sont pas forcément connectés à tous ceux de la population B. Quand A transmet une information à B, on peut donc considérer que seule une partie des neurones qui composent A émet l'information, et que seule une sous-population de B la reçoit. Par la suite, l'information est propagée et traitée au sein de B (à la façon d'un RNF). Il résulte de ce traitement et de cette propagation une certaine activité neuronale. Puis, en fonction du résultat de ce traitement, une sous-population de B émettra (ou pas) une réponse vers d'autres populations neuronales qui lui sont anatomiquement liées.

Cette description est très schématique, la réalité est plus complexe. En effet, une population neuronale reçoit en général des informations provenant de plusieurs sources, ce qui implique que la couche d'entrée de B réalise une intégration spatiale de l'information. De plus, le neurone étant un objet dynamique, le comportement de la population de neurones dépend également de son état passé. En d'autre terme, la population réalise aussi une intégration temporelle de l'information.

L'état actuel des connaissances ne permet pas de se prononcer quant au mécanisme d'intégration des entrées. Il existe deux possibilités : soit la région cérébrale intègre dans un premier temps ses entrées, effectue un traitement sur le résultat, pour finalement émettre ; soit, au contraire, elle effectue un traitement sur chaque entrée séparément avant de les combiner et éventuellement de décharger. La première option est identique au mode de fonctionnement d'un neurone formel, alors que la seconde se rapproche au principe utilisé dans BioCaEn (chapitre III.1.2). Mais la deuxième hypothèse revient à décomposer la région en plusieurs nœuds distincts appliquant la première hypothèse, et à utiliser un dernier nœud pour intégrer le résultat de leur traitement. Nous avons donc choisi de ne considérer que le premier des deux mécanismes.

En définitive, le fonctionnement d'une population neuronale peut être partagé en deux étapes : l'activation et l'émission.

Activation : le terme d'activation de l'aire cérébrale désigne l'activité issue de l'intégration des entrées de l'aire et du traitement interne à l'aire.

Emission : on parle d'émission quand une couche de sortie de l'aire émet une information vers d'autres populations neuronales.

Cette décomposition du traitement est représentée par la Figure IV.3.1. Si une population cérébrale, à la suite de la réception d'une information, émet une réponse pouvant être considérée comme une validation sémantique de cette entrée, alors on dit qu'elle a *reconnu* l'information.

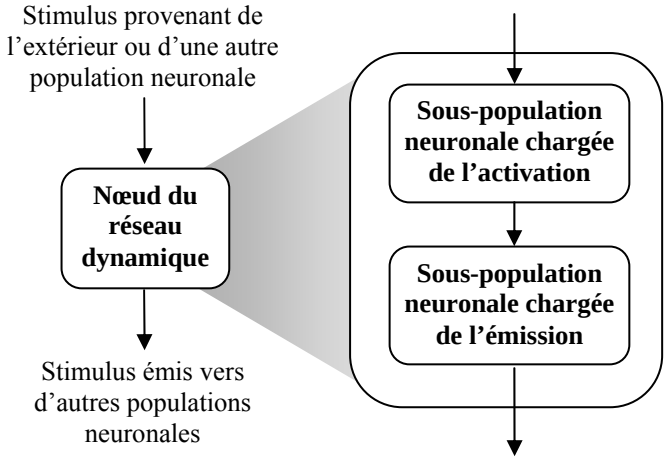

Figure IV.3.1 : décomposition d'un nœud du réseau dynamique due au traitement de l'information cérébrale.

Cette décomposition est biologiquement plausible, et offre en outre un intérêt certain en ce qui concerne la comparaison de données simulées avec les données issues de la neuroimagerie. En effet, ces dernières reflètent plus le niveau d'activation d'une zone que son niveau d'émission. Les aires localisées par neuroimagerie ne sont pas forcément des aires ayant répondu, il peut s'agir également d'aires ayant reçu une information qui a déclenché une forte activité interne, mais aucune réponse.

3.2. Description des mécanismes

Nous avons choisi de décomposer le processus de traitement et de propagation de l'information cérébrale en deux phases distinctes : l'activation et l'émission. Ces phases, implémentées par les populations de neurones, prennent la forme de fonctions. On distingue des *fonctions d'activation*, correspondant à la première phase, et des *fonctions d'émission* pour la seconde. Ces fonctions sont utilisées pour calculer l'information résidant dans la population neuronale ou émise par elle.

L'information manipulée par une population neuronale étant modélisée par un couple type/magnitude, il est nécessaire d'associer, pour chaque phase, deux fonctions à une population neuronale : une pour traiter les types, et une autre pour les magnitudes. On obtient donc au total, pour décrire une même population neuronale à un instant donné : une magnitude et un type d'activation, et une magnitude et un type d'émission, soit quatre données représentant l'information (Figure IV.3.2).

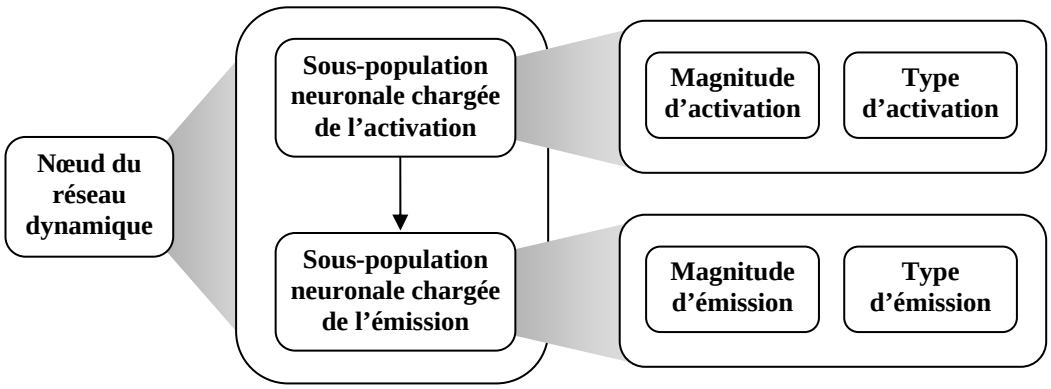

Figure IV.3.2 : l'information décrivant l'état d'un nœud dynamique est constituée de quatre valeurs : magnitude et type d'activation, et magnitude et type d'émission.

3.2.1. Activation

Le principal rôle de la fonction d'activation est de réaliser l'intégration spatiotemporelle des entrées du nœud. A ce titre, la fonction prend pour arguments les valeurs émises par les nœuds parents du nœud concerné afin de pouvoir réaliser l'intégration spatiale. L'intégration temporelle est réalisée en incluant les valeurs d'activation antérieures du nœud considéré (Figure IV.3.3). Nous faisons l'hypothèse qu'il n'y a pas d'interaction entre les deux composantes de l'information à ce niveau du traitement. Par conséquent, la fonction dédiée au calcul du type d'activation ne prend en compte que la partie type de l'information provenant de ces parents, et de même pour la fonction dédiée au calcul de la magnitude d'activation.

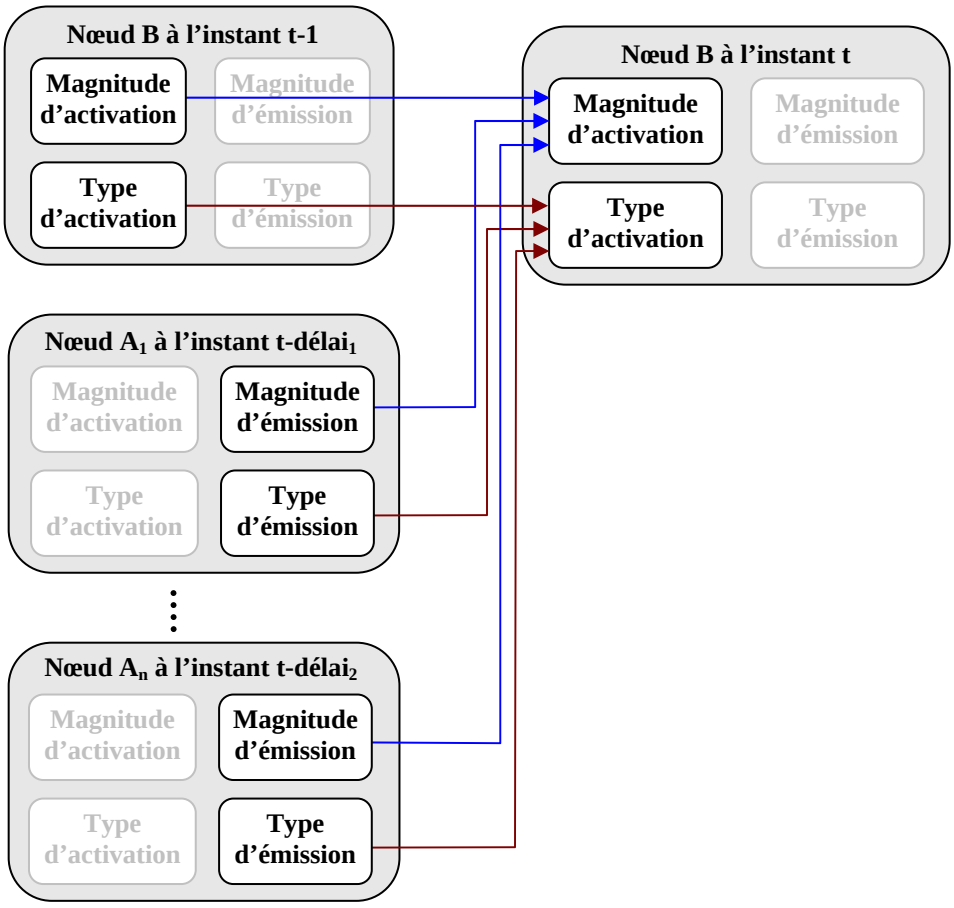

Figure IV.3.3 : Calcul des valeurs d'activation d'un nœud B possédant *n* parents $A_1,\ldots,A_n$.
Les arguments utilisés sont les valeurs d'émission des parents et les valeurs d'activation précédentes du nœud.

La combinaison de ces valeurs est différente suivant qu'il s'agit de magnitudes ou de types. Pour la magnitude, il est possible de distinguer deux façons de combiner deux ou plusieurs entrées : une combinaison linéaire ou une combinaison non-linéaire. Combiner des magnitudes en les

additionnant ou en les soustrayant revient respectivement à traiter des relations d'excitation ou d'inhibition, de la même manière qu'au niveau neuronal. Une combinaison multiplicative donne un rôle prépondérant à l'aire qui émet la magnitude servant de facteur. Ce type de combinaison est, par exemple, adapté à la modélisation d'un mécanisme de contrôle d'une aire sur une autre.

Le cas des types est particulier, en raison de leur nature symbolique. On distingue également deux sortes de combinaisons : entre types de même champ catégoriel, et entre types de champs catégoriels différents. Considérons les symboles d'après leur valeur sémantique. Un type étant une distribution de tous les symboles d'un champ catégoriel, le résultat de la combinaison de types de même champ catégoriel est alors un type défini également sur ce champ catégoriel. On combine des informations comparables sémantiquement.

Dans le cas où les types sont hétérogènes, c'est-à-dire définis sur des champs catégoriels différents, les informations qu'ils représentent ne sont pas comparables. Il n'est pas possible de considérer que le résultat de cette combinaison est un type défini sur l'union des champs catégoriels, car cela irait à l'encontre de la définition du champ catégoriel (symboles de même plan sémantique et de même granularité). En fait, la combinaison de types hétérogènes aboutit à une information d'une nature différente, dans le sens où le résultat est caractérisé par un plan sémantique plus élevé ou une granularité plus importante. Ce changement implique l'utilisation d'un champ catégoriel différent pour représenter l'information. Par exemple, la combinaison d'une information auditive et d'une information visuelle va correspondre à la description d'un objet (changement de plan sémantique). Mais ce changement de champ catégoriel ne se produit qu'au niveau de l'information émise par l'aire. A ce titre, le type d'activation correspond à une étape intermédiaire de la combinaison : ce type est défini sur un *domaine multiple*, c'est-à-dire sur un ensemble de champs catégoriels. Il s'agit d'une information composée, qui va être utilisée lors de la phase d'émission, pour construire un type défini sur un seul champ catégoriel, caractéristique du nœud.

3.2.2. Emission

La phase d'émission permet de déterminer si le nœud émet, et la nature de l'information qui est transmise aux nœuds situés en aval. L'information émise est une fonction de l'information qui a activé l'aire. A noter que le type et la magnitude d'émission peuvent dépendre à la fois du type et de la magnitude d'activation. La fonction d'émission permet d'exprimer une sorte de condition d'émission, et également de réaliser un traitement sur les valeurs d'activation.

Considérer la valeur de la magnitude comme condition d'émission revient à supposer qu'une aire cérébrale n'émet que si son activation atteint un niveau suffisant, par exemple si elle dépasse un certain seuil, ou en lui appliquant un fonction sigmoïdale, à l'instar de certains neurones formels. En raison de la définition de la magnitude, on peut considérer que ce seuil représente à la fois le nombre minimal de neurones devant s'activer, et le niveau d'activation moyen minimal, pour que l'aire émette. Ce choix est critiquable, car dans le cas de nombreux neurones activés à un faible niveau, on peut tout de même avoir émission. D'où l'intérêt de tenir également compte du type d'activation pour définir les conditions d'émission. En ce qui concerne le type, il s'agit de déterminer si la configuration des neurones activés peut déclencher une émission. En termes cognitifs, cela revient à savoir si l'aire reconnaît l'information intégrée lors de la phase d'activation. Par exemple, considérons une aire dont le rôle est de catégoriser des phonèmes. On peut supposer qu'elle émettra quand un phonème en particulier est représenté de façon prépondérante dans le type d'activation. Par contre, si l'information représentée par ce type correspond à un mélange de phonèmes dans lequel aucun ne ressort franchement, l'aire n'émettra pas de réponse même si la magnitude d'activation est importante. Bien sûr, il est possible d'avoir une condition d'émission dépendant à la fois de la magnitude et du type d'activation (Figure IV.3.4).

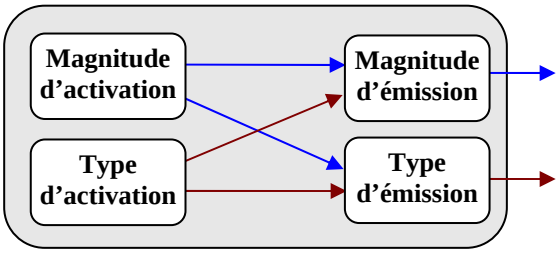

Figure IV.3.4 : Calcul des valeurs d'émission d'un nœud.
Les arguments utilisés sont les valeurs d'activation du nœud au même instant.

Le traitement sur les valeurs d'activation permet de calculer les valeurs d'émission. Pour la magnitude, il peut s'agir de n'importe quelle fonction définie sur les réels, puisque cette partie de l'information est numérique. En revanche, la transformation appliquée au type est plus intéressante, elle constitue un aspect essentiel dans la définition du rôle du nœud. On peut concevoir deux transformations possibles sur un type : soit simplement modifier la répartition des symboles en conservant le domaine de définition, soit créer à partir du type d'activation un type d'émission défini sur un autre domaine de définition, par exemple pour transformer un type d'activation multiple en un type d'émission simple (propre à être propagé).

La première option peut s'appliquer dans le cas où une zone intègre différentes sources d'informations de même type. On peut considérer le type d'activation comme une intégration brute des sources, et le type d'émission comme un raffinement de cette intégration. Ce raffinement peut consister, par exemple, en l'application d'une fonction qui va contraster la répartition des symboles. Dans ce cas, le type d'émission est bien défini sur le même champ catégoriel. Bien qu'elle puisse aussi s'appliquer dans le cas de types homogènes, la deuxième option est intéressante dans le cas où le nœud a intégré, lors de l'activation, des information de types hétérogènes, c'est-à-dire définis sur des champs catégoriels différents. En effet, comme il a été mentionné dans le paragraphe sur la phase d'activation, le type d'activation possède alors un domaine de définition multiple. Pour être cohérent avec la contrainte portant sur la nature du type émis par un nœud, ce type doit être transformé en un type simple. On peut prendre l'exemple d'un nœud recevant des informations visuelles et auditives décrivant un objet : le type d'activation est défini sur un domaine multiple. La transformation opérée lors de la phase d'émission va consister à traiter le type d'activation de manière à obtenir un type d'émission défini sur un champ catégoriel simple dont les symboles représentent des objets.

La deuxième sorte de transformation appliquée au type nécessite l'utilisation d'un paramètre du nœud appelé table de préférence des types (TPT, c.f. Figure IV.3.5).

> **Table de préférence des types** : la table de préférence des types (TPT) est le paramètre utilisé dans la fonction d'émission qui permet de définir la sensibilité de la population neuronale à certains types.

La TPT associe un type d'émission, défini sur le champ catégoriel propre à la sortie du nœud, à certaines combinaisons d'informations issues de l'activation. Elle permet également d'exprimer les différences de sensibilité de la population neuronale envers différents types d'activations.

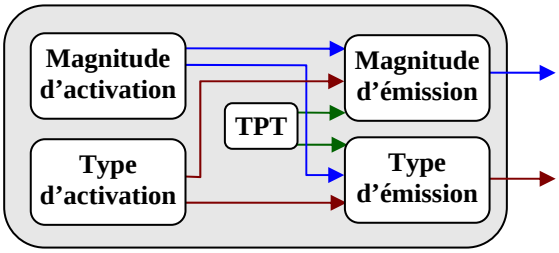

Figure IV.3.5 : La table de préférence des types (TPT) est un paramètre pouvant intervenir lors du calcul des valeurs d'émission.

Dans le cerveau, une aire qui a appris à reconnaître une information en particulier est en général très sensible à un stimulus la représentant, mais également un peu sensible à des stimuli représentant des information voisines. La TPT permet de représenter ce type de propriétés. Elle joue un rôle essentiel dans le mécanisme d'apprentissage, qui est basé sur une modification de la sensibilité du nœud.

4. PROCESSUS D'APPRENTISSAGE

Cette partie est consacrée à la description d'un ensemble de mécanismes destinés à simuler l'apprentissage à un niveau intégré. L'apprentissage considéré ici est de type adaptatif et non-supervisé. Il faut souligner que tous les nœuds d'un modèle n'utiliseront pas forcément ces mécanismes. En effet, suivant les caractéristiques de la fonction cognitive étudiée, les objectifs de modélisation ou la décomposition fonctionnelle de nœuds structurels, il peut être inutile, voire indispensable, que les nœuds représentant certaines populations n'apprennent pas.

Au niveau neuronal, l'apprentissage est réalisé essentiellement par des modifications qui interviennent dans la nature et l'importance des connexions, en fonction des activations neuronales. En raison de la nature duale de l'information manipulée, et notamment du type, les mécanismes qui implémentent l'apprentissage au niveau des populations de neurones sont différents.

4.1. Définitions

L'apprentissage consiste à modifier la réponse de la population lorsqu'elle reçoit une certaine information, dans certaines circonstances. Par circonstances, nous voulons souligner le fait que l'apprentissage nécessite la répétition (c.f. chapitre I.1.3.2) : il faut que l'information soit reçue un nombre de fois suffisant. Il n'est pas nécessaire que l'information soit exactement la même, il faut plutôt parler de catégorie d'information, englobant un ensemble de signaux aux caractéristiques proches, c'est pourquoi il est fait mention d'une *certaine* information.

Dans le modèle, une catégorie d'information correspond à un ensemble de types proches. D'après la description faite dans la partie précédente, l'information reçue par une population neuronale est le résultat de l'intégration de ses entrées, il s'agit donc de ses valeurs d'activation. De plus, la réponse de la population est issue du traitement de ces valeurs d'activation, effectué lors de la phase

d'émission. On peut donc considérer que l'apprentissage consiste à modifier la fonction d'émission suivant les valeurs d'activation, notamment le type, qui représente la partie sémantique. Or, la TPT est utilisée dans la fonction d'émission pour calculer l'information émise par la population neuronale. L'apprentissage peut finalement se définir comme suit :

> **Apprentissage** : ensemble de processus destinés à modifier la TPT en fonction des type et magnitude d'activation.

La Figure IV.4.1 illustre notre définition de l'apprentissage. La modification de la TPT a une influence sur le calcul des valeurs d'activation futures, ce qui explique le délai d'un instant représenté dans la figure.

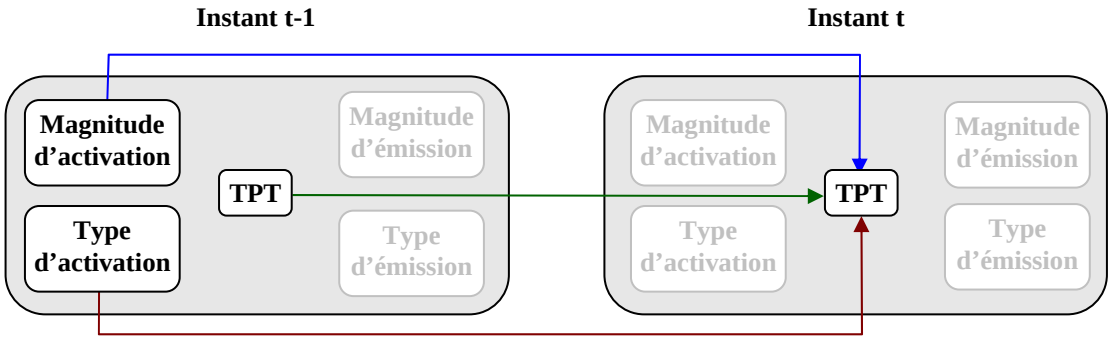

Figure IV.4.1 : L'apprentissage est réalisé en modifiant un paramètre du nœud, appelé table de préférence des types (TPT).
Cette modification est guidée par les valeurs d'activation précédentes du nœud.

4.2. Description des mécanismes

Le processus d'apprentissage se compose de quatre mécanismes complémentaires : le renforcement, l'introduction, l'oubli, et le glissement. Ils reposent tous sur le concept d'archétype :

> **Archétype** : un archétype est un type utilisé dans la TPT pour représenter une catégorie d'information.

Dans la TPT, une préférence est associée à chaque archétype :

> **Préférence** : valeur qui, dans la TPT, est associée à un archétype. Elle symbolise le niveau de reconnaissance de cette information par l'aire.

La répétition d'une information est à la base de l'apprentissage. Le mécanisme de renforcement permet d'implémenter cette propriété :

Renforcement : mécanisme consistant à augmenter la préférence d'un archétype dans la TPT, à chaque fois que le type d'activation reçu peut être considéré comme faisant partie de la catégorie d'information représentée par cet archétype.

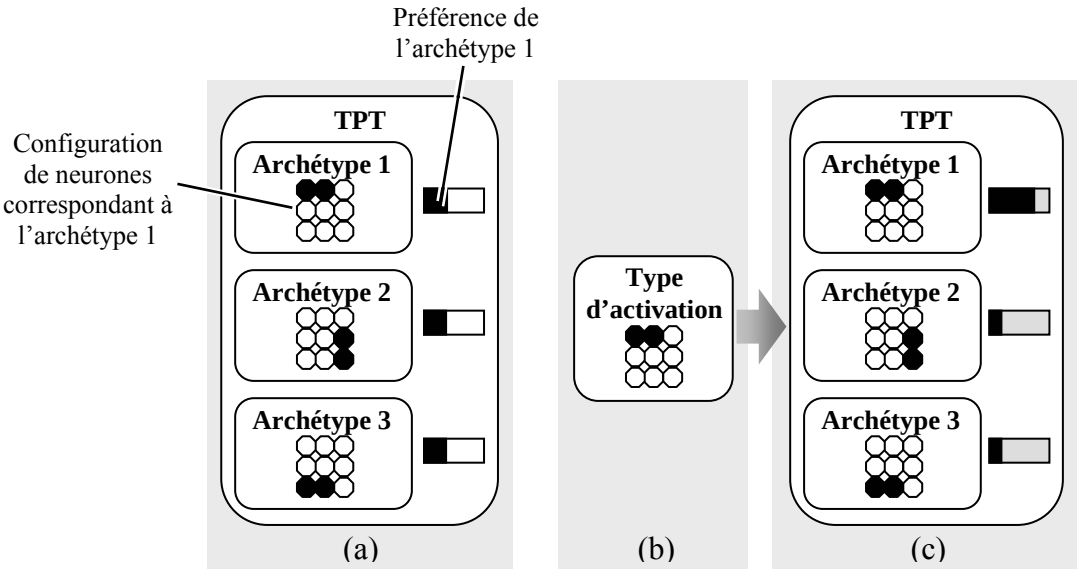

Figure IV.4.2 : le mécanisme de renforcement.
(a) : la TPT initiale. (b) : une information dont le type correspond à un archétype est répétée maintes fois. (c) : modifications des préférences.

En d'autres termes, à chaque fois que la configuration neuronale représentée par le type d'activation est suffisamment proche du modèle que constitue l'archétype, la préférence de cet archétype augmente. L'augmentation de la préférence envers un archétype entraîne la baisse de la préférence pour les autres archétypes. La conséquence directe est que les types les plus fréquents (et donc les plus répétés) sont les mieux reconnus.

Introduction : mécanisme consistant à introduire un nouvel archétype dans la TPT.

Le mécanisme d'introduction intervient lorsque le type d'activation représente une information inconnue, ou plutôt nouvelle, dans le sens où elle n'a jamais été traitée par le nœud jusque là. On considère que le type est inconnu quand il ne correspond à aucun archétype. Ce nouveau type étant susceptible d'être appris, il est nécessaire d'introduire un nouvel archétype lui correspondant, dans la TPT (Figure IV.4.3). Si le type est répété, la préférence pour l'archétype augmentera grâce au renforcement, confirmant ainsi l'apprentissage initié par son introduction.

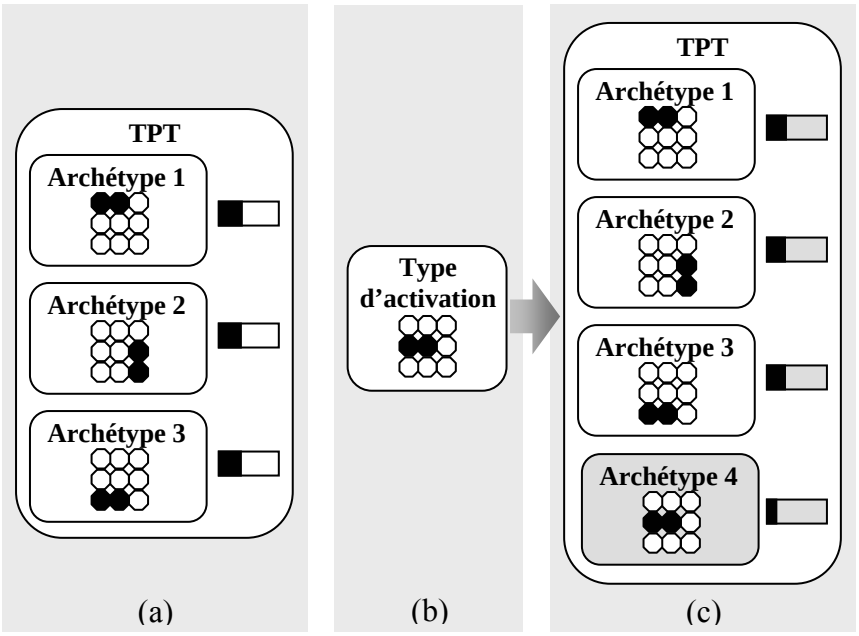

Figure IV.4.3 : le mécanisme d'introduction.
(a) : la TPT initiale. (b) : une information d'un nouveau type est soumise au nœud. (c) : un nouvel archétype est créé pour représenter cette nouvelle catégorie d'information.

Oubli : mécanisme consistant à supprimer de la TPT les archétypes à la préférence trop faible, c'est-à-dire trop peu fréquents.

Le mécanisme d'oubli est la contrepartie du mécanisme d'introduction. Grâce à lui, si un type nouvellement introduit n'est pas répété, il n'est pas conservé dans la TPT.

Glissement : mécanisme consistant à modifier les répartitions de symboles définissant les archétypes.

Le mécanisme de glissement concerne la nature même des archétypes. Des expériences comportant des tâches de catégorisation ont montré que les représentations internes de catégories pouvaient être modifiées au cours du temps (c.f. cbapitre I.1). Dans le modèle, cela revient à modifier les archétypes, au niveau de la configuration de neurones qui les définit. La modification se fait par rapport au type d'activation : s'il est proche mais néanmoins différent d'un archétype, celui-ci est légèrement modifié de façon à tendre vers le type d'activation (Figure IV.4.4). Ce mécanisme permet notamment de modéliser la propriété de plasticité cérébrale concernant le glissement d'une aire fonctionnelle (c.f. chapitre I.1.3).

Fusion : mécanisme consistant à réunir les archétypes assimilés à la même catégorie.

En raison du mécanisme de glissement, il est possible d'observer le rapprochement de la nature de certains archétypes. Pour éviter la présence de plusieurs archétypes représentant la même catégorie, nous introduisons le mécanisme de fusion, qui consiste à réunir les archétypes concernés ainsi que leurs préférences.

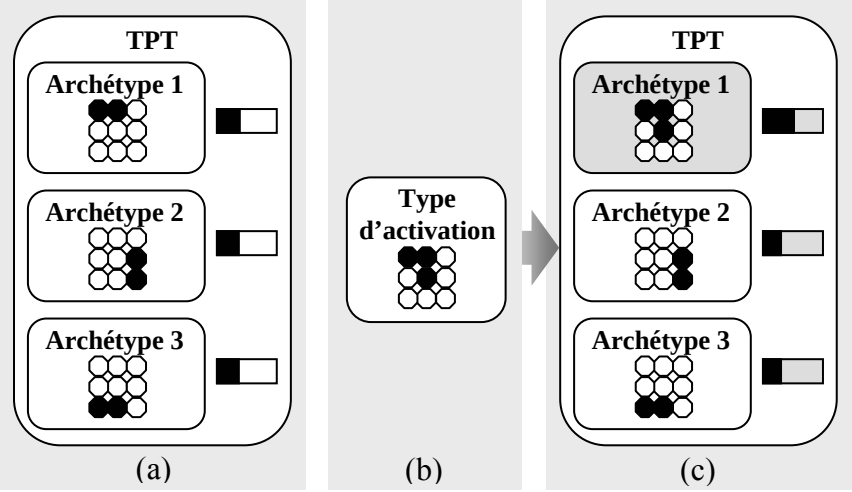

Figure IV.4.4 : le mécanisme de glissement.
(a) : la TPT initiale. (b) : une information dont le type est proche, mais différent, de l'archétype n°1 de la TPT, est présentée au nœud. (c) : au bout d'un grand nombre de répétitions, la nature de cet archétype est modifiée.

La magnitude intervient également dans ces différents mécanismes, bien que son importance soit moindre. En effet, on peut considérer que la signification d'un type n'est valable que si la magnitude associée est suffisamment élevée. En considérant le type d'activation, cela traduit la contrainte qu'un nombre de neurones assez important doit réagir à l'information. Cette contrainte est essentielle dans le cas du mécanisme d'introduction, car elle permet de limiter le nombre de nouveaux archétypes rajoutés dans la TPT.

5. CONCLUSION

Les mécanismes de propagation de l'information et d'apprentissage qui ont été décrits dans cette partie reposent sur un certain nombre de relations entre les différents composants (magnitude et type d'activation, magnitude et type d'émission, TPT) des nœuds fonctionnels. La Figure IV.5.1 donne une représentation graphique de ces relations.

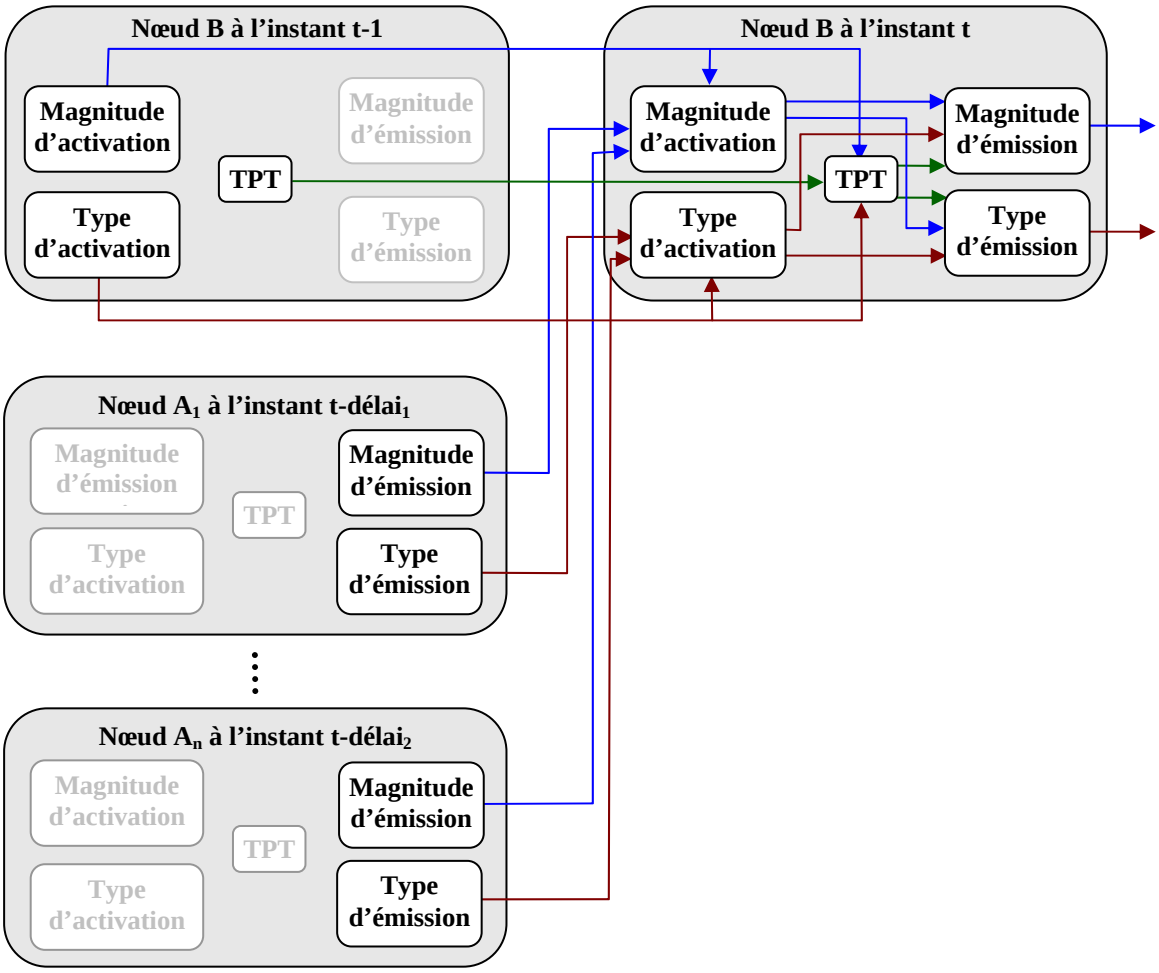

Figure IV.5.1 : ensemble des mécanismes d'apprentissage, de représentation et de traitement de l'information.

Chapitre V

RAGE : DEFINITION DU FORMALISME

Dans le chapitre précédent, les contraintes décrites au chapitre II.4 ont été exprimées sous la forme d'un certain nombre de concepts de modélisation. Dans le présent chapitre, nous allons traduire formellement ces concepts. Le résultat est RAGE (Réseaux Artificiels à Grande Echelle), un formalisme basé sur les réseaux bayésiens dynamiques, qui permettent de respecter au maximum ces contraintes. Nous avons essayé de définir des mécanismes suffisamment généraux, pour deux raisons. La première est d'obtenir un outil de modélisation ouvert et facilement adaptable. En effet, toute évolution des connaissances en neurosciences est susceptible d'entraîner des changements dans nos hypothèses de modélisation, qui auraient des répercussions sur l'outil lui-même. La seconde est de pouvoir éventuellement effectuer des extensions de notre formalisme, afin de l'adapter à la modélisation dans d'autres domaines, différents des neurosciences.

Si l'on considère les propriétés des réseaux dynamiques bayésiens (RBD), décrits dans le chapitre III.1.3.2, on s'aperçoit que le réseau dynamique du chapitre précédent est directement exprimable sous la forme d'un RBD. Le réseau dynamique présente une décomposition sous forme de tranches temporelles, ce qui permet de conserver la même structure dans le RBD. Les nœuds de ce RBD sont des variables représentant des magnitudes et des types, qui décrivent les états de populations neuronales. Les traitements implémentés par les nœuds sont des fonctions qui portent sur l'information transitant dans le réseau, c'est-à-dire les valeurs de magnitude et de type. Ces fonctions peuvent nécessiter de manipuler un paramètre variant au cours du temps, la table de préférence des types (TPT). Ce paramètre permet également d'implémenter des mécanismes d'apprentissage.

1. REPRESENTATION DE L'INFORMATION

1.1. Magnitude

La magnitude est une variable aléatoire réelle.

1.2. Type

1.2.1. Définitions préliminaires

Un *symbole*, tel qu'il est défini dans le chapitre IV.2.2, sera noté s. $\mathbb{S}$ est l'ensemble dénombrable de *tous les symboles existants*. Nous supposons qu'il existe un nombre infini de symboles.

Un *champ catégoriel*, noté S, est un sous ensemble de $\mathbb{S}$. Les sens (les valeurs sémantiques) associés aux symboles qui composent un champ catégoriel doivent respecter les contraintes de même plan sémantique et de même granularité, décrites dans le chapitre IV.2.2. L'ensemble de tous les champs catégoriels $\left\{S_1,\ldots,S_{n_{CC}}\right\}$ constitue une partition de $\mathbb{S}$, c'est-à-dire que l'on a les propriétés suivantes :

$$S_i \neq \varnothing \quad \left(1 \leq i \leq n_{CC}\right) \qquad \text{(eq. V.1.1)}$$

$$S_i \cap S_j = \varnothing \quad \left(1 \leq i, j \leq n_{CC} \text{ et } i \neq j\right) \qquad \text{(eq. V.1.2)}$$

$$\mathbb{S} = \bigcup_{i=1}^{n_{CC}} S_i \qquad \text{(eq. V.1.3)}$$

Un *domaine de définition simple* correspond à un champ catégoriel. Soit D un tel domaine, on a :

$$D = S_i \quad \left(1 \leq i \leq n_{CC}\right) \qquad \text{(eq. V.1.4)}$$

En revanche, un domaine *multiple* est le produit cartésien de plusieurs champs catégoriels :

$$D = S_{i_1} \times \ldots \times S_{i_m} \quad \left(1 \leq i_1,\ldots,i_m \leq n_{CC}\right) \qquad \text{(eq. V.1.5)}$$

Ici, D est le produit de m champs catégoriels. On note d un élément d'un domaine D. Si D est simple, d est un symbole, sinon (i.e. si D est multiple) il s'agit d'un m-uplet de symboles.

Considérons, par exemple, deux champs catégoriels S_a et S_b, l'un désignant des couleurs et l'autre des formes. Les symboles qui les composent sont respectivement *blanc*, *bleu*, *rouge* et *vert* ; et *circ* (ulaire), *rect* (angulaire) et *tri* (angulaire) :

$$S_a = \{blanc; bleu; rouge; vert\} \quad \text{(eq. V.1.6)}$$

$$S_b = \{circ; rect; tri\} \quad \text{(eq. V.1.7)}$$

Soit D_a le domaine simple tel que $D_a = S_a$. Soit D_b le domaine multiple tel que $D_b = S_a \times S_b$. On a alors :

$$D_a = \{blanc; bleu; rouge; vert\} \quad \text{(eq. V.1.8)}$$

$$D_b = \begin{Bmatrix} (blanc; cir); (blanc; rect); (blanc; tri); \\ (bleu; circ); (bleu; rect); (bleu; tri); \\ (rouge; circ); (rouge; rect); (rouge; tri); \\ (vert; circ); (vert; rect); (vert; tri) \end{Bmatrix} \quad \text{(eq. V.1.9)}$$

Les éléments de D_a sont des symboles, et ceux de D_b sont des couples de symboles.

1.2.2. Définition du type

Un *type*, noté T, décrit la répartition des symboles de son domaine de définition. On note $\mathbb{T}_D$ l'ensemble des types définis sur un domaine $D = \{d_1, \ldots, d_n\}$ comptant n éléments. Que son domaine soit simple ou multiple, un type est, formellement, une application de ce domaine dans $[0,1]$, avec la contrainte suivante :

$$\sum_{i=1}^{n} T(d_i) = 1 \ (\textit{contrainte de sommation}) \quad \text{(eq. V.1.10)}$$

Le type permet donc d'associer un poids p_i compris entre 0 et 1 à chaque élément d_i du domaine de définition. La répartition définie par un type T est notée de la façon suivante :

$$T = \{(d_1; p_1); \ldots; (d_n; p_n)\} \quad \text{(eq. V.1.11)}$$

Soulignons qu'il n'est pas nécessaire de définir explicitement *a priori* tous les éléments d'un domaine pour manipuler des types définis sur ce domaine. Par commodité, nous négligerons dorénavant de représenter les symboles de poids nul lors de la description d'un type. La contrainte de sommation vient du fait que les symboles se concurrencent dans la description de l'information (c.f. chapitre IV.2.2). De ce fait, si un symbole prend plus d'importance, il est normal que les autres symboles perdent de l'importance en retour. Cette notion est étendue aux types multiples, qui constituent une étape intermédiaire de l'intégration d'informations définies sur différents domaines (c.f. chapitre IV.3).

Reprenons les domaines D_a et D_b de l'exemple du paragraphe précédent (équations (eq. V.1.8) et (eq. V.1.9)). Soient T_a et T_b, deux types définis respectivement sur D_a et D_b. Ces types ont la forme suivante :

$$T_a = \{(blanc; p_1); (bleu; p_2); (rouge; p_3); (vert; p_4)\} \quad \text{(eq. V.1.12)}$$

$$T_b = \begin{Bmatrix} ((blanc; cir); p_{1,1}); ((blanc; rect); p_{1,2}); ((blanc; tri); p_{1,3}); \\ ((bleu; circ); p_{2,1}); ((bleu; rect); p_{2,2}); ((bleu; tri); p_{2,3}); \\ ((rouge; circ); p_{3,1}); ((rouge; rect); p_{3,2}); ((rouge; tri); p_{3,3}); \\ ((vert; circ); p_{4,1}); ((vert; rect); p_{4,2}); ((vert; tri); p_{4,3}) \end{Bmatrix} \quad \text{(eq. V.1.13)}$$

1.2.3. Combinaison de types

Nous définissons deux façons de combiner les types, à travers les opérateurs $\oplus$ et $\otimes$. Le premier concerne uniquement les types définis sur le même domaine, et le résultat est également défini sur ce même domaine. Le second permet de combiner des types qui ne sont pas forcément définis sur le même domaine, et le résultat est un type défini sur un domaine multiple. En raison de la façon dont les poids des types sont combinés, on nomme ces opérateurs respectivement opérateurs de combinaison linéaire et de combinaison non-linéaire.

a. Combinaison linéaire

Soient $T_1,\ldots,T_n$, n types définis sur un domaine D, et T le résultat de leur combinaison par l'opérateur de combinaison linéaire $\oplus$, lui aussi défini sur D. Soient $c_1,\ldots,c_n \geq 0$, n réels utilisés comme paramètres. On note :

$$T = \oplus\left(\left(T_1;c_1\right),\ldots,\left(T_n;c_n\right)\right) \qquad \text{(eq. V.1.14)}$$

Au niveau de la répartition des éléments du domaine, cette combinaison se traduit par :

$$\forall d \in D : T\left(d\right) = \frac{1}{\sum_{i=1}^{n} c_i} \sum_{i=1}^{n} c_i T_i\left(d\right) \qquad \text{(eq. V.1.15)}$$

Ce calcul revient à faire une moyenne arithmétique pondérée des poids associés aux d. On a :

$$\sum_{d \in D} T\left(d\right) = \sum_{d \in D} \left[\frac{1}{\sum_{i=1}^{n} c_i} \sum_{i=1}^{n} c_i T_i\left(d\right) \right] \qquad \text{(eq. V.1.16)}$$

$$\sum_{d \in D} T\left(d\right) = \frac{1}{\sum_{i=1}^{n} c_i} \sum_{i=1}^{n} \left[c_i \sum_{d \in D} T_i\left(d\right) \right] \qquad \text{(eq. V.1.17)}$$

D'après la contrainte de sommation propre aux types (eq. V.1.10), on a :

$$\sum_{d \in D} T\left(d\right) = \frac{1}{\sum_{i=1}^{n} c_i} \sum_{i=1}^{n} c_i = 1 \qquad \text{(eq. V.1.18)}$$

Par conséquent, l'opérateur de combinaison linéaire des types conserve bien la contrainte de sommation (eq. V.1.10).

Prenons l'exemple de deux types T_c et T_d définis sur le même domaine D_a, décrit précédemment (eq. V.1.8) :

$$T_c = \{(blanc;0,2),(bleu;0,8)\} \quad \text{(eq. V.1.19)}$$

$$T_d = \{(blanc;0,1),(bleu;0,2),(rouge;0,7)\} \quad \text{(eq. V.1.20)}$$

Soit T_e le résultat de la combinaison linéaire de T_c et T_d en utilisant les paramètres $c_c = 0,4$ et $c_d = 0,6$, on a alors :

$$T_e = \oplus((T_c;0,4),(T_d;0,6)) = \{(blanc;0,14),(bleu;0,44),(rouge;0,42)\} \quad \text{(eq. V.1.21)}$$

A noter que par souci de simplification, lorsque les coefficients c_i sont tous égaux au sein d'une même combinaison linéaire, ils sont omis dans la notation :

$$T = \oplus(T_1,\ldots,T_n) \quad \text{(eq. V.1.22)}$$

b. Combinaison non-linéaire

Soient $T_1,\ldots,T_n$, n types de domaines respectifs $D_1,\ldots,D_n$, et T le résultat de leur combinaison par l'opérateur de combinaison non-linéaire $\otimes$. Les domaines $D_1,\ldots,D_n$ sont en général tous différents, mais il peut tout aussi bien y avoir des domaines identiques. T est un type défini sur le domaine multiple D, tel que :

$$D = D_1 \times \ldots \times D_n \quad \text{(eq. V.1.23)}$$

On note :

$$T = \otimes(T_1,\ldots,T_n) \quad \text{(eq. V.1.24)}$$

Au niveau de la répartition des éléments du domaine, cette combinaison se traduit par :

$$\forall d \in D : T(d) = \prod_{i=1}^{n} T_i(d_i) \quad \text{(eq. V.1.25)}$$

avec $d = (d_1,\ldots,d_n)$ et $d_i \in D_i$. On a :

$$\sum_{d \in D} T(d) = \sum_{d \in D} \prod_{i=1}^{n} T_i(d_i) \qquad \text{(eq. V.1.26)}$$

En développant puis en factorisant, on obtient :

$$\sum_{d \in D} T(d) = \prod_{i=1}^{n} \sum_{d_i \in D_i} T_i(d_i) \qquad \text{(eq. V.1.27)}$$

Or d'après la contrainte de sommation propre aux types (eq. V.1.10), on a :

$$\sum_{d \in D} T(d) = \prod_{i=1}^{n} 1 = 1 \qquad \text{(eq. V.1.28)}$$

L'opérateur de combinaison non-linéaire des types préserve donc bien la contrainte de sommation.

Reprenons le type T_c, utilisé dans l'exemple du paragraphe précédent et défini par (eq. V.1.19). Soit le domaine simple D_c, défini par le champ catégoriel S_b décrit précédemment (eq. V.1.7). Soit T_f le type défini sur D_c, tel que :

$$T_f = \{(circ; 0,9), (rect; 0,1)\} \qquad \text{(eq. V.1.29)}$$

Soit T_g le résultat de la combinaison non-linéaire de T_c et T_f :

$$T_g = \otimes(T_c, T_f) = \begin{Bmatrix} ((blanc, circ); 0,18); ((blanc; rect); 0,02); \\ ((bleu; circ); 0,72); ((bleu; rect); 0,08) \end{Bmatrix} \qquad \text{(eq. V.1.30)}$$

c. Comparaison des opérateurs

Le premier opérateur peut être vu comme un moyen de combiner des types décrivant un même aspect de l'information. Il s'agit de types concurrents, définis sur le même domaine. L'opérateur non-linéaire permet d'intégrer des aspects complémentaires de l'information. Le type obtenu est donc un type complexe, caractérisé par un domaine regroupant plusieurs champs catégoriels. Les types combinés de façon non-linéaire ne sont pas forcément de domaines différents, il est possible de combiner non-linéairement des types de même domaine. Supposons par exemple que deux types représentent chacun une note de musique : ils sont donc définis sur le même domaine. On

veut réaliser leur intégration de manière à traiter un accord musical. Pour décrire cette information, les deux notes sont complémentaires, et on va donc utiliser la combinaison non-linéaire. S'il s'était agi d'informations concurrentielles, il aurait fallu utiliser l'opérateur linéaire.

2. DESCRIPTION DES NŒUDS

Un nœud du réseau dynamique est représenté par cinq nœuds dans le RBD : quatre d'entre eux représentent l'état de ce nœud à un instant donné, et le dernier est la table de préférence des types.

2.1. Etat d'un nœud

Soit un X^t, nœud du réseau dynamique. Son état est décrit par l'information qui y transite. Cette information se constitue de quatre valeurs différentes : magnitude et type d'activation et magnitude et type d'émission. Dans le RBD, chacune de ces valeurs est modélisée par un nœud distinct, selon les principes décrits dans la partie précédente. On note respectivement $\tilde{M}_X^t$ et M_X^t les variables représentant les magnitudes d'activation et d'émission, et $\tilde{T}_X^t$ et T_X^t celle dédiées aux types d'activation et d'émission. Il n'y a pas de différence de représentation entre la magnitude d'activation et la magnitude d'émission. Ce n'est pas le cas pour les types, puisque les types d'activation peuvent être simples ou multiples, alors que les types d'émission sont forcément simples. De plus, nous avons fait, au chapitre précédent, l'hypothèse que le domaine de définition du type d'émission était invariable. Or, un type d'activation est lui-même directement calculé à partir de différents types d'émission. Par conséquent, son domaine ne varie pas non plus.

2.2. Table de préférence des types

La table de préférence des types (TPT) peut être considérée comme un paramètre qui évolue au cours du temps. Il intervient dans les mécanismes d'émission utilisant les types, et lors de l'apprentissage. Le terme de TPT regroupe, en fait, trois entités distinctes : deux fonctions et un ensemble de types. L'utilisation de fonctions rend la description de la TPT plus compréhensible et sa manipulation plus aisée.

2.2.1. Fonction de préférence et archétypes

Considérons un nœud X^t du réseau dynamique. Soit $\tilde{D}_X$, le domaine de définition de ses types d'activation. L'ensemble des types d'activation possibles est donc noté $\mathbb{T}_{\tilde{D}_X}$. La *fonction de préférence*, notée $pref$, associe à chaque type de $\mathbb{T}_{\tilde{D}_X}$ une valeur dans $[0,1]$, avec la contrainte de sommation suivante :

$$\sum_{T \in \mathbb{T}_{\tilde{D}X}} pref_X^t(T) = 1 \text{ (\textit{contrainte de sommation})} \qquad \text{(eq. V.2.1)}$$

La valeur $pref_X^t(T)$ représente le niveau de reconnaissance d'un type T par le nœud X^t. En limitant la fonction aux types définis sur $\tilde{D}_X$, on se restreint aux seuls types susceptibles d'être soumis au nœud (puisque nous avons fait l'hypothèse que $\tilde{D}_X$ n'évoluait pas). La contrainte de sommation est utile pour modéliser la propriété décrite dans le chapitre IV.4, qui veut que les préférences soient liées : l'augmentation de la préférence d'un type fait baisser celles des autres types. En résumé, cette fonction permet de quantifier le niveau de reconnaissance pour tout type d'activation possible. Ces préférences peuvent évoluer au cours du temps, grâce aux mécanismes d'apprentissage (c.f. chapitre IV.4).

Bien sûr, la plupart des types ont une préférence nulle, car seul un nombre restreint d'entre eux sont pertinents pour la population. Il s'agit des archétypes, décrits dans le chapitre IV.4. On note $Arch_X^t$ l'*ensemble des archétypes* pour un nœud X^t du réseau dynamique :

$$Arch_X^t = \left\{ T : T \in \mathbb{T}_{\tilde{D}_X} \wedge pref_X^t(T) \neq 0 \right\} \qquad \text{(eq. V.2.2)}$$

Puisqu'il est défini par rapport à $pref$, l'ensemble des archétypes est lui aussi susceptible d'évoluer au cours du temps, dans le cadre de l'apprentissage.

Pour reprendre l'exemple du domaine D_b (équation (eq. V.1.9)), considérons un type d'activation de la même forme que T_b (équation (eq. V.1.13)). On peut imaginer que le nœud dynamique considéré est chargé d'intégrer des informations visuelles de deux sortes différentes, puisque certaines concernent la forme d'un objet alors que d'autres en décrivent la couleur. Supposons qu'il

reconnaisse quatre catégories d'information : les ronds rouges, les ronds bleus, les triangles rouges et les rectangles blancs. On peut les représenter respectivement par les quatre archétypes suivants :

$$A_a = \{((rouge; circ); 1)\} \quad \text{(eq. V.2.3)}$$

$$A_b = \{((bleu; circ); 1)\} \quad \text{(eq. V.2.4)}$$

$$A_c = \{((rouge; tri); 1)\} \quad \text{(eq. V.2.5)}$$

$$A_d = \{((blanc; rect); 1)\} \quad \text{(eq. V.2.6)}$$

Pour des archétypes également préférés, on obtient alors :

$$pref_X^t(T) = \begin{cases} 0,25 \text{ si } T \in \{A_a; A_b; A_c; A_d\} \\ 0 \quad \text{sinon} \end{cases} \quad \text{(eq. V.2.7)}$$

$$Arch_X^t = \{A_a; A_b; A_c; A_d\} \quad \text{(eq. V.2.8)}$$

2.2.2. Fonction d'association

Un des rôles de la TPT est de permettre un changement de domaine de définition entre le type d'activation et le type d'émission. Ce traitement correspond à une intégration sémantique de l'information en entrée (c.f. chapitre IV.4), et permet d'avoir un type simple en sortie (rappel : le domaine de définition d'un type d'émission est toujours simple, c'est-à-dire qu'il s'agit d'un seul champ catégoriel).

C'est la *fonction d'association* de symboles, notée $symb$, qui permet d'implémenter ce mécanisme. Elle associe à chaque archétype un symbole appartenant au domaine de définition du type d'émission. En effet, chaque archétype est censé désigner une catégorie d'information en entrée du nœud. Cette catégorie est représentée par un symbole approprié dans le type d'émission. Soit D_X le domaine de définition du type d'émission de X^t. Formellement, la fonction $symb$ est donc définie de $Arch_X^t$ dans D_X, c'est-à-dire qu'à une répartition de symboles correspond un seul symbole.

La fonction peut varier au cours du temps, si des mécanismes d'apprentissage sont présents. Ceci est dû en particulier au glissement, qui va modifier la correspondance entre un archétype et un symbole du domaine d'émission (c.f. chapitre V.4.2).

Reprenons l'exemple de TPT du paragraphe précédent. Supposons que les informations visuelles dont le nœud dynamique dispose décrivent des panneaux de signalisation routière. Le rôle du nœud est d'intégrer les deux plans sémantiques différents (couleur et forme) et de produire une infirmation synthétique, par exemple concernant le sens du panneau. Considérons le champ catégoriel S_c décrivant ce sens, composé des symboles *avert* (issement), *indic* (ation), *inter* (diction) et *oblig* (ation) :

$$S_c = \{avert; indic; inter; oblig\} \quad \text{(eq. V.2.9)}$$

Supposons que le type d'émission du noeud soit défini sur le domaine simple $D_d = S_c$. La fonction d'association doit associer un symbole à chaque archétype, par exemple :

$$symb_X^t(\mathrm{A_a}) = inter \quad \text{(eq. V.2.10)}$$

$$symb_X^t(\mathrm{A_b}) = oblig \quad \text{(eq. V.2.11)}$$

$$symb_X^t(\mathrm{A_c}) = avert \quad \text{(eq. V.2.12)}$$

$$symb_X^t(\mathrm{A_d}) = indic \quad \text{(eq. V.2.13)}$$

On exprime ici le fait que les panneaux rond et rouges signalisent une interdiction, les ronds et bleus une obligation, et les triangulaires et rouges un avertissement.

3. PROPAGATION ET TRAITEMENT DE L'INFORMATION

Le traitement effectué sur l'information est représenté par plusieurs fonctions mathématiques. Ces fonctions sont statiques, c'est-à-dire que leur forme ne varie pas dans le temps. Néanmoins, l'utilisation et la modification dynamique de paramètres permettent de modifier le comportement d'une fonction. Ici, ces fonctions modélisent les mécanismes de traitement et de propagation de

l'information décrits au chapitre précédent. Ces mécanismes sont dissociés en deux phases : activation et émission, chaque phase étant caractérisée par un traitement particulier. Ceci a des répercussions sur la forme et les propriétés des fonctions utilisées pour la modélisation, c'est pourquoi les fonctions utilisées pour représenter la phase d'activation et celles dédiées à la phase d'émission vont être décrites séparément.

3.1. Activation

Les fonctions d'activation associées à un nœud du réseau dynamique ont pour rôle de réaliser l'intégration spatiale et temporelle de ses entrées. Dans le contexte du RBD, cela consiste à calculer la valeur d'une variable représentant une valeur d'activation (magnitude ou type). A ce niveau, il n'y a pas d'interaction entre magnitude et type. Par conséquent, une magnitude d'activation résulte exclusivement de la combinaison de magnitudes, et le même principe s'applique aux types.

En ce qui concerne la combinaison de magnitudes, il s'agit de traiter des valeurs numériques, par conséquent il est possible d'utiliser n'importe quelle fonction réelle : somme pondérée, multiplications, combinaisons des deux, ou autres fonctions, linéaires ou non-linéaires. La combinaison de types repose exclusivement sur les deux opérateurs définis précédemment.

Pour un nœud X^t du réseau dynamique, on note respectivement $\tilde{f}_{M_X}$ et $\tilde{f}_{T_X}$ les fonctions permettant de calculer la magnitude et le type d'activation $\tilde{M}_X^t$ et $\tilde{T}_X^t$. Ces fonctions n'évoluent pas au cours du temps. Soient $Y_1^{t-\partial_1},\ldots,Y_n^{t-\partial_n}$ les nœuds qui influent sur X^t dans le réseau dynamique. Les $\partial_1,\ldots,\partial_n$ sont des délais associés à ces influences. La forme générale d'une fonction d'activation implique l'utilisation des valeurs d'émissions des $Y_1^{t-\partial_1},\ldots,Y_n^{t-\partial_n}$, ainsi que de la valeur d'activation précédente de X^t :

$$\tilde{M}_X^t = \tilde{f}_{M_X}\left(M_{Y_1}^{t-\partial_1},\ldots,M_{Y_n}^{t-\partial_n},\tilde{M}_X^{t-1},\tilde{u}_X^t\right) \qquad \text{(eq. V.2.14)}$$

$$\tilde{T}_X^t = \tilde{f}_{T_X}\left(T_{Y_1}^{t-\partial_1},\ldots,T_{Y_n}^{t-\partial_n},\tilde{T}_X^{t-1}\right) \qquad \text{(eq. V.2.15)}$$

Dans l'équation (eq. V.2.14), $\tilde{u}_X^t$ est une variable aléatoire (v.a.) normale de moyenne nulle (c.f. annexe A.1.1.5), indépendante de toute autre v.a. Elle représente les influences non-modélisées et

les erreurs, et correspond à la variable W utilisée dans le chapitre III.1.3.1 pour définir une relation dans un RBD (eq. III.1.16). En général, $\tilde{u}_X^t$ sera utilisée comme une variable additive, mais cela n'est pas obligatoire.

Les variables représentant des types ne sont pas des v.a., d'où l'absence d'une v.a. équivalente à $\tilde{u}_X^t$ dans l'équation (eq. V.2.15). La combinaison des types se fait en utilisant les deux opérateurs précédemment définis, $\oplus$ et $\otimes$. Il est possible de les composer, tant que cela respecte leurs définitions. On peut par exemple calculer la combinaison linéaire des résultats de deux combinaisons non-linéaires, à condition que ces derniers soient définis sur le même domaine. Rappelons que nous avons fait l'hypothèse que les domaines de définition des types d'activation et d'émission associés à un nœud ne varient pas au cours du temps.

3.2. Emission

Les fonctions d'émission associées à un nœud du réseau dynamique utilisent les valeurs d'activation pour déterminer les valeurs d'émission. On peut utiliser à la fois la magnitude et le type d'activation pour déterminer chacune des deux valeurs d'émission. En effet, on considère que l'information émise par la population peut dépendre à la fois de son niveau d'activation et de sa nature (répartition). Les fonctions d'émission sont notées f_{M_X} et f_{T_X}, et leur forme générale est :

$$M_X^t = f_{M_X}\left(\tilde{M}_X^t, \tilde{T}_X^t, u_X^t\right) \qquad \text{(eq. V.2.16)}$$

$$T_X^t = f_{T_X}\left(\tilde{M}_X^t, \tilde{T}_X^t\right) \qquad \text{(eq. V.2.17)}$$

où la variable u_X^t est une v.a. normale de valeur nulle. Les mécanismes de traitement des types nécessitent de comparer le type d'activation aux archétypes de la TPT. Pour cela, une fonction spécifique appelée fonction de similitude est utilisée. Elle intervient dans le calcul du type d'émission, et peut également jouer un rôle lors du calcul de la magnitude d'émission.

3.2.1. Fonction de similitude

La *fonction de similitude* associe à un couple de types forcément définis sur le même domaine, une valeur appartenant à $[0,1]$. Cette valeur numérique représente une mesure de la similitude entre les deux types : plus elle est élevée, plus les types sont semblables. Il est possible de définir plusieurs fonctions de similitudes différentes.

Ainsi, la fonction *cosinus* est utilisée entre autres dans les réseaux ART2 (c.f. paragraphe III.2.2) et dans d'autres formalismes plus anciens [Levine '00] pour mesurer la similitude. Les deux informations à comparer sont d'abord exprimées sous la forme de deux vecteurs. Puis, on calcule le cosinus de l'angle formé par ces deux vecteurs, et on utilise cette valeur comme mesure de similitude. Dans notre cas, cela revient à considérer les poids de deux types comme les coordonnées définissant deux vecteurs. Le cosinus de l'angle formé par deux vecteurs s'obtient en faisant le rapport de leur produit scalaire par le produit de leurs normes. Pour deux types T_1 et T_2 définis sur le même domaine D, on a :

$$sim_{\cos}(T_1,T_2) = \cos\left(\widehat{\vec{T_1},\vec{T_2}}\right) = \frac{\vec{T_1}\vec{T_2}}{\left\|\vec{T_1}\right\|\left\|\vec{T_2}\right\|} \qquad \text{(eq. V.2.18)}$$

Si les types sont égaux, l'angle formé par les vecteurs est nul et la valeur de son cosinus est 1. S'ils sont complètement différents, les vecteurs forment un angle droit, dont le cosinus vaut zéro (Figure V.3.1). Cette mesure de similitude repose sur une définition euclidienne de la distance.

Nous définissons également une fonction de similitude linéaire, basée sur une définition linéaire de la distance entre deux types T_1 et T_2 :

$$dist(T_1,T_2) = \frac{1}{2}\sum_{d\in D}\left|T_1(d) - T_2(d)\right| \qquad \text{(eq. V.2.19)}$$

Il s'agit en fait d'une version normalisée de la distance de Manhattan [Skiena '90]. La distance compare les deux répartitions de symboles en considérant les poids associés à un même symbole dans chaque type. La similitude est alors exprimée par :

$$sim_{lin}(T_1,T_2) = 1 - dist(T_1,T_2) \qquad \text{(eq. V.2.20)}$$

Mais il s'est révélé, à l'utilisation, qu'une fonction offrant des mesures de similitude plus contrastées était nécessaire. Notre choix s'est porté sur l'utilisation d'une fonction sigmoïdale, qui permet, lorsqu'on l'applique à notre distance (eq. V.2.19), d'avoir une courbe plus aplatie près des extrêmes que ce n'est le cas avec la fonction linéaire précédente (eq. V.2.20). La formule générale d'une sigmoïde est :

$$\sigma(x) = \frac{1}{1+e^{-a(x-b)}} \qquad \text{(eq. V.2.21)}$$

où a et b sont deux réels. Le coefficient a détermine la pente de la sigmoïde (croissante si $a > 0$ et décroissante pour $a < 0$) et le paramètre b l'abscisse de son centre de symétrie. Dans notre cas, x correspond à la distance entre deux types (eq. V.2.19). Finalement, la fonction de similitude sigmoïdale, notée sim_{sig}, est définie par :

$$sim_{sig}(T_1, T_2) = \frac{1}{1+e^{-a(dist(T_1,T_2)-b)}} \qquad \text{(eq. V.2.22)}$$

La Figure V.3.1 est un comparatif des trois fonctions de similitude présentées ici. La similitude linéaire correspond à (eq. V.2.20), la sigmoïde à (eq. V.2.22), et la fonction cosinus à (eq. V.2.18). La fonction sigmoïdale a pour paramètres $a = -13$ (pente modérée) et $b = 0,5$ (courbe centrée à mi-chemin entre les valeurs minimale et maximale de distance).

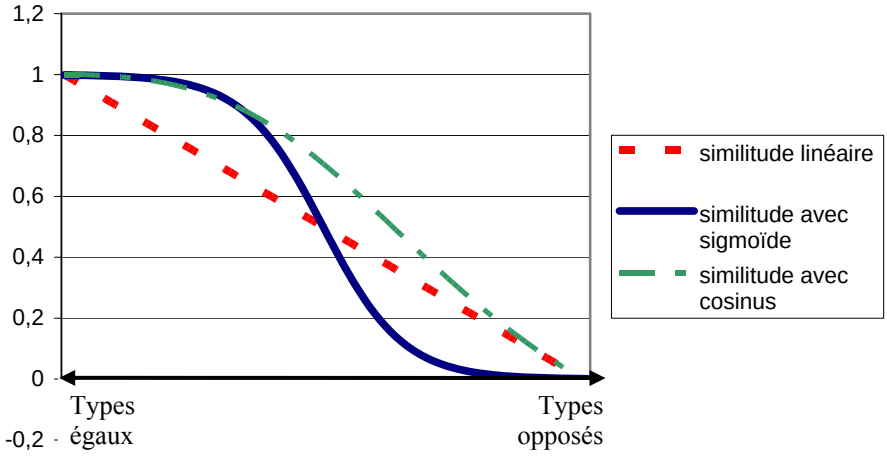

Figure V.3.1 : comparaison entre trois mesures de similitude.

Les valeurs en ordonnée sont des mesures de similitude incluses dans $[0,1]$. Elles sont calculées en fonction de deux types définis sur un domaine composé de deux symboles s_1 et s_2. A l'origine, les deux types sont égaux à $((s_1;0);(s_2;1))$ (similitude maximale). Puis, le premier type est fixé à cette valeur, pendant que le second tend vers $((s_1;1);(s_2;0))$ (similitude minimale).

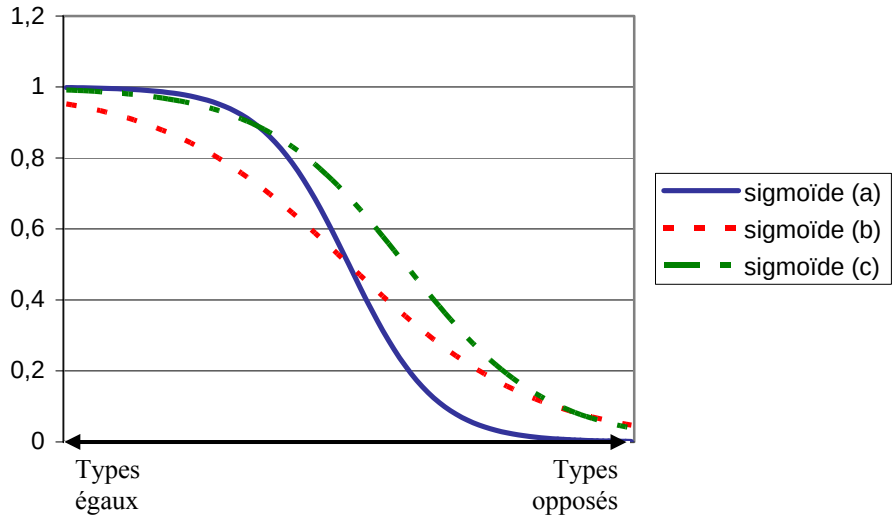

Figure V.3.2 : comparaison entre trois mesures de similitude basées sur des sigmoïdes.
Les trois fonctions sont basées sur (eq. V.2.22), avec des paramètres différents. (a) : $a=-13$ et $b=0,5$. (b) : $a=-6$ et $b=0,5$. (c) : $a=-8$ et $b=0,6$.

La pente détermine l'écrasement de la courbe. Plus la pente est forte, plus on se rapproche d'une fonction binaire (i.e. renvoyant des valeurs proches de 1 ou de 0). Plus elle est faible, plus on se rapproche d'une fonction linéaire semblable à (eq. V.2.20). Le fait de déplacer le centre de symétrie permet en quelque sorte d'avantager certaines distances. Si l'abscisse est proche de 0, seule des distances très faibles engendreront des mesures de similitude maximales (et inversement pour une abscisse proche de 1). En jouant sur ces paramètres, il est possible, en utilisant seulement des sigmoïdes, de se rapprocher des similitudes linéaires (Figure V.3.2, courbe b) et cosinus (Figure V.3.2, courbe c).

A noter qu'il est possible de définir une fonction différente pour chaque nœud d'un modèle. Dans l'exemple suivant, nous avons choisi d'utiliser une sigmoïde avec une pente $a=-13$, qui permet de contraster la fonction tout en conservant une certaine subtilité. La valeur $b=0,5$ permet de centrer la fonction sur la mesure de similitude médiane, comme c'était le cas pour (eq. V.2.20). Considérons

les types T_c et T_d définis précédemment par les équations (eq. V.1.19) et (eq. V.1.20). La distance entre ces types est :

$$dist(T_c, T_d) = \frac{1}{2}\sum_{d \in D} \left| T_c(d) - T_d(d) \right| \quad \text{(eq. V.2.23)}$$

$$dist(T_c, T_d) = \frac{1}{2}\left(\begin{array}{l} \left| T_c(blanc) - T_d(blanc) \right| + \left| T_c(bleu) - T_d(bleu) \right| \\ + \left| T_c(rouge) - T_d(rouge) \right| \end{array} \right) \quad \text{(eq. V.2.24)}$$

$$dist(T_c, T_d) = \frac{1}{2}(0,1 + 0,6 + 0,7) = 0,7 \quad \text{(eq. V.2.25)}$$

La similitude est alors donnée en appliquant (eq. V.2.22) :

$$sim_{sig}(T_c, T_d) = \frac{1}{1 + e^{13(0,7-0,5)}} = 0,06913842 \quad \text{(eq. V.2.26)}$$

3.2.2. Type d'émission

Pour un nœud donné, le domaine de définition du type d'émission est défini par la TPT. Le type d'émission est calculé à partir des magnitude et type d'activation. Le type d'activation est utilisé pour calculer la répartition de symboles qui correspondra au type d'émission. La magnitude d'activation intervient pour moduler le type résultant de ce calcul. En effet, si la magnitude d'activation est trop faible, on suppose que le nœud ne peut pas traiter le type d'activation, et qu'il émet donc un type très bruité. Au contraire, si la magnitude est suffisamment élevée, le type d'émission doit être intelligible.

Pour représenter cette modulation, nous utilisons une combinaison linéaire :

$$T_X^t = \oplus\left(\left(bruit_X^t ; 1 - f_{T_X}^{(1)}\left(\overline{\tilde{M}_X^t} \right) \right); \left(f_{T_X}^{(2)}\left(\tilde{T}_X^t \right); f_{T_X}^{(1)}\left(\overline{\tilde{M}_X^t} \right) \right) \right) \quad \text{(eq. V.2.27)}$$

Cette combinaison linéaire fait intervenir, d'une part, un type calculé en fonction du type d'activation et grâce à la TPT $f_{T_X}^{(2)}\left(\tilde{T}_X^t \right)$, et d'autre part, un type représentant du bruit, noté $bruit_X^t$. Le type représentant du bruit est calculé grâce à la TPT du nœud. En effet, le nœud est capable d'émettre des types contenant un nombre limité de symboles, correspondant au nombre

d'archétypes qu'il reconnaît. Le type $bruit_X^t$ est créé en affectant un poids équivalent à tous ces symboles. Soient $Arch_X^t = \{A_1, \ldots, A_n\}$ l'ensemble des archétypes de TPT_X^t, et $s_1, \ldots, s_n$ les symboles associés à ces archétypes par la fonction d'association $symb_X^t$. On a alors :

$$bruit_X^t = \left\{ \left(s_i ; \frac{1}{Card\left(Arch_X^t\right)} \right) \right\} \qquad \text{(eq. V.2.28)}$$

La combinaison linéaire (eq. V.2.27) est pondérée par une fonction de la magnitude d'activation $f_{T_X}^{(1)}\left(\overline{\tilde{M}_X^t}\right)$.

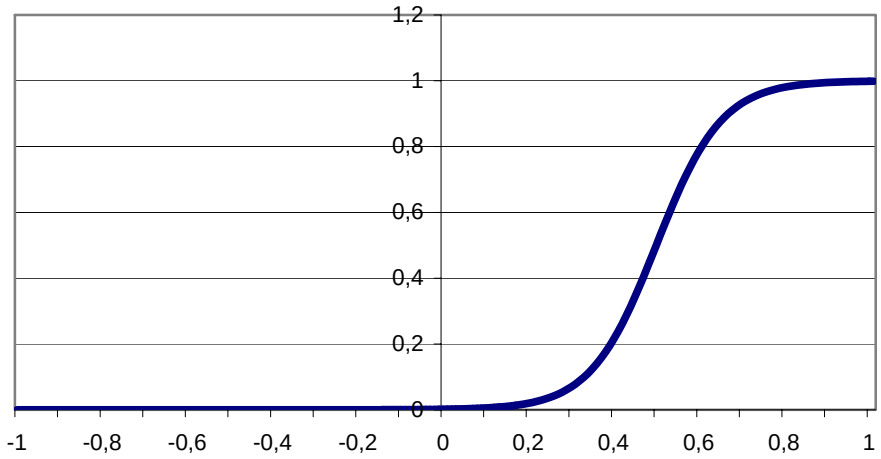

Figure V.3.3 : exemples de fonctions $f_{T_X}^{(1)}$ dans (eq. V.2.27).
Les valeurs en abscisse sont des magnitudes, et les valeurs en ordonnée représentent des coefficients réels inclus dans $[0,1]$.

Une magnitude étant représentée par une variable aléatoire et un type par une variable classique, nous utilisons ici seulement la moyenne de la v.a. représentant la magnitude d'activation. La fonction $f_{T_X}^{(1)}$ doit être définie de $\mathbb{R}$ dans $[0,1]$ et doit être croissante. De cette façon, plus la magnitude est grande, moins la part prise par le bruit dans l'élaboration de T_X^t est importante. Il est par exemple possible d'utiliser une fonction sigmoïdale telle que celle présentée dans la Figure V.3.3.

L'intérêt de la fonction sigmoïdale est qu'il est possible d'utiliser le paramètre déterminant le centre de symétrie comme une sorte de seuil (c.f. paragraphe 3.2.1) en dessous duquel le nœud émet surtout du bruit, tout en conservant une fonction continue. Par exemple, dans la Figure V.3.3, le seuil est fixé à 0,5.

La fonction $f_{T_X}^{(2)}$ s'applique au type d'activation $\tilde{T}_X^t$. Néanmoins, à la différence de la fonction utilisée pour calculer le type d'activation, elle n'est pas une combinaison des opérateurs $\oplus$ et $\otimes$. Notons U_X^t le résultat de $f_{T_X}^{(2)}\left(\tilde{T}_X^t\right)$. Il est défini sur le domaine de définition D_X (qui est celui du type d'émission) par la répartition de symboles suivante :

$$\forall d \in D_X : U_X^t(d) = \begin{cases} \dfrac{sim\left(\tilde{T}_X^t, A_i\right) pref_X^t\left(A_i\right)}{\sum_{k=1}^{n} sim\left(\tilde{T}_X^t, A_k\right) pref_X^t\left(A_k\right)} & \text{si } \exists s_i : d = s_i \\ 0 & \text{sinon} \end{cases} \qquad \text{(eq. V.2.29)}$$

où $Arch_X^t = \{A_1, \ldots, A_n\}$ est l'ensemble des archétypes de TPT_X^t, et s_i est le symbole associé à l'archétype A_i. Le principe est le suivant : dans un premier temps, les similitudes entre le type d'activation et chaque archétype de la TPT sont calculées :

$$sim\left(\tilde{T}_X^t, A_i\right) \quad (1 \le i \le n) \qquad \text{(eq. V.2.30)}$$

Puis, ces valeurs sont pondérées par les préférences de la TPT, en utilisant la fonction de préférence $pref_X^t$ (paragraphe 2.2.1). On obtient un ensemble de produits :

$$sim\left(\tilde{T}_X^t, A_i\right) pref_X^t\left(A_i\right) \qquad \text{(eq. V.2.31)}$$

Ces valeurs sont ensuite normalisées par rapport à leur somme. Notons p_i les résultats de la normalisation :

$$p_i = \frac{sim\left(\tilde{T}_X^t, A_i\right) pref_X^t\left(A_i\right)}{\sum_{k=1}^{n} sim\left(\tilde{T}_X^t, A_k\right) pref_X^t\left(A_k\right)} \qquad \text{(eq. V.2.32)}$$

Les p_i vont constituer les poids du type d'émission. Les symboles correspondant à ces poids sont obtenus grâce à la fonction d'association de la TPT (paragraphe 2.2.2) :

$$s_i = symb_X^t\left(A_i\right) \qquad \text{(eq. V.2.33)}$$

Quant aux symboles de D_X qui ne correspondent à aucun archétype, le poids qui leur est associé est zéro.

Finalement, le type d'émission s'obtient alors en réalisant les substitutions adéquates dans (eq. V.2.27). A noter qu'il est possible de négliger l'influence de la magnitude d'activation lors du calcul du type d'émission. L'équation (eq. V.2.27) se simplifie alors en :

$$T_X^t = f_{T_X}^{(2)}\left(\tilde{T}_X^t\right) \qquad \text{(eq. V.2.34)}$$

Plaçons-nous dans ce cas de figure, et reprenons comme exemple la TPT définie dans le paragraphe 2.2 par les équations (eq. V.2.3) à (eq. V.2.12), à l'exception de la fonction de préférence, définie par :

$$pref_X^t\left(T\right) = \begin{cases} 0,2 & \text{si } T = A_a \\ 0,3 & \text{si } T = A_b \\ 0,3 & \text{si } T = A_c \\ 0,2 & \text{si } T = A_d \\ 0 & \text{sinon} \end{cases} \qquad \text{(eq. V.2.35)}$$

De plus, supposons que le type d'activation $\tilde{T}_X^t$ induise les valeurs de similitudes qui constituent la première ligne du Tableau V.3.1.

$A_\bullet$	A_a	A_b	A_c	A_d
$sim\left(\tilde{T}_X^t, A_\bullet\right)$	0	0,6	0,1	0,3
$sim\left(\tilde{T}_X^t, A_\bullet\right) pref_X^t\left(A_\bullet\right)$	0	0,18	0,03	0,06
$p_\bullet$	0	0,67	0,11	0,22
$symb_X^t\left(A_\bullet\right)$	*inter*	*oblig*	*avert*	*indic*

Tableau V.3.1 : détail du calcul d'un type d'émission.

Le Tableau V.3.1 décrit également la suite des calculs intermédiaires permettant d'aboutir au type d'émission :

$$T_X^t = \{(avert;0,11);(indic;0,22);(oblig;0,67)\} \quad \text{(eq. V.2.36)}$$

3.2.3. Magnitude d'émission

A la différence de la fonction utilisée pour le calcul de la magnitude d'activation, f_{M_X} est appliquée à une seule magnitude (la magnitude d'activation), et cette fois le type d'activation peut intervenir dans le calcul :

$$M_X^t = f_{M_X}^{(1)}\left(simax_X^t\right) f_{M_X}^{(2)}\left(\tilde{M}_X^t, u_X^t\right) \quad \text{(eq. V.2.37)}$$

La fonction $f_{M_X}^{(1)}$ représente l'hypothèse, faite au chapitre précédent, qu'un type d'activation qui n'a pas suffisamment de sens pour le nœud engendrera une magnitude d'émission faible. En d'autres termes, moins le type d'activation est reconnu, plus la magnitude d'émission sera diminuée. Les mesures de similitude (eq. V.2.30) utilisées dans le calcul du type d'émission constituent le meilleur indice de cette reconnaissance. Soit $simax_X^t$ le maximum de ces valeurs :

$$simax_X^t = \max_{A \in Arch_X^t}\left(sim\left(\tilde{T}_X^t, A\right)\right) \quad \text{(eq. V.2.38)}$$

On peut considérer que si $simax_X^t$ ne dépasse pas un certain seuil, le nœud n'est pas en position d'émettre. Ceci traduit le fait que le type d'activation n'est suffisamment proche d'aucun archétype pour pouvoir considérer qu'il sera reconnu. La fonction $f_{M_X}^{(1)}$ joue un rôle de modulation, analogue à celui de la fonction $f_{T_X}^{(1)}$ dans le calcul du type d'émission (eq. V.2.27). Elle doit être définie de $[0,1]$ dans $[0,1]$, et être croissante. Ceci permet d'exprimer le fait que moins le type d'activation est reconnu, plus la magnitude doit être diminuée. Là encore, une fonction sigmoïdale peut s'avérer utile.

La fonction $f_{M_X}^{(2)}$ s'applique à la magnitude d'activation, elle doit donc être définie sur les réels. Elle peut être linéaire ou non-linéaire, à l'instar de la fonction permettant de calculer la magnitude d'activation.

A noter qu'il est également possible de négliger l'impact de la reconnaissance du type d'activation sur la magnitude d'émission. L'expression (eq. V.2.37) est alors simplifiée en :

$$M_X^t = f_{M_X}^{(2)}\left(\tilde{M}_X^t, u_X^t\right) \qquad \text{(eq. V.2.39)}$$

3.3. Conclusion

Le Tableau V.3.2 récapitule les différentes formules principales permettant de calculer les valeurs d'activation et d'émission, dans le cas général.

valeur	formule	équation
type d'activation	$\tilde{T}_X^t = \tilde{f}_{M_X}\left(T_{Y_1}^{t-\partial_1}, \ldots, T_{Y_n}^{t-\partial_n}, \tilde{T}_X^{t-1}\right)$	(eq. V.2.15)
magnitude d'activation	$\tilde{M}_X^t = \tilde{f}_{M_X}\left(M_{Y_1}^{t-\partial_1}, \ldots, M_{Y_n}^{t-\partial_n}, \tilde{M}_X^{t-1}, \tilde{u}_X^t\right)$	(eq. V.2.16)
type d'émission	$T_X^t = \oplus\left(\left(bruit_X^t; 1 - f_{T_X}^{(1)}\left(\overline{\tilde{M}_X^t}\right)\right); \left(f_{T_X}^{(2)}\left(\tilde{T}_X^t\right); f_{T_X}^{(1)}\left(\overline{\tilde{M}_X^t}\right)\right)\right)$	(eq. V.2.27)
magnitude d'émission	$M_X^t = f_{M_X}^{(1)}\left(simax_X^t\right) f_{M_X}^{(2)}\left(\tilde{M}_X^t, u_X^t\right)$	(eq. V.2.37)

Tableau V.3.2 : récapitulatif des fonctions de traitement de l'information.

Soulignons que dans certains cas particuliers, il n'est pas nécessaire qu'un nœud du réseau dynamique traite les deux composantes de l'information. Par exemple, dans un nœud modélisant une population de neurones GABA, il n'est pas forcément utile de définir un traitement sur le type, si on suppose que seule la nature quantitative (puissance de l'inhibition) de l'information importe. Formellement, dans ces cas là, la composante de l'information dont on ne veut pas tenir compte est ignorée parmi les valeurs reçues par le nœud dynamique. Le calcul de l'activation et de l'émission se fait également sans cette composante. Au niveau de l'émission, cela signifie que l'on utilise les fonctions d'émission spécifiques (eq. V.2.34) et (eq. V.2.39) (ces fonctions correspondent au cas où l'on souhaite ignorer les influences croisées des magnitude et type d'activation sur les type et magnitude d'émission). De plus, le nœud n'émet pas la composante que l'on souhaite ignorer.

4. MECANISMES D'APPRENTISSAGE

Les mécanismes d'apprentissage reposent uniquement sur des modifications apportées à la TPT. Dans le chapitre précédent, nous avons décrit qu'il existait quatre mécanismes distincts : renforcement, introduction, oubli et glissement. Les trois premiers s'appuient exclusivement sur des modifications de la fonction de préférence, alors que le dernier correspond à des changements apportés à la nature même des archétypes.

4.1. Renforcement, introduction et oubli

Le *mécanisme de renforcement* consiste à modifier les préférences des archétypes en fonction du type et de la magnitude d'activation. Les archétypes proches du type d'activation voient leur préférence légèrement augmenter, alors que les autres voient la leur diminuer. Le *mécanisme d'introduction* consiste à rajouter un archétype dans la TPT, tandis que celui *d'oubli* supprime des archétypes. Dans les deux cas, une mise à jour de la fonction de préférence est également nécessaire. Ces trois mécanismes sont donc extrêmement imbriqués, et c'est pour cette raison qu'ils vont être décrits ensemble.

Considérons que la TPT d'un nœud X possède n archétypes $A_1,\ldots,A_n$. On note $\Delta pref_X^t\left(A_i\right)$ la variation apportée à la préférence de l'archétype A_i :

$$\Delta pref_X^t\left(A_i\right) = cr_X\, f_{TPT_X}^{(1)}\left(\overline{\tilde{M}_X^t}\right) sim\left(\tilde{T}_X^t, A_i\right) \qquad \text{(eq. V.3.1)}$$

où cr_X est un paramètre constant de X appelé *coefficient de renforcement*. Utiliser $sim\left(\tilde{T}_X^t, A_i\right)$ en facteur permet d'exprimer le fait que plus l'archétype A_i est proche du type d'activation, plus sa préférence augmente. La fonction $f_{TPT_X}^{(1)}$, appliquée à la moyenne de la magnitude, permet d'obtenir un coefficient traduisant le fait que la variation de préférence dépend également du niveau d'activation. La fonction doit être croissante et à valeur de $\mathbb{R}$ dans $[0,1]$, elle est comparable, dans son rôle et ses propriétés, à la fonction $f_{T_X}^{(1)}$ utilisée pour calculer le type d'émission dans l'équation (eq. V.2.27). Les variations apportées aux préférences sont forcément positives, mais un mécanisme

de normalisation appliqué à la fin du processus d'apprentissage permet d'éviter un accroissement permanent de toutes les préférences (c.f. (eq. V.3.14)).

Le type d'activation constitue lui-même un nouvel archétype potentiel, sous certaines conditions. C'est le cas quand (1) la magnitude d'activation est suffisamment élevée et (2) le type d'activation n'est suffisamment proche d'aucun archétype pour pouvoir lui être assimilé. Il est donc nécessaire de calculer une variation de préférence également pour le type d'activation, qui permettra d'implémenter ce *mécanisme d'introduction* :

$$\Delta pref_X^t\left(\tilde{T}_X^t\right) = cr_X\, f_{TPT_X}^{(1)}\left(\overline{\tilde{M}_X^t}\right) f_{TPT_X}^{(2)}\left(simax_X^t\right) \qquad \text{(eq. V.3.2)}$$

L'utilisation de la moyenne de la magnitude d'activation comme argument de la fonction $f_{TPT_X}^{(1)}$ représente la première condition (1). La fonction $f_{TPT_X}^{(2)}$ est appliquée à la similitude maximale calculée précédemment (eq. V.2.38), elle représente la deuxième condition (2). Elle doit être définie de $[0,1]$ dans $[0,1]$, et être décroissante pour exprimer le fait que moins le type d'activation est reconnu, plus il est susceptible de devenir un nouvel archétype de la TPT. Les notions de magnitude *suffisamment* élevée et de type *suffisamment* proche exprimées dans (1) et (2) sont représentées par le mécanisme d'oubli, décrit ci-après.

Pour tout autre type T défini sur le domaine $\tilde{D}_X$ du type d'activation, la variation est bien évidemment nulle :

$$\Delta pref_X^t\left(T\right) = 0 \qquad \text{(eq. V.3.3)}$$

En appliquant ces variations de préférence aux préférences existant à l'instant t, on obtient une première version de ce que seront les préférences à l'instant $t+1$. Ce résultat est intermédiaire et n'est pas normalisé (la somme des préférences n'est plus égale à 1), c'est pourquoi il est noté $pref_1$:

$$\forall T \in \mathbb{T}_{\tilde{D}_X} :$$
$$pref_1\left(T\right) = \begin{cases} pref_X^t\left(T\right) + \Delta pref_X^t\left(T\right) & \text{si } pref_X^t\left(T\right) + \Delta pref_X^t\left(T\right) > so_X \\ 0 & \text{sinon} \end{cases} \qquad \text{(eq. V.3.4)}$$

Cette opération correspond également au *mécanisme d'oubli*, puisque les préférences inférieures à un certain *seuil d'oubli*, noté so_X , sont ramenées à zéro. Si la préférence du type d'activation est en dessous du seuil, il ne sera donc pas introduit dans la TPT. A partir de cette fonction $pref_1$, il est possible de calculer un nouvel ensemble des archétypes :

$$Arch_1 = \left\{ T : T \in \mathbb{T}_{\tilde{D}_X} \wedge pref_1(T) > 0 \right\} \quad \text{(eq. V.3.5)}$$

Cet ensemble constitue également un résultat intermédiaire. Par rapport à l'ensemble initial $Arch_X^t = \{A_1, \ldots, A_n\}$, le nouvel ensemble $Arch_1 = \{B_1, \ldots, B_m\}$ est susceptible de contenir le type d'activation comme nouvel archétype (mécanisme d'introduction), mais il est également possible que certains archétypes aient disparu (mécanisme d'oubli).

4.2. Glissement et fusion

Le *mécanisme de glissement* consiste à modifier la nature de tous les archétypes de façon à les faire tendre vers le type d'activation. L'importance de la modification est déterminée par un *coefficient de glissement* cg_X , et dépend également de la magnitude d'activation et de la similitude entre le type d'activation et l'archétype considéré. Le résultat du glissement d'un archétype B_i de $Arch_1$ est le type C_i également défini sur $\tilde{D}_X$, qui est le résultat de la combinaison linéaire suivante :

$$C_i = \oplus\left(\left(B_i, 1 - cg_X f^{(1)}_{TPT_X}\left(\overline{\tilde{M}^t_X}\right) sim\left(\tilde{T}^t_X, B_i\right)\right), \left(\tilde{T}^t_X, cg_X f^{(1)}_{TPT_X}\left(\overline{\tilde{M}^t_X}\right) sim\left(\tilde{T}^t_X, B_i\right)\right)\right) \quad \text{(eq. V.3.6)}$$

L'utilisation de la similitude permet de faire varier en priorité les archétypes proches du type d'activation, tandis que la magnitude a le même rôle que dans les mécanismes de renforcement et d'introduction. Il est évident que si le type d'activation a été incorporé dans l'ensemble des archétypes par le mécanisme d'introduction, il ne sera pas modifié par cette opération (puisqu'il sera combiné à lui-même).

Les C_i vont constituer les nouveaux archétypes de la TPT. La préférence (non définitive) associée à un C_i est celle du B_i correspondant avant le glissement, c'est-à-dire $pref_1(B_i)$. Celle associée

aux autres types (y compris les B) est nulle. On obtient donc un nouveau résultat intermédiaire en ce qui concerne la fonction de préférence :

$$\forall T \in \mathbb{T}_{\tilde{D}_X} :$$
$$pref_2(T) = \begin{cases} pref_1(B_i) & \text{si } T = C_i \\ 0 & \text{sinon} \end{cases} \qquad \text{(eq. V.3.7)}$$

Ce résultat permet de déterminer un nouvel ensemble des archétypes, lui aussi intermédiaire, qui ne contient en fait que les C_i :

$$Arch_2 = \left\{ T : T \in \mathbb{T}_{\tilde{D}_X} \wedge pref_2(T) > 0 \right\} = \{C_1, \ldots, C_m\} \qquad \text{(eq. V.3.8)}$$

A noter qu'ici, le nombre d'archétypes ne varie pas, seule leur nature varie. La fonction d'association $symb_2$ constitue également un résultat intermédiaire. Si le type d'activation a été précédemment introduit dans $Arch_1$ en tant que nouvel archétype, un élément d de $\tilde{D}_X$ doit lui être associé, et cet élément ne doit pas déjà être associé à un autre archétype. Pour les autres archétypes, on associe à un C_i le symbole qui était déjà associé, par la fonction $symb_X^t$, à l'archétype B_i correspondant :

$$\forall C_i \in Arch_2 : symb_2(C_i) = \begin{cases} d & \text{si } C_i = \tilde{T}_X^t \\ symb_X^t(B_i) & \text{sinon} \end{cases} \qquad \text{(eq. V.3.9)}$$

A la suite de l'opération de glissement des archétypes, il est possible d'obtenir des types C_i reflétant des répartitions si proches qu'elles pourraient être assimilées à la même information. D'où la nécessité d'introduire un mécanisme de régulation, appelé *mécanisme de fusion*. Il permet de limiter la prolifération de ces archétypes similaires. La fusion consiste, lorsque deux archétypes sont suffisamment similaires, à supprimer le moins préféré, et à donner sa préférence à l'archétype qui est conservé.

On considère que deux archétypes C_i et C_j sont suffisamment similaires quand leur mesure de similitude dépasse un certain *seuil de fusion* sf_X :

$$sim(C_i, C_j) > sf_X \qquad \text{(eq. V.3.10)}$$

Si c'est le cas, en supposant que $pref_2(C_i) > pref_2(C_j)$, les conséquences sur la fonction de préférence, l'ensemble des archétypes et la fonction d'association, sont les suivantes :

$$\forall T \in \mathbb{T}_{\tilde{D}_X} :$$
$$pref_3(T) = \begin{cases} pref_2(C_i) + pref_2(C_j) & \text{si } T = C_i \\ 0 & \text{si } T = C_j \\ pref_2(T) & \text{sinon} \end{cases} \quad \text{(eq. V.3.11)}$$

$$Arch_3 = \left\{ T : T \in \mathbb{T}_{\tilde{D}_X} \wedge pref_3(T) > 0 \right\} = \{D_1, \ldots, D_k\} \quad \text{(eq. V.3.12)}$$

$$\forall D_i \in Arch_3 : symb_3(D_i) = symb_2(D_i) \quad \text{(eq. V.3.13)}$$

Ce filtrage de la TPT doit être renouvelé jusqu'à ce qu'il n'y ait plus deux archétypes assimilables. Le traitement réalisé revient à dire que l'on fusionne les parties du nœud qui traitent la même catégorie d'information. Finalement, les derniers résultats intermédiaires permettent d'aboutir à la TPT qui sera utilisée par le nœud à l'instant suivant. Pour cela, la préférence doit être normalisée afin de respecter la contrainte (eq. V.2.1) :

$$\forall T \in \mathbb{T}_{\tilde{D}_X} :$$
$$pref_X^{t+1}(T) = \frac{pref_3(T)}{\sum_{U \in \mathbb{T}_{\tilde{D}_X}} pref_3(U)} \quad \text{(eq. V.3.14)}$$

Puis il est possible de définir l'ensemble des archétypes et la fonction d'association :

$$Arch_X^{t+1} = \left\{ T : T \in \mathbb{T}_{\tilde{D}_X} \wedge pref_X^{t+1}(T) > 0 \right\} \quad \text{(eq. V.3.15)}$$

$$\forall A \in Arch_X^{t+1} : symb_X^{t+1}(A) = symb_3(A) \quad \text{(eq. V.3.16)}$$

4.3. Conclusion

Les différentes étapes de notre processus d'apprentissage sont résumées dans le Tableau V.4.1.

mécanisme	principe	équation
renforcement	la préférence de chaque archétype est modifiée en fonction de sa similitude avec le type d'activation, et en fonction de la magnitude d'activation	(eq. V.3.1)
introduction	une préférence est calculée pour le type d'activation	(eq. V.3.2)
oubli	tous les archétypes de préférence inférieure à un certain seuil sont supprimés de la TPT si c'est le cas pour le type d'activation, il n'est pas introduit dans la TPT	(eq. V.3.3) à (eq. V.3.5)
glissement	la nature de chaque archétype est modifiée de façon à le faire tendre vers le type d'activation, en fonction de la similitude entre l'archétype et le type d'activation, et en fonction de la magnitude d'activation	(eq. V.3.6) à (eq. V.3.9)
fusion	les archétypes dont la similitude dépasse un certain seuil sont réunis en un seul archétype dont la préférence est la somme des préférences des archétypes fusionnés	(eq. V.3.10) à (eq. V.3.13)
normalisation	les préférences sont normalisées, et la TPT est finalisée	(eq. V.3.14) à (eq. V.3.16)

Tableau V.4.1 : récapitulatif des mécanismes d'apprentissage.

4.4. Exemple

Dans cet exemple, nous utilisons la TPT définie en exemple dans la partie 2.2, à l'exception du troisième archétype qui est modifié en $A_c = \{((rouge;tri);0,1);((bleu;circ);0,9)\}$ (ceci afin de pouvoir illustrer le fonctionnement du mécanisme de fusion), et de la préférence du premier archétype, qui est diminuée et prend la valeur $0,01$ (ceci afin d'illustrer le mécanisme d'oubli), le autres préférences étant révisées en conséquence. Finalement, on obtient la TPT représentée dans le Tableau V.4.2.

$A_\bullet$	détail	$symb_X^t(A_\bullet)$	$pref_X^t(A_\bullet)$
A_a	$\{((rouge;circ);1)\}$	*inter*	0,01
A_b	$\{((bleu;circ);1)\}$	*oblig*	0,35
A_c	$\begin{Bmatrix}((rouge;tri);0,1);\\((bleu;circ);0,9)\end{Bmatrix}$	*avert*	0,30
A_d	$\{((blanc;rect);1)\}$	*indic*	0,34

Tableau V.4.2 : TPT initiale.

Le coefficient de renforcement a pour valeur $cr_X = 0,05$, le coefficient de glissement vaut $cg_X = 0,05$, le seuil d'oubli est $so_X = 0,02$ et le seuil de fusion $sf_X = 0,9$. La fonction $f_{TPT_X}^{(1)}$ (appliquée à une magnitude) est une sigmoïde de paramètres $a = 14$ et $b = 0,5$. La fonction et $f_{TPT_X}^{(2)}$ est définie par :

$$f_{TPT_X}^{(2)}\left(simax_X^t\right) = 1 - simax_X^t \qquad \text{(eq. V.3.17)}$$

Enfin, la fonction de similitude que nous employons est une sigmoïde de paramètres $a = -14$ et $b = 0,5$. Supposons que le type d'activation soit :

$$\tilde{T}_X^t = \{((blanc;rect);0,1);((vert;rect);0,9)\} \qquad \text{(eq. V.3.18)}$$

et que la moyenne de la magnitude d'activation soit $\overline{\tilde{M}_X^t} = 1$. On calcule les variations de préférence des archétypes d'après l'équation (eq. V.3.1). Les résultats sont donnés dans le Tableau V.4.3.a. Nous pouvons alors calculer la variation de préférence pour le type d'activation (candidat à l'introduction dans la TPT) grâce à (eq. V.3.2) :

$$\Delta pref_X^t\left(\tilde{T}_X^t\right) = 0,049908936 \qquad \text{(eq. V.3.19)}$$

En appliquant (eq. V.3.4), on obtient les résultats du Tableau V.4.3.b.

$A_\bullet$	$sim\left(\tilde{T}_X^t, A_\bullet\right)$	$\Delta pref_X^t\left(A_\bullet\right)$
A_a	0,000911051	0,000045511
A_b	0,000911051	0,000045511
A_c	0,000911051	0,000045511
A_d	0,00368424	0,000184044

type/archétype	$pref_1$
A_a	0,010045511
A_b	0,350045511
A_c	0,300045511
A_d	0,340184044
$\tilde{T}_X^t$	0,049908936

Tableau V.4.3 : renforcement, introduction et oubli.
(a) : similitudes entre le type d'activation et chaque archétype, et variations de préférence calculées avec (eq. V.3.1). (b) : nouvelles préférences calculées avec (eq. V.3.4)

La préférence du premier archétype est sous le seuil d'oubli so_X, par conséquent il doit être supprimé de la TPT. En revanche, celle du type d'activation dépasse ce seuil, donc celui-ci est introduit dans la TPT. On a donc d'après (eq. V.3.5) (les archétypes sont renommés en $B_\bullet$) :

$$Arch_1 = \left\{A_b, A_c, A_d, \tilde{T}_X^t\right\} = \left\{B_b, B_c, B_d, B_e\right\} \quad \text{(eq. V.3.20)}$$

L'application de (eq. V.3.6), permet de faire glisser les archétypes de $Arch_1$, et d'obtenir les résultats du Tableau V.4.4.

D'après (eq. V.3.7), on calcule $pref_2$ en associant à chaque $C_\bullet$ la préférence $pref_1$ du $B_\bullet$ d'indice correspondant. De plus, d'après (eq. V.3.8), on a :

$$Arch_2 = \left\{C_b, C_c, C_d, C_e\right\} \quad \text{(eq. V.3.21)}$$

avant le glissement		après le glissement	
$B_\bullet$	détail	$C_\bullet$	détail
B_b	$\{((bleu;circ);1)\}$	C_b	$\begin{Bmatrix}((bleu;circ);0{,}999954489);\\((blanc;rect);0{,}000004551);\\((vert;rect);0{,}000040960)\end{Bmatrix}$
B_c	$\begin{Bmatrix}((rouge;tri);0{,}1);\\((bleu;circ);0{,}9)\end{Bmatrix}$	C_c	$\begin{Bmatrix}((rouge;tri);0{,}099995449);\\((bleu;circ);0{,}899834360);\\((blanc;rect);0{,}000004551);\\((vert;rect);0{,}000040960)\end{Bmatrix}$
B_d	$\{((blanc;rect);1)\}$	C_d	$\begin{Bmatrix}((blanc;rect);0{,}999834360);\\((vert;rect);0{,}000165640)\end{Bmatrix}$
B_e	$\begin{Bmatrix}((blanc;rect);0{,}1);\\((vert;rect);0{,}9)\end{Bmatrix}$	C_e	$\begin{Bmatrix}((blanc;rect);0{,}1);\\((vert;rect);0{,}9)\end{Bmatrix}$

Tableau V.4.4 : glissement.

$C_\bullet$	$symb_2(C_\bullet)$
C_b	*oblig*
C_c	*avert*
C_d	*indic*
C_e	*direct*

Tableau V.4.5 : fonction d'association.

En utilisant (eq. V.3.9), on affecte un nouveau symbole appelé ***direct*** à l'archétype nouvellement introduit C_e (Tableau V.4.5).

Afin de pouvoir effectuer les éventuelles fusions, nous devons calculer les similitudes des archétypes de $Arch_2$ entre eux (Tableau V.4.6).

$C_\bullet$	C_b	C_c	C_d	C_e
C_b	-	0,99631576	0,000911051	0,000911051
C_c	0,99631576	-	0,000911051	0,000911051
C_d	0,000911051	0,000911051	-	0,00368424
C_e	0,000911051	0,000911051	0,00368424	-

Tableau V.4.6 : mesures de similitude entre les archétypes de $Arch_2$.

D'après (eq. V.3.10), les archétypes C_b et C_c sont trop similaires pour être conservés tous les deux. Des deux archétypes, c'est C_b qui a la préférence la plus élevée, c'est donc lui qui est conservé, alors que C_c est supprimé de la TPT. D'après (eq. V.3.11), sa préférence est rajoutée à celle de C_b :

$$pref_3(C_c) = 0 \qquad \text{(eq. V.3.22)}$$

$$pref_3(C_b) = 0{,}650091022 \qquad \text{(eq. V.3.23)}$$

Les préférences des autres archétypes ne changent pas. On a donc maintenant l'ensemble d'archétypes suivant :

$$Arch_2 = \{C_b, C_d, C_e\} \qquad \text{(eq. V.3.24)}$$

Aucune autre fusion ne s'avérant nécessaire, on peut maintenant normaliser les préférences en utilisant (eq. V.3.14). On obtient ainsi la TPT définitive donnée dans le Tableau V.4.7.

Sur les 4 archétypes de départ, un (A_a) a été oublié, un autre (A_c) a disparu pour cause de fusion (avec A_b), et le type d'activation a été introduit en tant que nouvel archétype dans la TPT.

$C_\bullet$	détail	$symb_X^{t+1}(C_\bullet)$	$pref_X^{t+1}(C_\bullet)$
C_b	$\left\{\begin{array}{l}((bleu;circ);0,999);\\((blanc;rect);0,0001);\\((vert;rect);0,0009)\end{array}\right\}$	*oblig*	0,624976947
C_d	$\left\{\begin{array}{l}((blanc;rect);0,9982);\\((vert;rect);0,0018)\end{array}\right\}$	*indic*	0,327042180
C_e	$\left\{\begin{array}{l}((blanc;rect);0,1);\\((vert;rect);0,9)\end{array}\right\}$	*direct*	0,047980873

Tableau V.4.7 : TPT finale.

5. DISCUSSION SUR LE FORMALISME

5.1. Propriétés du formalisme

Dans cette partie, nous détaillons certains traits caractéristiques de RAGE, en ce qui concerne l'information manipulée, mais également les mécanismes de traitement de cette information. En effet, le choix des paramètres et des fonctions utilisés lors de la propagation et de l'apprentissage dans un nœud est déterminant pour définir le comportement de ce dernier.

5.1.1. Information

La magnitude pouvant prendre des valeurs négatives ou positives, il est nécessaire d'expliquer son interprétation dans le cadre de la modélisation cérébrale. En fait, une magnitude nulle ne représente pas une absence d'activation, mais l'activation cérébrale au repos. Par conséquent, une valeur positive représente une activation supérieure à l'activation au repos (hyperactivation), tandis qu'une valeur négative représente une activation inférieure (hypoactivation).

L'activité représentée par la magnitude représente une activité *neuronale*. En appliquant certaines transformations à la magnitude, il est possible de la confronter aux mesures d'imagerie, qui sont elles-mêmes des mesures indirectes de l'activité neuronale. Par exemple, dans un réseau fonctionnel modélisant une aire cérébrale, un nœud fonctionnel peut être dédié à la simulation de la variation du débit sanguin cérébral de l'aire, dans le but de comparer les valeurs issues de ce nœud à des valeurs issues d'expériences réelles en TEP ou IRMf (c.f. les applications du chapitre VI). De plus, notre modèle permet de décomposer l'activité d'une population neuronale en une valeur d'activation et une valeur d'émission. Ainsi, il nous est possible de distinguer l'activation infructueuse (qui n'aboutit pas à une émission de la part d'une population neuronale) et l'activation reflétant un réel traitement de l'information. Ceci est d'une grande utilité en ce qui concerne l'interprétation de données issues de la neuroimagerie, notamment de techniques tomographiques.

La magnitude est une variable aléatoire, alors que ce n'est pas le cas du type. En effet, la magnitude constitue la seule composante de l'information qui peut être comparée aux données statistiques issues de la neuroimagerie, ce qui explique l'emploi d'une représentation incertaine. Au contraire, le type est utile lors de la simulation, mais n'est pas comparable aux mesures d'activation cérébrale réelles. Dès lors, il est inutile de le représenter de façon incertaine, ce qui alourdirait le traitement.

5.1.2. Fonctions

Dans RAGE, on peut distinguer trois types de fonctions implémentant un traitement défini par l'utilisateur. Tout d'abord, la fonction de similitude, qui est utilisée à la fois lors de la propagation (eq. V.2.29),(eq. V.2.38), et de l'apprentissage (eq. V.3.1),(eq. V.3.2),(eq. V.3.6),(eq. V.3.10) : elle constitue un cas à part. Puis, les fonctions de propagation, qui offrent une grande liberté d'expression. Il s'agit des fonctions $\tilde{f}_{M_X}$ (eq. V.2.14), $\tilde{f}_{T_X}$ (eq. V.2.15), et $f^{(2)}_{M_X}$ (eq. V.2.37), présentes lors de l'activation et de l'émission. Enfin, les fonctions de modulation, qui permettent de modéliser les influences croisées entre types et magnitudes, qui sont utilisées lors de l'émission et de l'apprentissage. Les fonctions concernées sont $f^{(1)}_{M_X}$ (eq. V.2.37), $f^{(1)}_{T_X}$ (eq. V.2.27), $f^{(1)}_{TPT_X}$ (eq. V.3.1) , (eq. V.3.2) et $f^{(2)}_{TPT_X}$ (eq. V.3.2).

a. Fonction de similitude

A l'utilisation, nous pouvons dire que les fonctions de similitude linéaire et cosinus présentent des comportements similaires. Globalement, les résultats calculés en les utilisant sont caractérisés par

une certaines sensibilité aux valeurs de départ. Par exemple, lors du calcul du type d'émission, la distribution de symboles qui définit ce type est relativement sensible à de faibles variations dans le type d'activation. Ceci est du au fait que les valeurs de similitude calculées grâce à ces deux fonctions sont réparties relativement régulièrement sur $[0,1]$.

En revanche, la fonction de similitude sigmoïdale a pour propriété d'écraser les valeurs de similitude, ce qui provoque une concentration de ces valeurs aux environs de 1 et de 0. De ce fait, les résultats sont beaucoup moins sensibles aux faibles variations. Ils sont également plus tranchés. Pour reprendre l'exemple du type d'émission, la distribution de symboles montre une plus grande disparité, c'est-à-dire qu'on voit plus facilement apparaître un ou plusieurs symboles dominants (de poids fort), que ce n'est le cas en utilisant les fonctions de similitude linéaire et cosinus. En fait, plus la pente de la sigmoïde est importante, plus les différences sont marquées dans la distribution du type d'émission, ce qui se traduit par la domination presque exclusive d'un symbole sur les autres. Le centre de la sigmoïde détermine une distance pivot, au-delà de laquelle deux types sont considérés comme « plutôt semblables ». En l'augmentant, on rend la fonction moins restrictive : les types seront considérés comme plutôt semblables bien que leur distance soit élevée. Au contraire, une valeur faible sera plus restrictive.

b. Fonctions de propagation

Dans les fonctions permettant le calcul des valeurs d'activation, c'est-à-dire $\tilde{f}_{M_X}$ et $\tilde{f}_{T_X}$, on cherche à réaliser l'intégration des entrées. Pour cette raison, bien que sa forme ne soit pas contrainte par le formalisme, $\tilde{f}_{M_X}$ se présente en général sous la forme d'une somme pondérée des différentes entrées. Les coefficients associés aux entrées permettent de déterminer leurs importances relatives. Lorsqu'un des parents du nœud dynamique considéré exerce un contrôle plutôt qu'une simple transmission d'information (c.f. contrainte de non-linéarité, chapitre II.4.2.3), on utilisera sa valeur sous la forme d'un coefficient multiplicatif. En ce qui concerne le calcul du type d'activation via $\tilde{f}_{T_X}$, la fonction est contrainte par les règles portant sur les combinaisons de types. Il faut noter que le formalisme n'interdit pas de négliger les valeurs en provenance de certains nœuds, aussi bien en ce qui concerne la magnitude que le type.

Pour le calcul de la magnitude d'émission, la fonction $f_{M_X}^{(2)}$ représente le traitement interne exercé sur la valeur d'activation. Il peut s'agir d'appliquer une fonction linéaire, par exemple baisser

systématiquement la magnitude d'activation, ou bien un traitement non-linéaire semblable à celui utilisé dans le neurone formel, qui consiste à n'émettre que si la valeur d'activation est suffisamment élevée. En ce qui concerne le calcul du type d'émission réalisé par $f_{T_X}^{(2)}$, rappelons qu'il est automatiquement défini par la TPT.

c. Fonctions de modulation

Les fonctions $f_{T_X}^{(1)}$ et $f_{TPT_X}^{(1)}$ s'appliquent à des magnitudes. Elles modélisent les influences respectives de la magnitude d'activation sur le type d'émission et sur la TPT. Leur rôle est de transformer la magnitude en une valeur normalisée comprise entre 0 et 1, pouvant jouer le rôle d'un coefficient qui viendra moduler le calcul du type d'émission et de la TPT. Une fonction sigmoïdale semble parfaitement adaptée dans ce cas-là. Comme précédemment (c.f. fonction de similitude), la modification de la pente de la sigmoïde se traduit par des changements plus radicaux. Par exemple, prenons le calcul du type d'émission, qui repose sur l'introduction de bruit en fonction du résultat de $f_{T_X}^{(1)}$. Si on utilise une sigmoïde à forte pente, on observe soit un type d'activation extrêmement bruité (cas où la magnitude est faible, et où le résultat de $f_{T_X}^{(1)}$ est donc proche de 0), soit un type d'activation sans bruit (le résultat de $f_{T_X}^{(1)}$ est proche de 1). Le déplacement du centre de la sigmoïde permet de déterminer pour quelle magnitude on passe d'un type majoritairement bruité à un type minoritairement bruité.

Les fonctions $f_{M_X}^{(1)}$ et $f_{TPT_X}^{(2)}$ s'appliquent à des similitudes. Elles modélisent respectivement l'influence du type d'activation sur la magnitude d'émission et sur la TPT. Contrairement au cas précédent, le rôle de ces fonctions n'est pas de normaliser les valeurs, puisqu'il s'agit de mesures de similitudes, qui sont par définition déjà comprises entre 0 et 1. Elles peuvent par contre être utiles pour contraster les mesures de similitudes. Ceci est superflu dans le cas où la fonction de similitude est sigmoïdale, on pourra alors utiliser une fonction linéaire. Mais ceci peut se révéler intéressant si l'on souhaite utiliser une fonction linéaire ou cosinus. On pourra alors utiliser, par exemple, une fonction sigmoïdale, et se ramener ainsi à des valeurs proches de ce que donnerait directement une fonction de similitude sigmoïdale (à noter que la fonction $f_{TPT_X}^{(2)}$ doit posséder en plus la propriété d'être décroissante). En jouant sur le centre de symétrie de la sigmoïde, il est également possible d'utiliser cette fonction pour implémenter un mécanisme de seuillage.

5.1.3. Paramètres

Il est possible de partager les paramètres d'apprentissage en deux catégories : d'un côté le coefficient de renforcement et seuil d'oubli, qui influencent les préférences des archétypes, et de l'autre le coefficient de glissement et le seuil fusion, qui sont plutôt dédiés à la modification de la nature des archétypes.

a. Coefficient de renforcement et seuil d'oubli

Le coefficient de renforcement cr_X intervient dans le calcul des variations des préférences, via l'équation (eq. V.3.1). Sa valeur détermine la vitesse de variation des préférences, et, influence donc la stabilité de la TPT (du moins en ce qui concerne les préférences). Un coefficient fort entraîne de l'instabilité car les préférences changent beaucoup à chaque instant. Dans le cas où le nœud est soumis à des entrées régulières (i.e. on a toujours le même type d'activation, ou presque) un archétype peut rapidement écraser les autres, qui vont finir par être oubliés. Au contraire, un coefficient faible est synonyme de TPT quasi-statique, et il faut recevoir de nombreuses fois la même entrée pour voir une évolution notable des préférences. Si les entrées sont irrégulières, les préférences de la TPT ne varient quasiment pas. De plus, en examinant l'équation (eq. V.3.1), on s'aperçoit que si le coefficient de renforcement est fort, il amoindrit, dans cette équation, l'importance de la magnitude d'activation. En effet, les écarts de magnitude deviennent négligeables par rapport au coefficient.

Le seuil d'oubli so_X est utilisé pour contrôler l'accroissement du nombre d'archétypes dans la TPT (eq. V.3.4). En utilisant la contrainte de sommation sur les préférences (eq. V.2.1), le nombre d'archétypes maximal n_{max} pour un seuil so_X peut être déterminé :

$$n_{max} \approx \left\lfloor \frac{1}{so_X} \right\rfloor \qquad \text{(eq. V.4.1)}$$

où la notation $\lfloor z \rfloor$ désigne la partie entière de z. Un seuil élevé permet de ne conserver dans la TPT qu'un petit nombre d'archétypes de préférences relativement fortes. Au contraire, un seuil bas n'autorise l'oubli que pour les archétypes de préférences faibles, et, de ce fait, le nombre d'archétypes dans la TPT est élevé. Le nombre d'archétypes est à mettre en rapport avec les types qu'ils représentent : un nombre réduit d'archétypes signifie que chaque archétype représente un ensemble assez large de types, alors que dans une TPT possédant de nombreux archétypes, un

archétype correspond à certains types bien précis. Dans le premier cas, on assimile un type d'activation donné à un certain archétype, même s'il est relativement différent (pas trop, tout de même, sinon un nouvel archétype serait introduit !) pour la simple raison qu'il n'existe pas d'autre archétype plus proche. Dans le deuxième cas, on peut imaginer qu'il existe un archétype plus proche. Dans le premier cas, on exprime donc une catégorisation grossière, alors que dans le second, il s'agit d'une catégorisation fine.

Le seuil d'oubli et le coefficient de renforcement interviennent également dans le mécanisme d'introduction, à travers les équations (eq. V.3.2) et (eq. V.3.4). En étudiant (eq. V.3.2), on voit qu'un fort coefficient de renforcement facilite l'introduction d'un nouvel archétype. En effet, l'importance (dans le produit) des variations de la magnitude d'activation et, dans une moindre mesure, de $simax_X^t$, sont négligeables face à un coefficient de renforcement très fort. Cela signifie qu'un nouvel archétype peut être introduit même si la magnitude d'activation est relativement faible, ou si la similitude maximale ($simax_X^t$) est élevée. A l'inverse, dans le cas d'un coefficient de renforcement très faible, on court le risque de voir le résultat du produit (eq. V.3.2) être inférieur au seuil d'oubli, même dans le cas d'une forte magnitude d'activation et d'une faible similitude maximale. Cela se traduirait par un oubli de l'archétype que l'on tente d'introduire. En d'autres termes, il serait impossible d'introduire un nouvel archétype dans la TPT. On en déduit que pour que le mécanisme d'introduction ait un sens, il faut avoir :

$$cr_X \geq so_X \qquad \text{(eq. V.4.2)}$$

De plus, il est évident qu'un type nouvellement introduit dans la TPT ne doit y rester que si le même type est présenté ensuite un nombre suffisant de fois : un type inconnu, présenté peu de fois, constitue un accident dont la TPT ne doit pas garder trace longtemps. Par exemple, en prenant cr_X et so_X tels que $cr_X = so_X$, on ne laisse que le minimum de chance à un archétype d'être introduit et conservé dans la TPT : si le type d'activation à l'instant suivant n'est pas plus proche du nouvel archétype que des autres, cet archétype est supprimé.

b. Coefficient de glissement et seuil de fusion

Le coefficient de glissement cg_X intervient dans le calcul des variations des poids des archétypes (eq. V.3.6). Le coefficient de glissement possède pour les poids des propriétés similaires à celles du coefficient de renforcement pour les préférences. Il quantifie la variation maximale pouvant être

apportée à la nature d'un archétype. S'il est fort, il a tendance à amoindrir l'impact de la magnitude d'activation et de la similitude dans l'équation (eq. V.3.6). De ce fait, tous les archétypes ont tendance à dériver rapidement vers le type d'activation. Par conséquent, si le type d'activation est toujours plus ou moins le même, on aboutit à des archétypes très proches, qui risquent finalement de ne faire plus qu'un à cause du mécanisme de fusion. Dans le cas d'un coefficient faible, les variations apportées ont peu d'impact, et concernent essentiellement les archétypes proches du type d'activation. Il faut recevoir un grand nombre de fois le même type d'activation pour voir une évolution de la TPT, concernant peu d'archétypes.

Le mécanisme de fusion permet de regrouper des archétypes légèrement différents, mais supposés exprimer la même information (eq. V.3.10). Il représente la mesure de similitude entre deux archétypes à partir de laquelle on considère que ces archétypes expriment la même information. On peut considérer que le seuil de fusion définit une sorte de *scope* ou de *portée* des archétypes, c'est-à-dire un ensemble de types englobant l'archétype et les types situés à une certaine distance de celui-ci. Le mécanisme de fusion consiste à faire disparaître l'un des deux archétypes lorsqu'il y a intersection de deux scopes. Par conséquent, un seuil élevé provoquera plus de fusions qu'un seuil faible. Il existe un lien entre les rôles du coefficient de glissement et du seuil de fusion. En effet, le premier quantifie la variation à apporter à un archétype, alors que le second quantifie l'écart minimal entre deux archétypes. De plus, la fusion permet de limiter le glissement des archétypes : si le seuil de fusion est suffisamment élevé, les archétypes n'ont pas l'opportunité de beaucoup glisser avant d'entrer en concurrence avec un autre archétype. Un seuil élevé permet donc d'empêcher la nature d'un archétype d'évoluer énormément (sauf si tous les archétypes évoluent conjointement de la même façon et à la même vitesse). Enfin, le seuil de fusion participe également à la régulation des archétypes nouvellement apparus. En effet, en présence d'un seuil d'oubli relativement bas et d'une forte magnitude d'activation, il possible que le mécanisme d'introduction fasse apparaître dans la TPT un nouvel archétype trop proche d'un archétype déjà existant. Le mécanisme de fusion va empêcher ce phénomène, en fusionnant ces deux archétypes. On ne peut pas dire pour autant que le seuil de fusion limite le nombre d'archétypes comme le fait déjà le seuil d'oubli. Il s'agit plutôt ici de limiter leur concentration, dans le sens où on va empêcher que plusieurs archétypes se centrent sur des informations similaires, ou, en d'autres termes, représentent quasiment la même catégorie.

5.2. Respect des contraintes

La seule utilisation des RBD nous permet de respecter la plupart de nos contraintes de départ : réseau orienté aux nœuds différenciés, temps explicitement représenté et discrétisé, utilisation de fonctions non-linéaires, et enfin, représentation explicite de l'incertitude. Grâce au formalisme que nous avons défini sur cette base, nous remplissons deux contraintes supplémentaires : la manipulation d'une information duale, et l'existence de mécanismes d'apprentissages non-supervisés. De façon plus générale, RAGE permet d'obtenir la plausibilité biologique nécessaire à son application dans un objectif d'interprétation des données de neuroimagerie en termes de traitement de l'information.

5.3. Comparaisons avec d'autres formalismes

En ce qui concerne les propriétés formelles de notre outil de modélisation, la comparaison peut essentiellement être établie avec BioCaEn, qui représente en quelque sorte la base de départ de notre travail, et les réseaux ART, avec lesquels notre approche présentent des similitudes en ce qui concerne les mécanismes d'apprentissage.

5.3.1. BioCaEn

La principale source d'inspiration pour la définition de RAGE est bien entendu BioCaEn, puisque nous nous situons dans le même projet global, et que notre but est d'améliorer ce premier outil de modélisation que constitue BioCaEn. On retrouve ainsi dans BioCaEn le concept d'information duale (type/magnitude), de relations causales entre les populations neuronales, de nœuds fonctionnellement différenciés, de temps représenté explicitement sous forme discrète. En outre, BioCaEn permet de manipuler implicitement des valeurs imprécises, et une certaine forme de non-linéarité peut être introduite dans les relations.

La différence la plus importante est que BioCaEn repose sur les réseaux causaux qualitatifs, alors que nous avons utilisés les RBD. Grâce à ce changement, nous pouvons exprimer des fonctions sans restriction de forme, ce qui n'était pas le cas dans BioCaEn. D'une part, l'intégration des entrées au niveau d'un nœud est contrainte, dans BioCaEn. Chacune de ces influences est, en effet, traitée séparément des autres, avant de les combiner toutes pour calculer une influence globale, au moyen d'une loi de combinaison qui doit elle-même suivre certaines contraintes. RAGE n'est pas

contraint à ce niveau, ce qui permet d'exprimer aussi bien une intégration à la BioCaEn, que d'autres types de traitements. D'autre part, dans BioCaEn, la seule façon d'introduire de la non-linéarité dans une relation est d'utiliser un gain dynamique, qui correspond en fait à la valeur d'une condition évaluée dans une logique multivaluée. On limite donc la non-linéarité à une linéarité avec discontinuité. A l'inverse, les filtres de Kalman non-linéaires, et donc RAGE, permettent une grande liberté dans l'expression des relations non-linéaires. Enfin, l'usage d'un formalisme probabiliste nous fait bénéficier d'une mesure de l'incertitude qui était absente de BioCaEn.

En dehors de ces différences liées aux formalismes de base, il existe plusieurs autres distinctions entre les deux formalismes. Tout d'abord, la définition formelle, aussi bien que l'interprétation des types ont été profondément remaniées. De ce fait, les opérateurs que nous utilisons pour manipuler les types sont également différents. Tout en conservant le concept de TPT, nous l'avons modifié de façon à l'adapter à des mécanismes de propagation différents de ceux de BioCaEn. Ceci comprend entre autres la décomposition de l'activité d'une population neuronale en valeurs d'activation et valeurs d'émissions. Grâce à cette décomposition, nous avons pu introduire un lien entre le calcul du type et de la magnitude, alors que ceux-ci se faisaient complètement en parallèle dans BioCaEn. Les modifications apportées à la TPT nous ont également permis de définir des mécanismes d'apprentissage, alors que BioCaEn n'en implémente aucun.

5.3.2. RNF et réseaux ART

On retrouve dans RAGE certains traits propres aux réseaux de neurones formels en général, et d'autres propres aux réseaux ART en particulier. Par rapport aux premiers, on trouve notamment la décomposition du traitement de l'information. Nos valeurs d'activation et d'émission se rapprochent du traitement en deux phases implémenté par le neurone formel général : intégration des entrées puis décharge. De plus, le rôle de notre coefficient de glissement est équivalent au facteur d'apprentissage α que l'on trouve dans la plupart des règles d'apprentissage des RNF (c.f. chapitre III.1.1.2).

RAGE présente certaines similitudes avec les réseaux ART (c.f. chapitre III.2.2), notamment en ce qui concerne les mécanismes d'apprentissage. Ainsi, la mémoire à long terme implémentée par les réseaux ART est proche de notre TPT. Elle permet au réseau de retenir un certain nombre de configurations d'activation des neurones de la couche d'entrée. On peut rapprocher ces configurations de notre concept d'archétype. Chaque configuration est associée à un neurone attracteur. Lorsque celui-ci émet, c'est que la configuration qui lui est associée a été reconnue. Dans

RAGE, l'attracteur correspond au symbole associé à un archétype : lorsque le type d'émission contient ce symbole, c'est que le type d'activation était proche de l'archétype concerné. Quand une configuration inconnue apparaît, un nouvel attracteur est créé. Cela ressemble à notre mécanisme d'introduction d'un nouvel archétype. Dans les réseaux ART, la création d'un nouvel attracteur dépend d'un seuil d'attention, qui peut être considéré comme équivalent à notre seuil d'introduction. On retrouve également dans les réseaux ART un mécanisme proche de notre glissement, puisque, à la suite de la phase lors de laquelle les attracteurs sont en concurrence, les poids de l'attracteur gagnant sont modifiés en fonction de l'entrée. Enfin, la comparaison entre les entrées et les configurations mémorisées au moyen des attracteurs se fait en calculant la distance existant entre elles en termes de répartition des neurones activés. Cela se rapproche de notre propre utilisation d'une fonction de similitude.

La principale différence entre l'information manipulée par ART et celle qui est caractéristique à notre propre formalisme est que nous séparons les aspects qualitatifs et quantitatifs. En ce qui concerne la composante qualitative, le type, nous utilisons une répartition de symboles, qui nous permet d'associer une valeur sémantique à l'information manipulée. Pour ce qui est des mécanismes d'apprentissage, de nombreuses différences existent malgré l'impression globale de similarité entre les deux formalismes. Ainsi, l'existence de préférences associées aux archétypes nous permet de moduler les processus d'émission et d'apprentissage, alors que dans les réseaux ART, tous les attracteurs sont traités sur le même plan. Ceci implique l'emploi d'un mécanisme de renforcement, qui est également absent des réseaux ART. De plus, le but des réseaux ART étant de réaliser de la classification de données, un seul attracteur s'active lors du traitement d'une information en entrée, et il correspond à la classe que le réseau associe à cette entrée. Dans notre cas, le type d'émission peut se composer de plusieurs symboles, correspondant à la position de plusieurs archétypes vis-à-vis de l'information en entrée, un peu comme si plusieurs attracteurs s'activaient à différents niveaux. Cette graduation du comportement de RAGE par rapport à celui, plus tranché, des réseaux ART, s'observe également au niveau du glissement : dans le cas du réseau ART, seul l'attracteur gagnant modifie ses poids. Dans RAGE, tous les archétypes glissent plus ou moins, en fonction de leur similitude avec l'entrée. On peut également remarquer qu'il n'existe pas, dans les réseaux ART, d'équivalent de notre mécanisme de fusion. Enfin, il faut souligner que nos mécanismes de propagation de l'information et d'apprentissage bénéficient de notre séparation de l'information, puisque la magnitude vient les moduler.

6. ETUDE DU COMPORTEMENT D'UN NOEUD

Nous avons réalisé une implémentation de RAGE, qui est décrite en détail dans l'annexe B. Nous l'avons l'utilisée pour étudier le comportement d'un nœud complet, c'est-à-dire en tenant compte à la fois des mécanismes de propagation et d'apprentissage. Nous nous intéressons à un nœud simple, possédant une seule entrée. Nous allons présenter quelques résultats issus de différentes simulations dans lesquelles nous soumettons le nœud à divers types et magnitudes.

6.1. Paramètres

Le type d'activation du nœud est défini sur le domaine $\tilde{D}_X = \{d_1, \ldots, d_5\}$, et le type d'émission est défini sur le domaine $D_X = \{s_i\}$. Sa TPT est donnée dans le Tableau V.6.1. A noter que pour chaque simulation effectuée, nous repartons de la TPT initiale.

archétype	$pref_X^t(A_i)$	$symb_X^t(A_i)$	$A_i(d_j)$			
			d_1	d_2	d_3	d_4
A_1	0,2	s_1	0,8	0,1	0,1	0
A_3	0,3	s_3	0,1	0,1	0,8	0
A_4	0,5	s_4	0,1	0,1	0	0,8

Tableau V.6.1 : TPT utilisée dans les exemples.

Comme le nœud n'a qu'une seule entrée, l'essentiel du traitement concerne le calcul des valeurs d'émission et l'apprentissage. Par conséquent, nous délaissons volontairement l'étude du calcul des valeurs d'activation en définissant $\tilde{f}_{M_X}$ et $\tilde{f}_{T_X}$ par la fonction identité. Le Tableau V.6.2 résume les les choix réalisés pour les autres fonctions, ainsi que les valeurs des différents paramètres du nœud.

notation	nom	implication dans les mécanismes	valeur / fonction utilisée ici
sim_{sig}	fonction de similitude	type d'émission (eq. V.2.32) magnitude d'émission (eq. V.2.37) renforcement (eq. V.3.1) introduction (eq. V.3.2) glissement (eq. V.3.6) fusion (eq. V.3.10)	sigmoïde : $a = 0,5$ et $b = -14$
$f_{T_X}^{(1)}$	-	type d'émission (eq. V.2.27)	sigmoïde : $a = 0,5$ et $b = 14$
$f_{M_X}^{(1)}$	-	magnitude d'émission (eq. V.2.37)	identité
$f_{M_X}^{(2)}$	-		
$f_{TPT_X}^{(1)}$	-	renforcement (eq. V.3.1) introduction (eq. V.3.2) glissement (eq. V.3.6)	identité
$f_{TPT_X}^{(2)}$	-	introduction (eq. V.3.2)	$1 - simax_X^t$
cr_X	coefficient de renforcement	renforcement (eq. V.3.1) introduction (eq. V.3.2)	0,03
cg_X	coefficient de glissement	glissement (eq. V.3.6)	0,05
so_X	seuil d'oubli	oubli (eq. V.3.4) introduction (eq. V.3.4)	0,02
sf_X	seuil de fusion	fusion (eq. V.3.10)	0,9

Tableau V.6.2 : récapitulatif des paramètres.

Comme nous utilisons une fonction de similitude sigmoïdale, nous nous permettons d'utiliser une fonction $f_{TPT_X}^{(2)}$ linéaire, et de prendre l'identité pour $f_{M_X}^{(1)}$ (c.f. paragraphe 5.1.2). Afin de simplifier la compréhension des résultats, $f_{M_X}^{(2)}$ est également l'identité (il n'y a pas de traitement de la magnitude d'activation pour calculer la magnitude d'émission, mise à part à la modulation exercée par le type d'activation via $f_{M_X}^{(1)}$).

6.2. Type d'activation constant

Dans ces premières simulations, le type est toujours constant, tandis que la magnitude d'activation qui lui associée est tirée au hasard, elle varie entre -1 et 1.

6.2.1. Proche d'un archétype

Nous utilisons tout d'abord un type d'activation constant extrêmement proche ($sim_{sig} = 0,99$) de l'archétype A_1 :

$$\tilde{T}_1 = \{(d_1;0,8);(d_2;0,07);(d_3;0,07);(d_4;0,06)\} \quad \text{(eq. V.5.1)}$$

Si le type d'activation est constant, on peut voir que ce n'est pas le cas du type d'émission (Figure V.6.1.a) : on observe des émissions de bruit (tous les symboles ont le même poids), par exemple juste avant l'itération 20. Elles correspondent à des magnitudes d'activation trop faibles, qui ont bruité le type d'émission.

Lorsque ce n'est pas du bruit qui est émis, c'est le symbole s_1 qui est largement dominant, ce qui est normal, puisque le type d'activation est proche de A_1 (qui est l'archétype associé à s_1). En ce qui concerne l'évolution des préférences de la TPT (Figure V.6.1.c), on peut observer que la préférence de l'archétype A_1 (qui est le plus proche de l'entrée) augmente, au détriment des autres archétypes.

6.2.2. Inconnu

On utilise à présent un type d'activation constant ne ressemblant à aucun archétype de la TPT :

$$\tilde{T}_2 = \{(d_1;0,05);(d_2;0,05);(d_3;0,05);(d_4;0,05);(d_5;0,8)\} \quad \text{(eq. V.5.2)}$$

Ce type est très différent de tous les archétypes déjà existants car il contient essentiellement d_5, un élément de $\tilde{D}_X$, qui n'est dans aucun des archétypes. La similitude entre $\tilde{T}_2$ et les archétypes est de $sim_{sig} = 0,0105$.

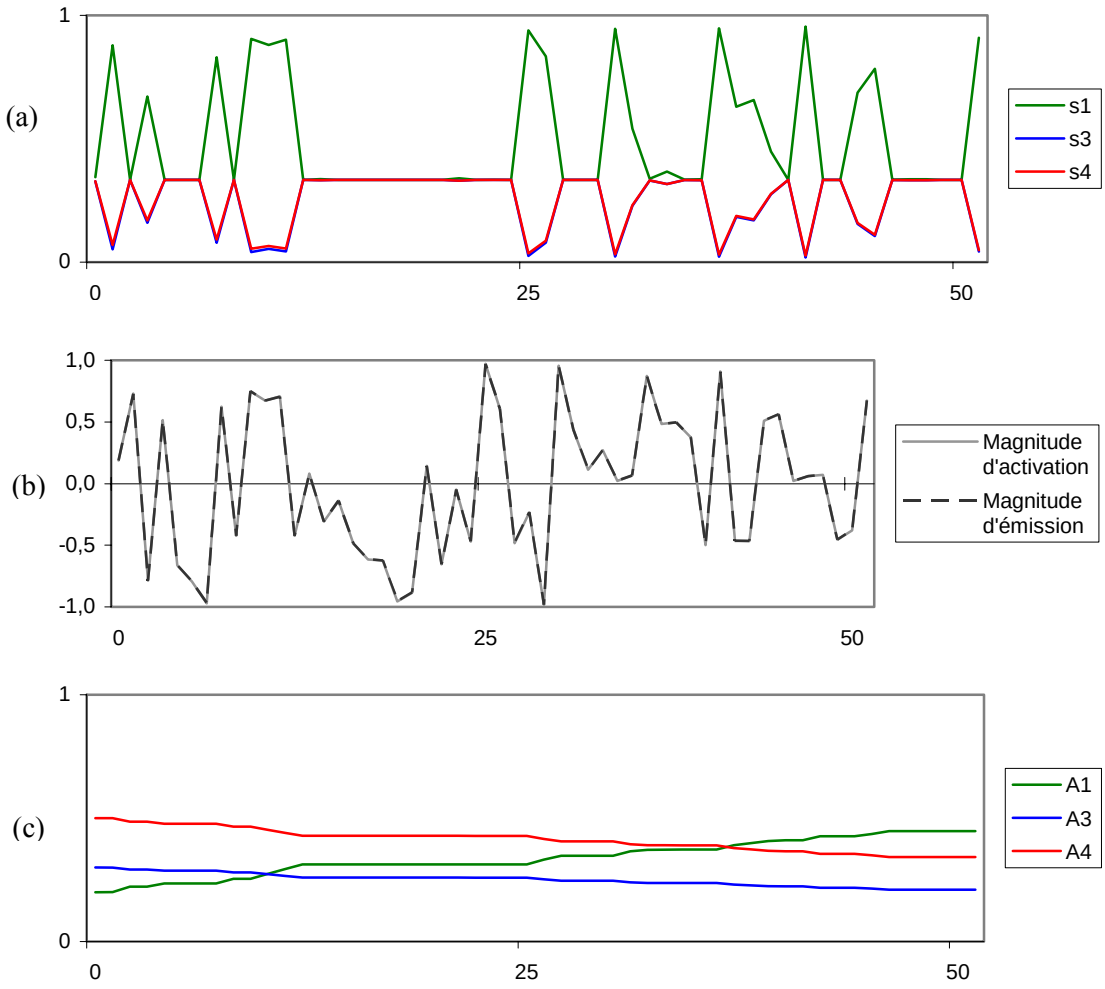

Figure V.6.1 : nœud soumis à un type d'activation constant proche de l'archétype A_1, et à une magnitude aléatoire.
(a) : évolution des poids du type d'émission. (b) : évolution des magnitudes d'activation et d'émission. (c) : évolution des préférences de la TPT.

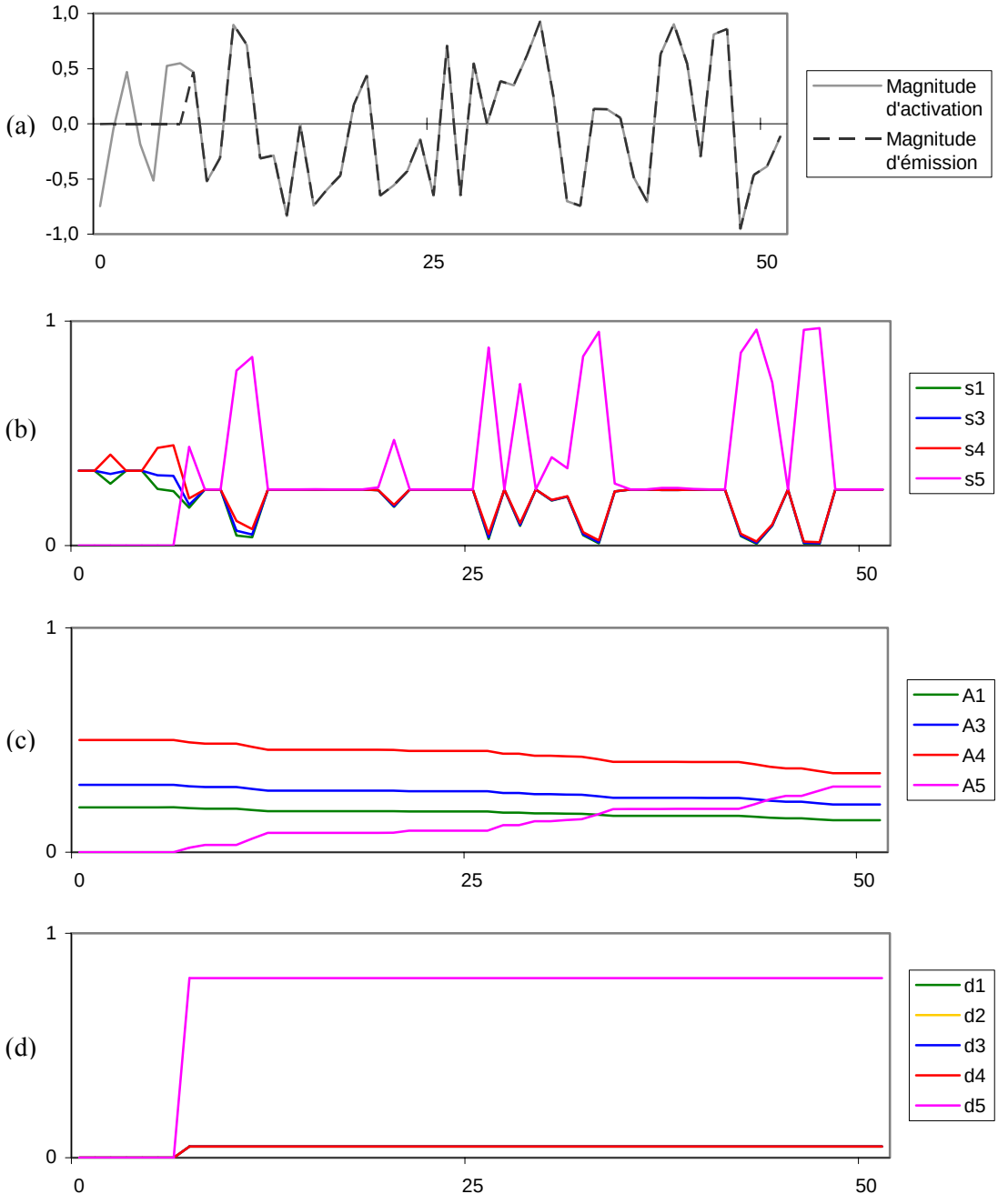

Figure V.6.2 : nœud soumis à un type d'activation constant différent de tous les archétypes de la TPT, et à une magnitude aléatoire.
(a) : évolution des magnitudes d'activation et d'émission. (b) : évolution des poids du type d'émission. (c) : évolution des préférences de la TPT. (d) : évolution des poids de l'archétype A_5 (créé pendant la simulation).

On observe, sur le graphique des préférences (Figure V.6.2.c), l'introduction d'un nouvel archétype noté A_5 vers la 10ème itération. La Figure V.6.2.d montre l'évolution des poids de cet archétype. On peut constater qu'il est caractérisé par le fort poids associé à d_5.

La Figure V.6.2.a représente l'évolution des magnitudes d'activation (trait plein) et d'émission (pointillés). En considérant la magnitude d'activation, on s'aperçoit que le nouvel archétype n'a pas été créé avant la 10 itération en raison d'une trop faible magnitude d'activation.

On peut également remarquer (Figure V.6.2.a) qu'à la suite de la création de ce nouvel archétype, la magnitude d'émission augmente enfin. Auparavant, les archétypes étaient tous très éloignés du type d'activation, ce qui rabaissait la magnitude. Cette création a également un effet sur le type d'émission (Figure V.6.2.b), qui, avant, était très bruité, et qui voit, après, le symbole associé au nouvel archétype prendre énormément d'importance. Cette sortie est modulée par la magnitude d'activation, comme nous l'avions déjà remarqué dans le cas précédent (Figure V.6.1.a).

6.2.3. Bruit

Enfin, nous utilisons ici un type d'activation constant représentant du bruit. Il contient les 4 éléments de $\tilde{D}_X$ présents dans les archétypes de la TPT, répartis de façon uniforme :

$$\tilde{T}_2 = \{(d_1;0,25);(d_2;0,25);(d_3;0,25);(d_4;0,25)\} \quad \text{(eq. V.5.3)}$$

Si on observe l'évolution de la magnitude d'émission (Figure V.6.3.a), on voit qu'elle se met progressivement à suivre la magnitude d'activation. Ceci est du au fait qu'au début de la simulation, aucun archétype n'est proche du type d'activation. De ce fait, la magnitude d'émission est diminuée. Mais les archétypes se rapprochent progressivement du type d'activation par glissement, et la magnitude d'activation est donc de moins en moins rabaissée.

En ce qui concerne le type d'émission, on voit qu'il s'agit le plus souvent de bruit (Figure V.6.3.b). A quelques reprises, le type émis s'éloigne du bruit. Ceci correspond à des pics de magnitude d'activation, qui font ressortir les différences de préférences entre les archétypes lors du calcul de type d'émission. Par exemple, on voit qu'à l'occasion de ces variations, le symbole de plus grand poids correspond à s_4. Or il s'agit du symbole associé à l'archétype A_4, qui a la plus forte préférence dans la TPT.

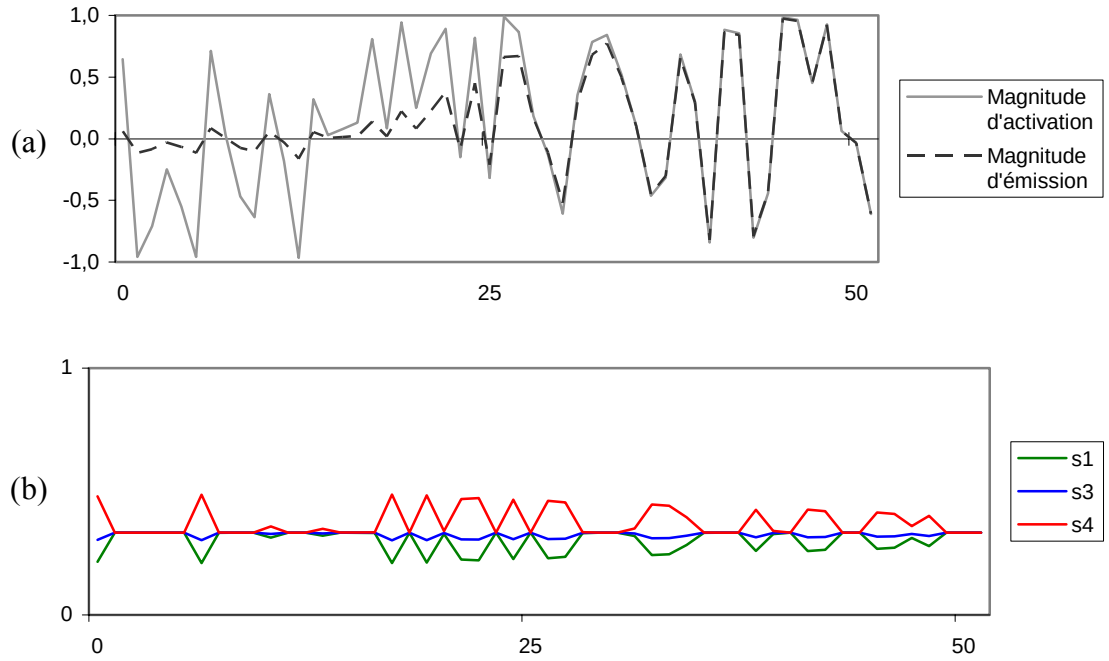

Figure V.6.3 : nœud soumis à un type d'activation constant $\tilde{T}_{Bruit}$, représentant du bruit, et à une magnitude aléatoire.
(a) : évolution des magnitudes d'activation et d'émission. (b) : évolution des poids du type d'émission.

6.3. Type d'activation variable

Dans cette deuxième série de simulations, la magnitude d'activation est tirée au hasard, comme précédemment. Mais maintenant, à la différence des simulations précédentes, le type d'activation varie également.

6.3.1. Aléatoire

Nous utilisons ici un type tiré au hasard. Il est construit à partir de $d_1,\ldots,d_5$, avec des poids tirés au hasard. La Figure V.6.4.a montre la grande variété des types d'activation présentés. En particulier, on voit apparaître à plusieurs reprises des types pour lesquels le poids associé à d_2 est très fort. Ceci a des répercussions, puisque la figure Figure V.6.4.c montre qu'un nouvel archétype noté A_5 apparaît, un peu avant la 25ème itération. Bien sûr, cet archétype est caractérisé par une

prépondérance de d_2 (Figure V.6.4.d). Il est intéressant de remarquer que le type d'émission (Figure V.6.4.b) n'offre pas du tout le même aspect aléatoire que le type d'activation. On distingue soit du bruit, soit des types dans lesquels un symbole domine franchement les autres.

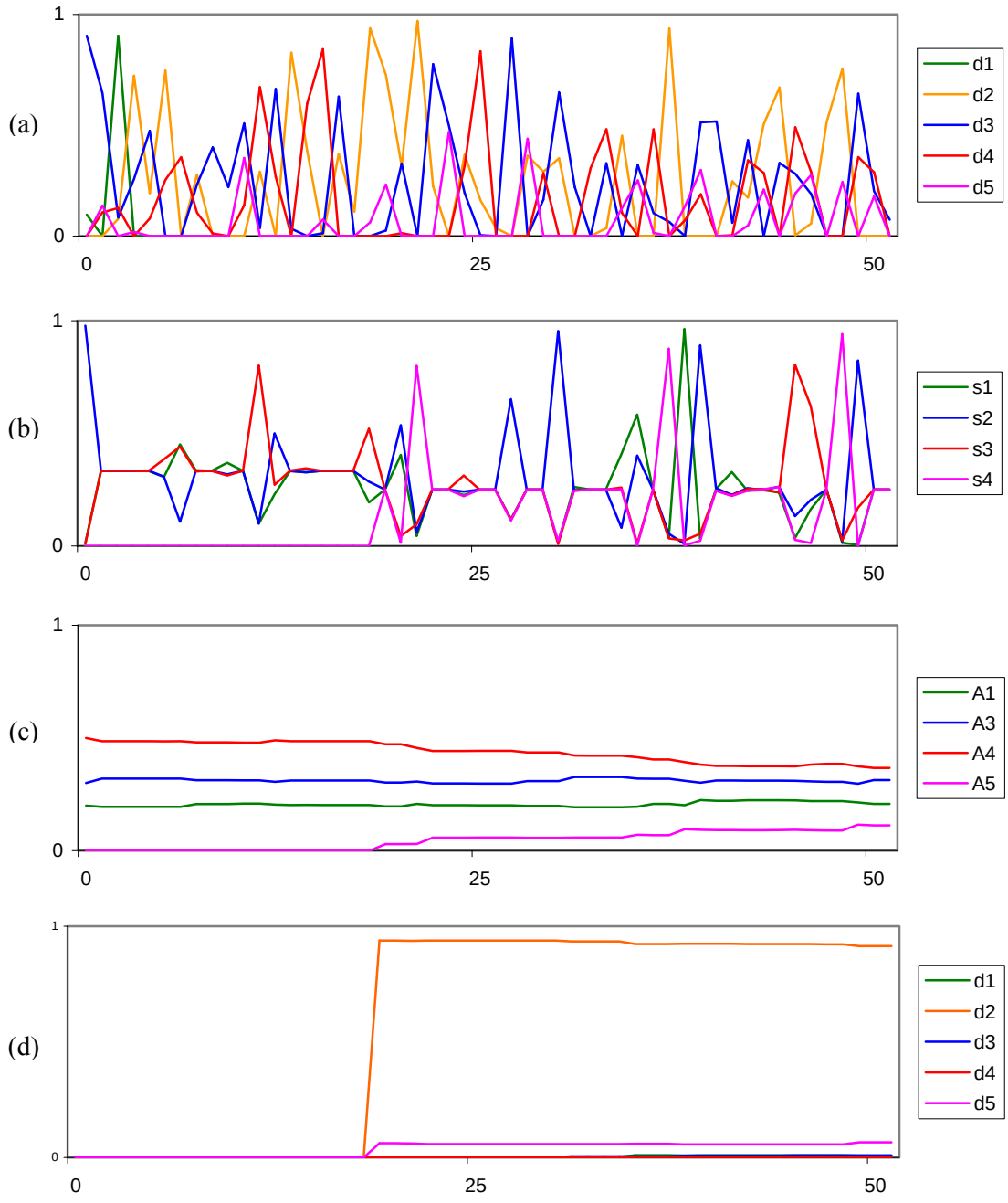

Figure V.6.4 : nœud soumis à des type et magnitude d'activation aléatoires.
(a) : évolution des poids du type d'activation. (b) : évolution des poids du type d'émission. (c) : évolution des préférences de la TPT. (d) : évolution des poids d' l'archétype A_5.

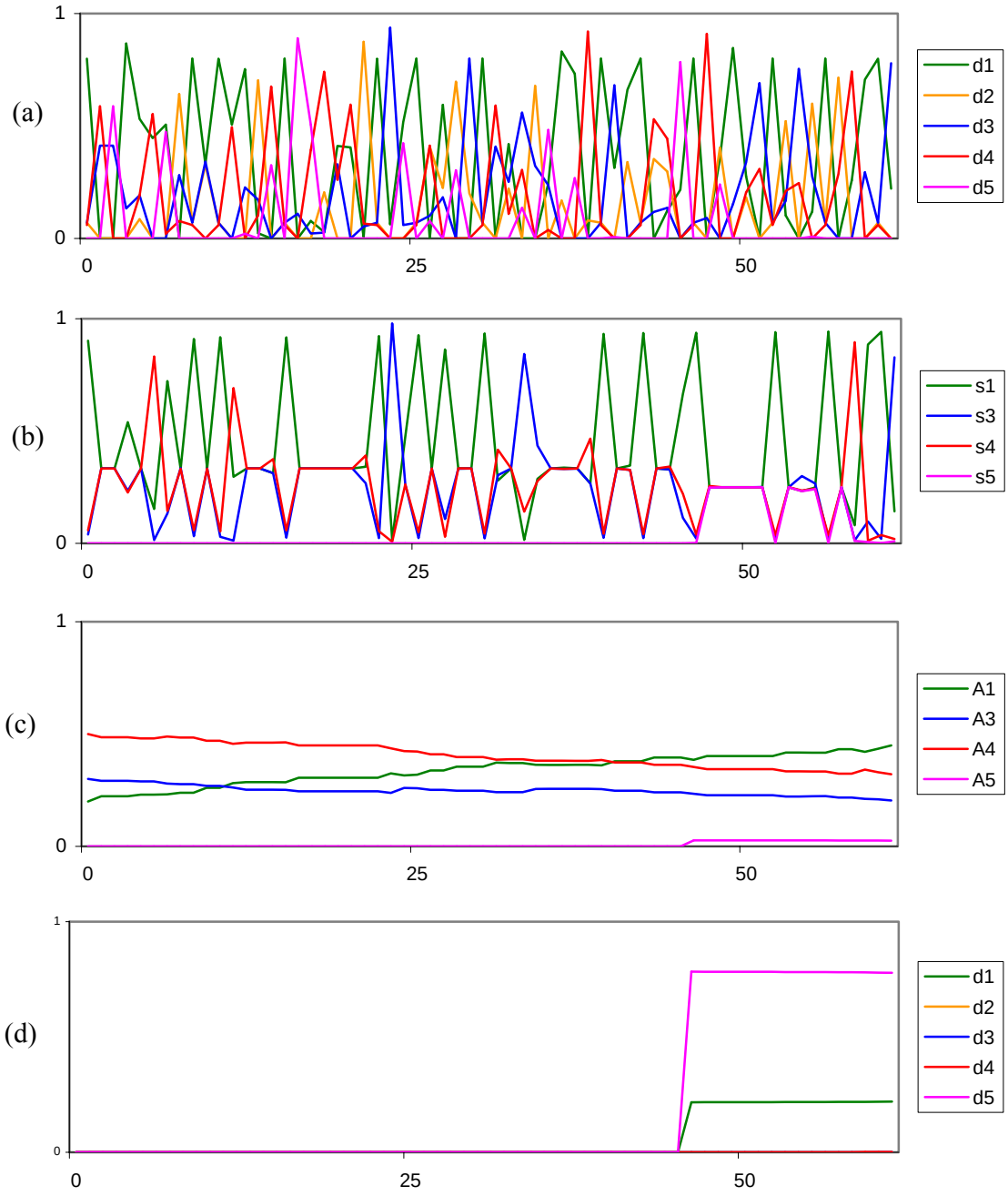

Figure V.6.5 : nœud soumis à des type et magnitude d'activation partiellement aléatoires. (a) : évolution des poids du type d'activation. (b) : évolution des poids du type d'émission. (c) : évolution des préférences de la TPT. (d) : évolution des poids de l'archétype A_5.

6.3.2. Partiellement aléatoire

Cette fois ci, le type utilisé n'est que partiellement tiré au hasard : nous imposons la présence régulière du type $\tilde{T}_1$ utilisé précédemment (paragraphe 6.2.1), et qui possède la propriété d'être proche de l'archétype A_1. Ce type apparaît en moyenne toutes les 5 itérations, et on lui associe une magnitude d'activation de 1. L'allure du type d'activation ressemble nettement à ce qui est observé lorsque les entrées sont entièrement tirées au hasard (Figure V.6.4.a et Figure V.6.5.a), sauf qu'ici nous observons des pics de d_1 relativement réguliers. En revanche, le type d'émission est beaucoup plus différent (Figure V.6.4.a et Figure V.6.5.b), puisque l'on voit très nettement apparaître les émissions correspondant à des $\tilde{T}_1$ imposés en entrée.

A noter que la préférence associée à A_1 augmente logiquement (Figure V.6.5.c), en raison de la fréquence d'apparition de $\tilde{T}_1$. Ceci n'empêche pas, toutefois, l'apparition d'un nouvel archétype, noté A_5, comme on l'avait déjà observé dans le cas précédent. Il est cette fois plutôt centré sur d_5 (Figure V.6.5.d), alors que c'était d_2 qui dominait dans la simulation précédente (Figure V.6.4.d). Ceci est dû au fait que les entrées sont tirées au hasard, et que l'apparition de d_5 avec un poids fort a coïncidé ici avec de fortes magnitudes d'activation, ce qui n'était pas le cas dans la simulation précédente.

6.4. Apprentissage à partir d'une TPT vierge

Dans les deux exemples suivants, nous reprenons le nœud précédent avec tous ses paramètres et fonctions. Il y a toutefois une différence en ce qui concerne la TPT, car celle-ci est vierge de tout archétype. L'intérêt de ces deux manipulations est d'étudier comment la TPT se forme à partir de rien. Pour cela, nous soumettons le nœud à une stimulation complètement aléatoire (type et magnitude), puis partiellement aléatoire (le type est régulièrement contraint à prendre une certaine valeur) (c.f. simulations précédentes).

6.4.1. Stimulation complètement aléatoire

Les valeurs d'activation utilisées ici sont les mêmes que dans le paragraphe 6.3. Pour faciliter la lecture des résultats, nous utilisons, dans les Figure V.6.6 et Figure V.6.7, un code de couleur qui nous permet d'établir des correspondances entre les archétypes et les types d'activation et

d'émission. Par exemple, dans la Figure V.6.6.d, l'archétype A_1 est représenté en rouge car il décrit une répartition dans laquelle d_4, lui-même représenté en rouge dans Figure V.6.6.b et Figure V.6.7, est dominant. Dans la Figure V.6.6.c, le poids de s_1 est également représenté en rouge, car il s'agit du symbole associé à A_1.

On remarque l'introduction de cinq archétypes différents (Figure V.6.6.d). Dans la Figure V.6.7, nous ne représentons que les quatre premiers apparus. Le premier archétype apparu domine largement les autres par sa préférence, ce qui est logique puisqu'il a bénéficié d'une préférence maximale lors de son introduction dans la TPT. Toutefois, en raison de la nature irrégulière des types d'activation, on observe une baisse progressive de sa préférence au profit des autres archétypes.

En ce qui concerne la nature des archétypes, il faut souligner que deux d'entre eux sont caractérisés par la prédominance de deux symboles (Figure V.6.7.b et Figure V.6.7.c), chose que l'on n'avait pas observée dans les simulations précédentes. En effet, auparavant, la TPT contenait déjà des archétypes spécifiquement dédiés à un symbole, ce qui empêchait l'apparition d'archétypes de ce genre.

6.4.2. Stimulation partiellement aléatoire

Ici, nous utilisons les mêmes valeurs d'activation que dans le paragraphe 0. Pour les figures, nous conservons le code de couleurs utilisé juste avant, dans le paragraphe 6.3.1. Nous observons cette fois l'apparition de quatre archétypes (Figure V.6.8.d). Dans la simulation précédente, l'apparition d'un archétype dans lequel d_1 possédait un poids fort apparaissait assez tard (A_4, Figure V.6.6.d). Ici, un type basé sur d_1 est régulièrement imposé en entrée, et de ce fait un archétype dédié à d_1 apparaît assez vite (A_2, Figure V.6.8.d et Figure V.6.9.b).

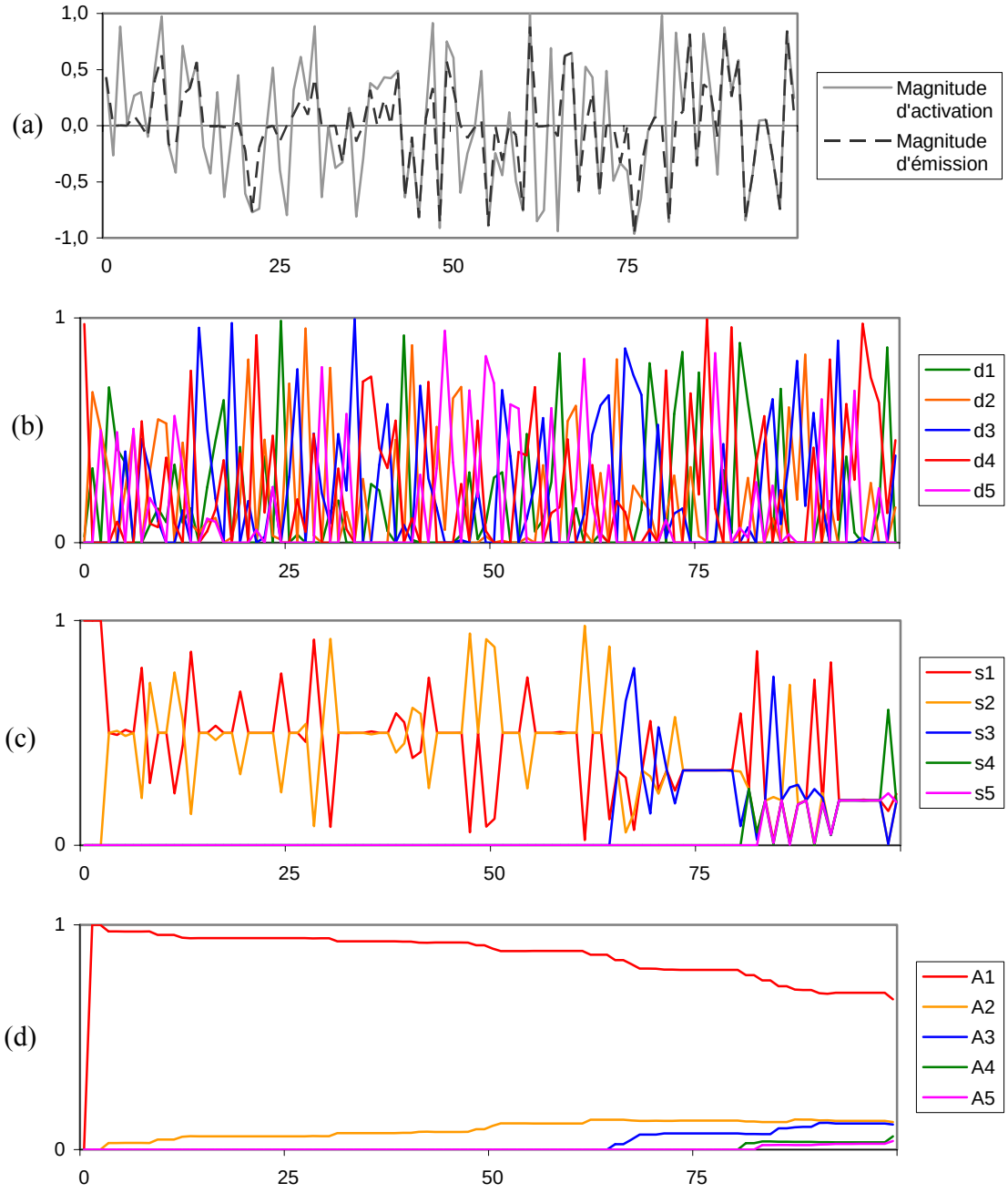

Figure V.6.6 : nœud soumis à des type et magnitude d'activation aléatoires, avec une TPT initiale vide.
(a) : évolution des magnitudes d'activation et d'émission. (b) : évolution des poids du type d'activation. (c) : évolution des poids du type d'émission. (d) : évolution des préférences de la TPT.

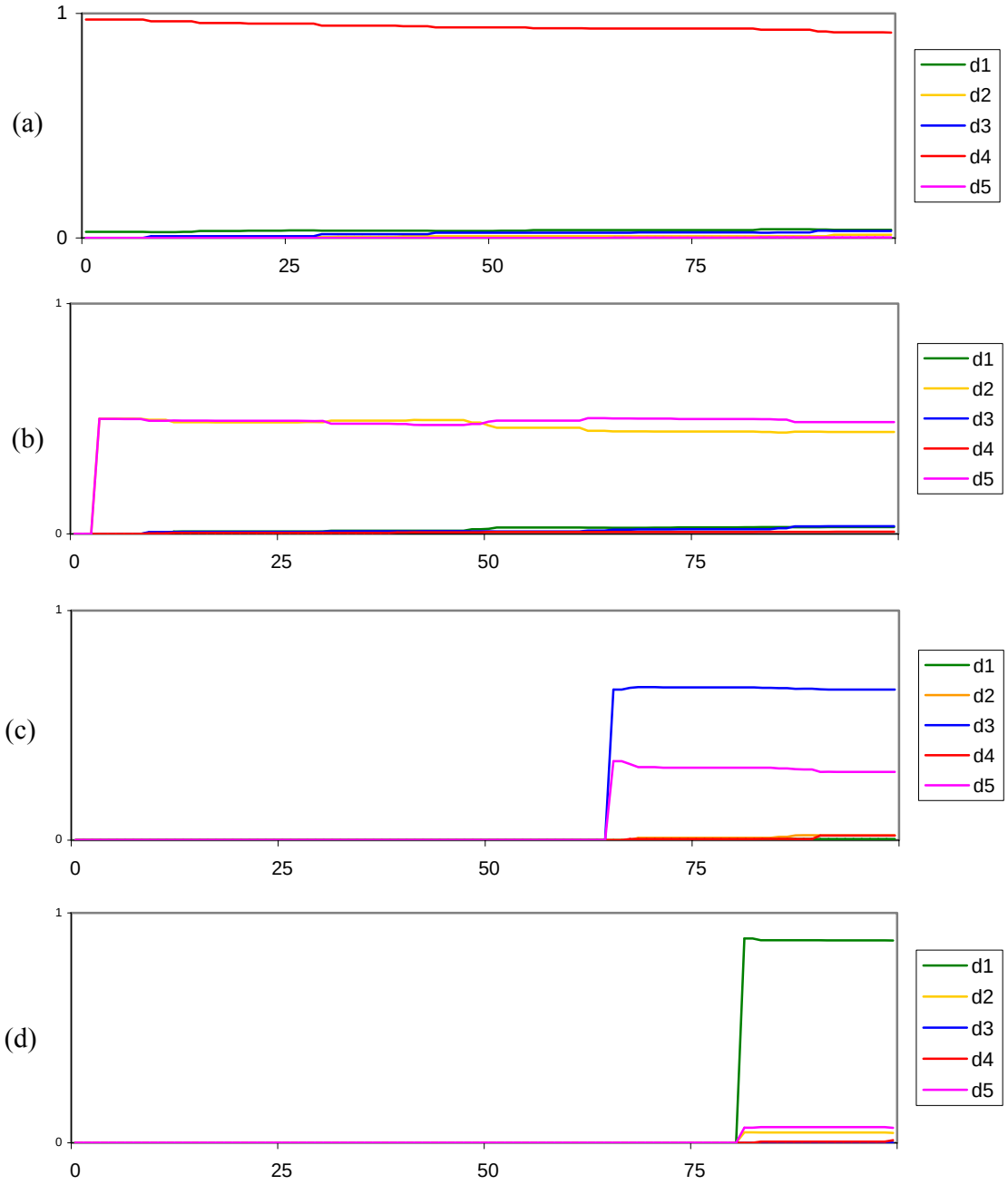

Figure V.6.7 : nœud soumis à des type et magnitude d'activation aléatoires, avec une TPT initiale vide.
(a) : évolution des poids du premier archétype introduit dans la TPT. (b)-(c)-(d) : même chose pour les 2ème, 3ème et 4ème archétypes créés par la suite.

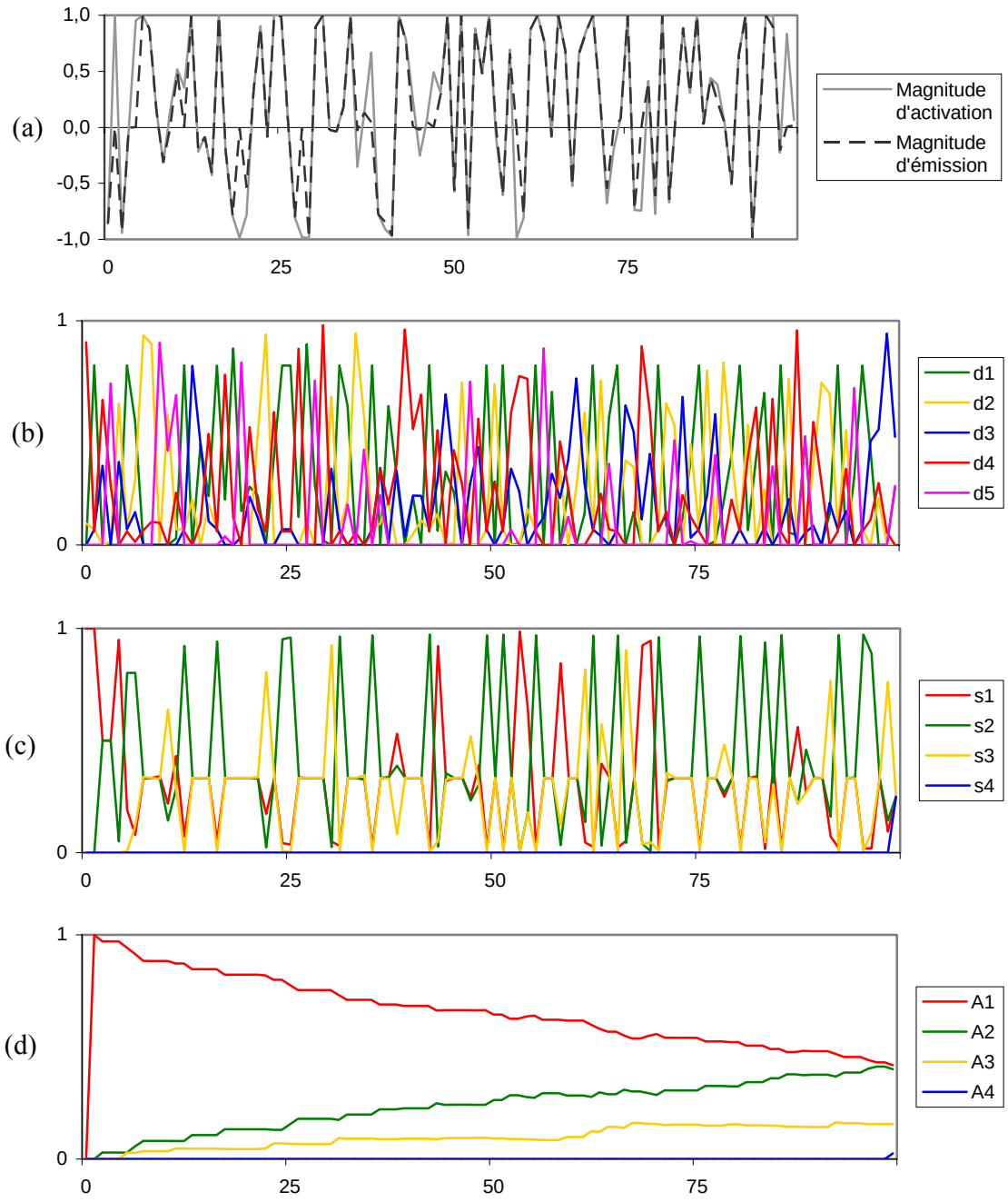

Figure V.6.8 : nœud soumis à des type et magnitude d'activation partiellement aléatoires, avec une TPT initiale vide.
(a) : évolution des magnitudes d'activation et d'émission. (b) : évolution des poids du type d'activation. (c) : évolution des poids du type d'émission. (d) : évolution des préférences de la TPT.

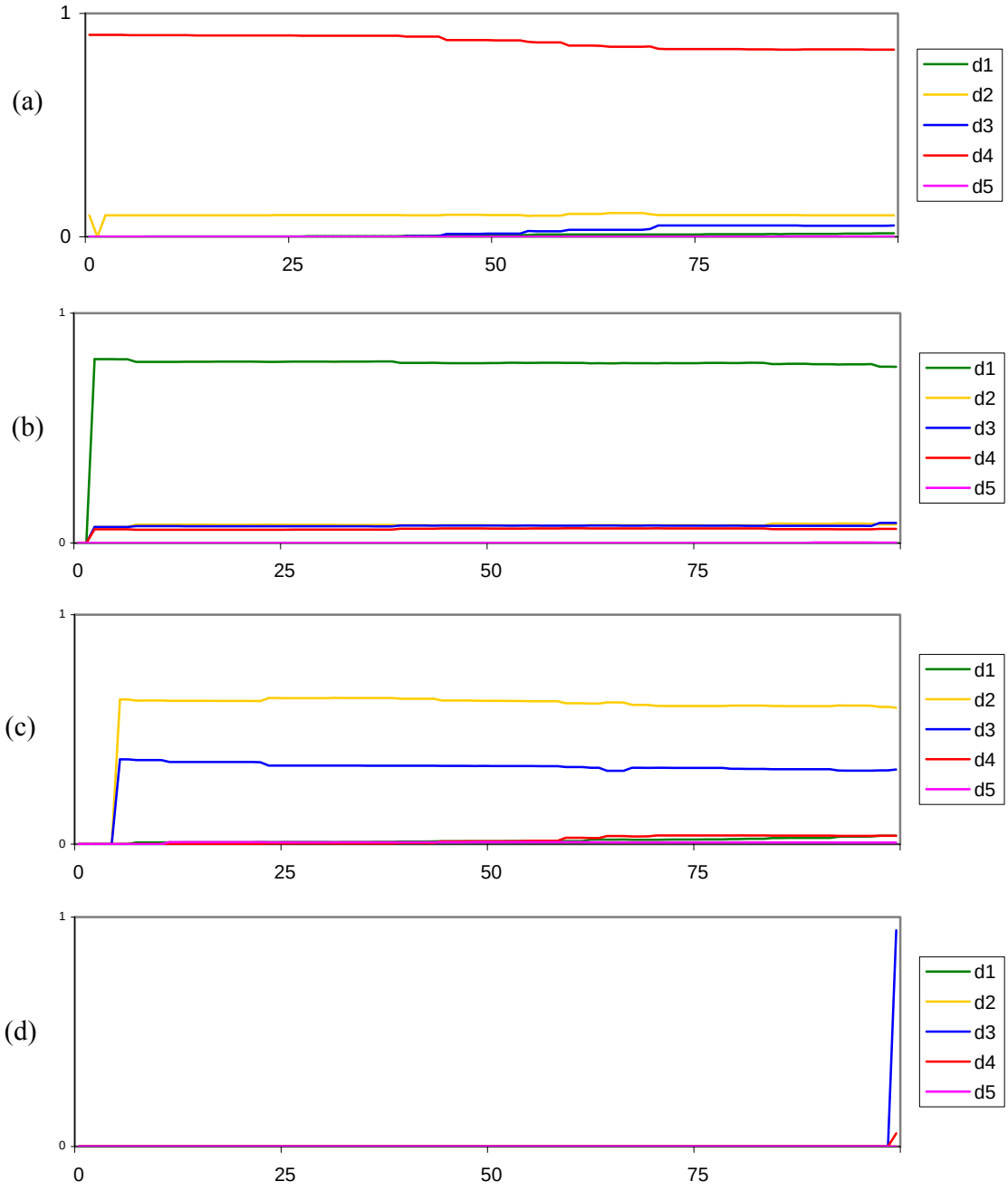

Figure V.6.9 : nœud soumis à des type et magnitude d'activation partiellement aléatoires, avec une TPT initiale vide.
(a) : évolution des poids du premier archétype introduit dans la TPT. (b)-(c)-(d) : même chose pour les 2ème, 3ème et 4ème archétypes créés par la suite.

Chapitre VI

APPLICATIONS

Le but de cette partie est de montrer concrètement comment RAGE peut être appliqué à la modélisation cérébrale. Pour cela, nous allons décrire deux modèles, définis grâce à l'implémentation de RAGE décrite dans l'annexe B. Ils sont utilisés pour étudier des mécanismes cérébraux bien précis. Ils n'exploitent pas toutes les possibilités du formalisme, nous sommes notamment limités à des cas ne nécessitant pas de modéliser des mécanismes d'apprentissage. De plus, notre démarche est à vocation uniquement illustrative, nous ne visons aucunement à présenter des modèles définitifs des tâches cognitives étudiées. Les modèles ont été définis à partir d'hypothèses sur le fonctionnement cérébral, et validés à partir de données de neuroimagerie. Dans notre description du processus de modélisation, nous nous attachons à décrire comment nous avons exprimé les hypothèses en termes de modèles. La définition des hypothèses elle-même relève de la neuropsychologie, et nous ne l'abordons donc pas ici.

La première application reprend un modèle utilisé par le passé pour valider BioCaEn, le précédent formalisme de modélisation cérébrale de notre groupe de travail. Cette application a été réalisée dans deux buts. Tout d'abord, vérifier que notre nouveau formalisme offrait au moins les mêmes possibilité de modélisation que son prédécesseur. Comme BioCaEn est limité aux processus automatiques et ne permet pas d'effectuer d'apprentissage, ce test a pu être effectué au début du développement de RAGE, avant d'implémenter des fonctions supplémentaires. Le deuxième but était d'effectuer une comparaison des résultats obtenus par BioCaEn et par RAGE sur le même modèle. En effet, une des raisons qui a causé l'abandon de BioCaEn est la grande imprécision des données obtenues en fin de simulation, à cause des propriétés de la simulation semi-qualitative (c.f. chapitre III).

La seconde application constitue un travail de modélisation original. Elle repose sur une étude neuropsychologique réalisée dans notre groupe de recherche, et a été effectuée en collaboration avec les neuropsychologues chargés de l'étude, en se basant sur leurs hypothèses quant au fonctionnement des mécanismes cérébraux étudiés. Le modèle présente la propriété de faire

intervenir un traitement des types plus complexe que ce n'est le cas dans la première application, puisque nous modélisons ici un processus de catégorisation.

1. MODELE DE LA REPONSE DE L'AIRE VISUELLE PRIMAIRE A UN STIMULUS SIMPLE

Le modèle original est un des modèles précédemment développé pour BioCaEn [Pastor *et al.* '00]. Le but était d'implémenter différentes hypothèses émises par Peter Fox et Marcus Raichle sur le fonctionnement de l'aire primaire visuelle, en se basant sur des résultats issus de deux études TEP [Fox & Raichle '84, '85]. Le modèle et la comparaison des résultats de BioCaEn avec ceux obtenus par RAGE sur cet exemple ont fait l'objet de plusieurs publications [Labatut & Pastor '03a, b; Labatut *et al.* '03b].

1.1. Etude en activation

Les deux expériences de Fox et Raichle [Fox & Raichle '84, '85] sont des études en activation faisant intervenir uniquement des processus automatiques, puisque le sujet ne doit effectuer aucune action particulière pendant l'acquisition, à l'exception du fait de rester attentif. La tâche à effectuer est donc passive. Le sujet est soumis à des stimuli visuels, le but est d'étudier les effets de la fréquence de présentation des stimuli sur le niveau d'activation.

Le nombre de sujets est de 9 pour la première expérience [Fox & Raichle '84], et de 4 pour la seconde [Fox & Raichle '85]. Chaque sujet reçoit une injection de marqueur radioactif avant d'être placé sous une caméra PET. La durée d'acquisition est de l'ordre de quarante secondes pour obtenir une image, ce qui correspond à ce que l'on appelle un *run*. La stimulation est maintenue pendant toute cette durée.

Les stimuli visuels sont délivrés au sujet par le biais d'écrans constitués de matrices de 36 diodes. Dans la première expérience, les stimuli sont des flashes, toutes les diodes s'allument puis s'éteignent en même temps. Dans la seconde expérience, une diode sur deux est toujours allumée, la matrice constitue un damier. Les stimuli consistent à inverser le damier, c'est-à-dire que les diodes allumées s'éteignent, et inversement. Dans les deux cas, les stimuli durent 5 millisecondes, avec une intensité lumineuse constante durant toute l'expérience.

Pendant tout le run, le stimulus soumis au sujet ne change pas, que ce soit sa nature (flash ou damier) ou bien sa fréquence de présentation. Par contre, plusieurs runs sont effectués pour les deux types de stimuli, avec différentes fréquences : 1 ; 3,9 ; 7,8 ; 15,5 ; 33,1 et 61 Hz.

Les mesures de TEP donnent la variation du *débit sanguin cérébral régional* (rCBF, *r*egional *C*erebral *B*lood *F*low) au niveau de l'aire visuelle primaire. La Figure VI.10 montre ces résultats sous la forme de variations moyennes pour l'ensemble des sujets, exprimées en pourcentages (ΔrCBF%), en fonction de la fréquence de stimulation. Les variations sont calculées par rapport à des mesures effectuées sur les sujets au repos, c'est-à-dire lorsque ceux-ci ne perçoivent pas de stimulation.

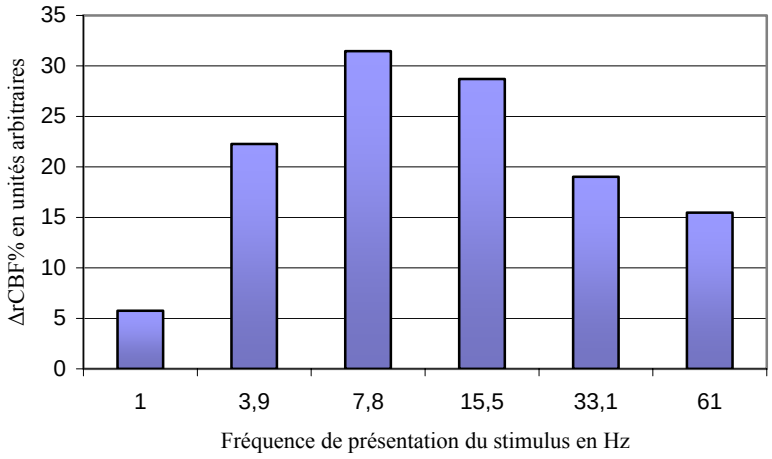

Figure VI.10 : résultats de l'expérience de Fox & Raichle utilisant des flashes lumineux comme stimuli [Fox & Raichle '84, '85].

La forme de la courbe montre que le débit sanguin cérébral régional ne varie pas linéairement en fonction de la fréquence de présentation du stimulus. On en déduit que le la réponse neuronale globale sous-jacente varie elle aussi de façon non-linéaire. Les auteurs décrivent une augmentation du ΔrCBF% jusqu'à 7.8 Hz. Il atteint ensuite son maximum entre 7.8 et 15 Hz, avant de décroître ensuite lentement pour se stabiliser largement en dessous de sa valeur maximale.

Les auteurs émettent trois hypothèses quant à la cause de la baisse d'activation observée. Première hypothèse : une baisse du niveau de l'activation neurale, due à une habituation à la stimulation. Deuxième hypothèse : une absence de réponse à certains stimuli, lorsque la fréquence est trop élevée, c'est-à-dire quand les stimuli sont trop rapprochés les uns des autres dans le temps.

Troisième hypothèse : une combinaison des deux premières hypothèses, celles-ci n'étant pas mutuellement exclusives.

Cette étude avait été retenue pour appliquer BioCaEn en raison de la simplicité du réseau d'aires cérébrales impliquées, et de la possibilité de confronter les différentes hypothèses en les implémentant chacune sous la forme d'un modèle, et en comparant les résultats obtenus par simulation. Ces modèles ont fourni une argumentation en faveur de la troisième hypothèse énoncée par Fox et Raichle, c'est-à-dire la conjonction des deux premières hypothèses.

1.2. Modèle de boucle thalamo-corticale

Notre but n'étant pas de refaire intégralement l'étude réalisée dans [Pastor *et al.* '00], nous nous sommes focalisés sur la troisième hypothèse et sur le modèle BioCaEn correspondant, que nous avons adapté sous forme de modèle RAGE. De plus, nous n'effectuerons pas d'interprétation détaillée des résultats vis-à-vis des données de TEP, puisque cela a déjà été fait pour la version BioCaEn du modèle.

La Figure VI.1.11 présente le réseau fonctionnel statique employé pour la modélisation. Les liens en gras représentent les liaisons structurelles, c'est-à-dire connectant les différentes structures. Le délai qui leur est affecté est 2 ms, alors que l'information met 1 ms à parcourir les autres liens. En raison de ces délais, le cycle orienté observé dans le réseau statique disparaît quand on déroule celui-ci pour obtenir le réseau dynamique (c.f. chapitre IV.1). Dans les équations qui vont suivre, les éléments notés $a_A^{(B)}$ représentent des paramètres, dont les valeurs sont données à la fin de cette partie, dans le Tableau VI.1.1. La signification de la notation est la suivante : $a_A^{(B)}$ est le paramètre affectant B utilisé dans le calcul de A.

Le nœud *OGN* est exogène, c'est-à-dire qu'il représente l'influence de l'extérieur sur le système modélisé. En l'occurrence, il s'agit de l'information en provenance des voies optiques. On manipule donc ici une représentation interne de l'information, et non pas du stimulus brut. Lors de la simulation, les valeurs que prend ce nœud sont prédéfinies en fonction de la fréquence de présentation du stimulus. Un flash lumineux est modélisé par une magnitude arbitraire de 1 et un type composé d'un seul symbole représentant la longueur d'onde de la lumière utilisée pour le flash

(orange). Une absence de stimulation correspond à une magnitude de 0 et un type représentant du bruit.

Le reste du réseau se découpe en deux : le cortex visuel, et une structure thalamique. Cette dernière est utilisée pour implémenter une boucle thalamo-corticale. En effet, chaque région corticale semble connectée à une partie du thalamus (c.f. chapitre I). Ces boucles sont supposées réguler l'excitabilité des cellules pyramidales des cortex, en réduisant leur seuil de décharge [Pastor *et al.* '00].

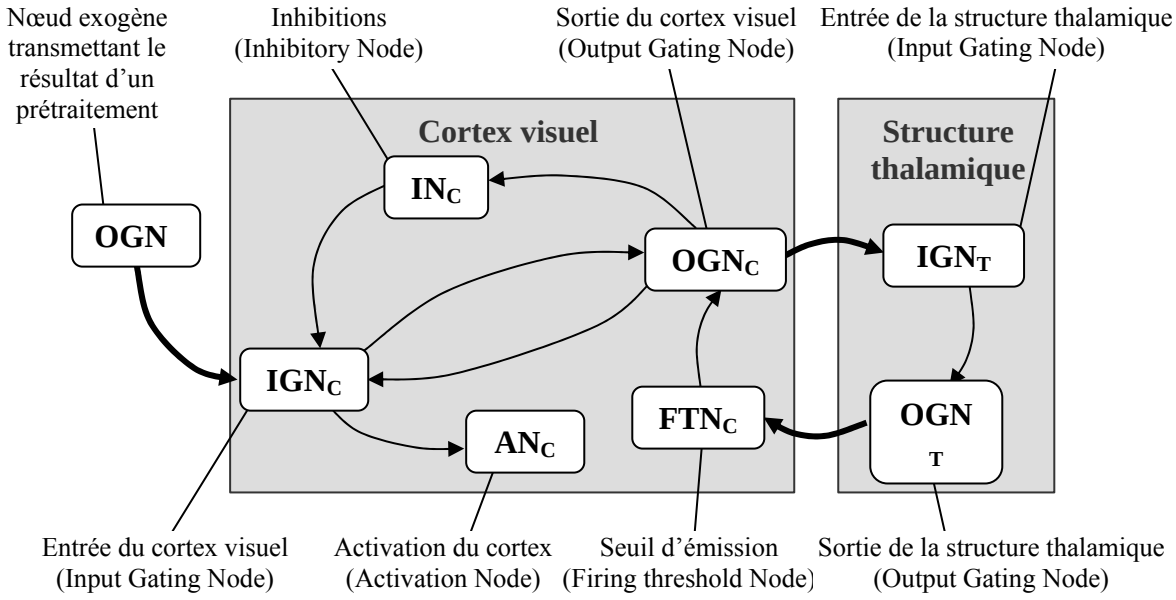

Figure VI.1.11 : modèle de cortex visuel incluant une boucle thalamo-corticale.

Le nœud IGN_C modélise l'entrée du cortex visuel. Il réalise l'intégration temporelle de l'information qui arrive au cortex visuel, et la combine avec les influences internes (IN_C et OGN_C). Ici, l'intégration n'est que temporelle, puisqu'il n'y a qu'une seule influence externe. Mais, dans le cas général, un tel nœud est également susceptible de réaliser une intégration spatiale des entrées. IN_C représente la population de neurones inhibiteurs du cortex visuel, il exerce une influence négative qui rabaisse le niveau d'activation de IGN_C, et se traduit par une soustraction dans (eq. VI.1.1). OGN_C est la sortie du cortex visuel, et, lorsqu'il émet, il empêche IGN_C de traiter les entrées de l'aire. Ce mécanisme permet d'implémenter la période réfractaire de la région. Nous faisons ici l'hypothèse qu'on trouve au niveau d'une aire cérébrale un phénomène similaire à celui qui existe pour le neurone. En d'autres termes, nous supposons qu'après avoir émis, une aire cérébrale ne traite plus ses entrées pendant un certain temps. Lorsque OGN_C cesse d'émettre, IGNc peut de nouveau traiter ses entrées. Le mécanisme est représenté dans (eq. VI.1.1) par une fonction

sigmoïdale décroissante, utilisée comme un coefficient multiplicatif de la magnitude du stimulus entrant :

$$M^{t}_{IGN_C} = \sigma\left(M^{t-1}_{OGN_C}\right)\left(a^{(Stim_C)}_{IGN_C} M^{t-2}_{Stim_C} f\left(simax\right)\right) + \left(a^{(IGN_C)}_{IGN_C} M^{t-1}_{IGN_C}\right) - \left(a^{(IN_C)}_{IGN_C} M^{t-1}_{IN_C}\right) \text{(eq. VI.1.1)}$$

IGN_C réalise également un filtrage de l'information entrante. Le filtrage est effectué en utilisant une TPT associée à ce nœud. Ici, la TPT ne possède qu'un seul archétype, lequel correspond au type de stimulation. Comme l'application présente n'utilise qu'un seul type de stimulus, toujours couplé à la même magnitude, et que l'apprentissage n'est pas implémenté, la TPT a ici un rôle relativement mineur, voire négligeable. De manière générale, on peut dire que le traitement du type n'a aucune influence sur le reste du modèle, et il sera donc omis pour les autres nœuds. Il faut toutefois souligner que si le modèle de cortex visuel était connecté à d'autres modèles de région, le type issus du cortex visuel aurait éventuellement son importance dans le traitement des régions situés en aval dans ce réseau à grande échelle.

L'information filtrée et intégrée par IGN_C est relayée vers les nœuds OGN_C et AN_C. Le nœud AN_C représente l'activation de la zone entière, intégrée sur le temps de simulation écoulé. C'est la valeur finale de ce nœud qui sera utilisée pour calculer une mesure d'activité comparable à celles de TEP.

$$M^{t}_{AN_C} = M^{t-1}_{IGN_C} + M^{t-1}_{AN_C} \text{(eq. VI.1.2)}$$

Le nœud OGN_C modélise la sortie du cortex visuel. Il dépend du nœud d'entrée, IGN_C, et d'un nœud de seuil, FTN_C. OGN_C ne peut émettre que si le niveau d'activation reçu de IGN_C dépasse celui du seuil. Cette condition est exprimée sous la forme d'une fonction sigmoïdale croissante, dont le résultat constitue un coefficient multiplicatif de la magnitude issue d'IGN_C :

$$M^{t}_{OGN_C} = \left(\sigma\left(M^{t-1}_{IGN_C} - M^{t-1}_{FTN_C}\right) a^{(IGN_C)}_{OGN_C} M^{t-1}_{IGN_C}\right) + \left(a^{(OGN_C)}_{OGN_C} M^{t-1}_{OGN_C}\right) \text{(eq. VI.1.3)}$$

Le nœud FTN_C est influencé par la sortie de la structure thalamique, OGN_T, qui rabaisse son activation. En l'absence d'émission de la part d'OGN_T, la magnitude de FTN_C tend vers la valeur maximale arbitraire 1. L'activation de FTN_C étant modulée au cours du temps, on peut parler ici de seuil dynamique :

$$M^{t}_{FTN_C} = 1 - \left(a^{(FTN_C)}_{FTN_C}\left(1 - M^{t-1}_{FTN_C}\right) + a^{(OGN_T)}_{FTN_C} M^{t-1}_{OGN_T}\right) \text{(eq. VI.1.4)}$$

Quand OGN_C émet, il influence IGN_C et IN_C, qui sont des nœuds appartenant au cortex visuel, et également IGN_T, dans la structure thalamique. L'influence exercée sur IGNc a déjà été décrite auparavant. En ce qui concerne le nœud d'inhibition IN_C, l'action d'OGN_C est excitatrice. Par conséquent, lorsque OGN émet, il active le nœud d'inhibition qui vient lui-même baisser l'activation du nœud d'entrée IGNc. Cela revient à dire que la zone s'auto-inhibe après avoir émis. Ce mécanisme d'auto-inhibition allonge la période réfractaire de la zone.

$$M_{IN_C}^{t} = a_{IN_C}^{(OGN_C)} M_{OGN_C}^{t-1} + a_{IN_C}^{(IN_C)} M_{IN_C}^{t-1} \qquad \text{(eq. VI.1.5)}$$

La structure thalamique ne comprend que deux nœuds, un nœud d'entrée IGN_T et un nœud de sortie OGN_T. Le nœud d'entrée IGN_T est influencé directement par OGN_C, la sortie du cortex visuel, et il influence lui-même le nœud de sortie, OGN_T. La structure corticale relaie l'information, et ne réalise pas de traitement particulier :

$$M_{IGN_T}^{t} = M_{OGN_C}^{t-1} \qquad \text{(eq. VI.1.6)}$$

$$M_{OGN_T}^{t} = \left(a_{OGN_T}^{(IGN_T)} M_{IGN_T}^{t-1} \right) + \left(a_{OGN_T}^{(OGN_T)} M_{OGN_T}^{t-1} \right) \qquad \text{(eq. VI.1.7)}$$

Le nœud de sortie OGN_T a des répercussions sur FTN_C, le nœud de seuil du cortex visuel, qui influence lui-même OGN_C. La structure thalamique, OGN_C et FTN_C forment la boucle thalamo-corticale. En raison des différents délais associés aux liens entre les nœuds formant la boucle, celle-ci achemine vers FTN_C une valeur proche de l'activation d'OGN_C peu de temps avant. Le seuil dynamique est baissé en fonction de cette valeur, empêchant ou pas OGN_C d'émettre. Cette boucle implémente un mécanisme d'habituation par facilitation de l'émission.

Ce modèle est aussi proche du modèle BioCaEn original que possible, toutefois, certaines différences subsistent. Tout d'abord, le traitement des types, bien que manipulant également un couple numérique/symbolique, n'est pas le même dans BioCaEn et RAGE. Toutefois, ceci ne pose pas véritablement de problème dans le cas présent, où le type du stimulus n'a aucun impact sur la simulation (il n'y a qu'un seul type). De plus, à la différence de RAGE, BioCaEn ne permet pas de distinguer les valeurs d'activation et d'émission d'un nœud. Afin d'obtenir des résultats comparables, nous avons donc choisi d'utiliser l'identité comme fonction d'émission dans le modèle version RAGE. Cela revient à décrire l'état d'un nœud à un instant donné en utilisant une seule magnitude et un seul type (au lieu de deux magnitudes et deux types). Enfin, BioCaEn permet

de spécifier des conditions applicables aux relations entre les nœuds du modèle. Ces conditions sont exprimées sous la forme de formules logiques, évaluées dans une logique multivaluée. Elles interviennent ensuite dans les calculs sous la forme de gains multiplicatifs compris entre 0 et 1. Par exemple, l'influence de FTN_C sur OGN_C est modélisée par la condition $FTN_C < OGN_C$. Nous avons choisi d'exprimer ces conditions au moyen de sigmoïdes (c.f. chapitre V.3.2.1), qui interviennent également dans notre modèle en tant que coefficients multiplicatifs, dans les équations (eq. VI.1.1) et (eq. VI.1.3).

paramètre	équation	valeur
$a_{IGN_C}^{(Stim_C)}$	(eq. VI.1.1)	0,469
$a_{IGN_C}^{(IGN_C)}$		0,6872
$a_{IGN_C}^{(IN_C)}$		0,6254
$a_{OGN_C}^{(IGN_C)}$	(eq. VI.1.3)	0,648
$a_{OGN_C}^{(OGN_C)}$		0,7408

paramètre	équation	valeur
$a_{FTN_C}^{(FTN_C)}$	(eq. VI.1.4)	0,98
$a_{FTN_C}^{(OGN_T)}$		0,5
$a_{IN_C}^{(OGN_C)}$	(eq. VI.1.5)	0,4
$a_{IN_C}^{(IN_C)}$		0,423
$a_{OGN_T}^{(IGN_T)}$	(eq. VI.1.7)	0,9502
$a_{OGN_T}^{(OGN_T)}$		0,0498

Tableau VI.1.1 : valeurs des paramètres du modèle de boucle thalamo-corticale.

Les paramètres utilisés dans le modèle RAGE sont directement issus de ceux utilisés dans le modèle BioCaEn. Ceux-ci étaient eux-mêmes issus pour la plupart de données de la littérature en neurophysiologie et neuroanatomie. Certaines des valeurs utilisées pour RAGE diffèrent car elles ont dû être adaptées à la forme différente des fonctions utilisées dans RAGE.

1.3. Simulation et discussion

A partir du même modèle, une simulation a été réalisée pour chacune des fréquences de présentation utilisées pendant l'expérience réelle. Le pas d'échantillonnage choisi est 1 ms. Dans l'expérience réelle, chaque stimulus durait 5 ms, ce qui revient pour nous à présenter le stimulus pendant 5 instants. Le run durant 40 s, notre simulation dure 40000 instants.

Les valeurs obtenues sont comparables à des mesures de l'activité de populations de neurones, dans la localisation et dans le temps. A partir de ces valeurs, nous souhaitons obtenir des valeurs comparables aux mesures du rCBF données par la PET lors de l'expérience réelle. Pour cela, nous utilisons la méthode employée par Josette Pastor et coauteurs [Pastor *et al.* '00]. Dans un premier temps, considérons la magnitude du nœud AN_C à la fin de la simulation (instant 40000) : elle correspond à la somme des activations successives du cortex visuel au cours de la simulation. On la note :

$$\Delta rCBF_{sim} = M_{AN_C}^{t=40000} \qquad \text{(eq. VI.1.8)}$$

Ces valeurs doivent être ramenées à la même échelle que les valeurs de l'expérience, c'est à dire à un rapport activité sur repos. Pour cela, on utilise le rapport entre la somme des pourcentages expérimentaux $\Delta rCBF\%_{exp}$ et la somme de nos valeurs obtenues par simulation $\Delta rCBF_{sim}$. On obtient ainsi une valeur normalisée $\Delta rCBF\%_{sim}$, calculée par rapport à la même mesure de repos qui avait permis d'obtenir $\Delta rCBF\%_{exp}$ [Lafon '00; Pastor *et al.* '00] :

$$\Delta rCBF\%_{sim} = 100 \times \Delta rCBF_{sim} \times \frac{\sum_{runs} \Delta rCBF\%_{exp}}{\sum_{runs} \Delta rCBF_{sim}} \qquad \text{(eq. VI.1.9)}$$

On obtient une mesure comparable, qualitativement, à celle de l'expérience initiale. Les valeurs simulées puis normalisées laissent apparaître des résultats très proches de ceux observés expérimentalement par Fox et Raichle, c'est-à-dire une phase d'augmentation jusqu'à environ 7,8 Hz, puis une lente décroissance et enfin une stabilisation au delà de 33,1 Hz.

La Figure VI.1.12 montre que les valeurs issues des simulations, que ce soit celles de BioCaEn ou celle de RAGE, suivent une évolution identique à celle des valeurs expérimentales, ce qui corrobore la troisième hypothèse. Les valeurs représentées sont, pour BioCaEn, les valeurs centrales des intervalles numériques représentant la magnitude, et pour RAGE, les moyennes de la v.a. représentant la magnitude de AN_C. Il faut souligner que si les résultats sont très proches, en revanche la dispersion observée dans le cas de BioCaEn est plus élevée. Cette dispersion élevée est une des raisons pour laquelle nous avions abandonné la méthode qualitative.

Ce modèle montre l'intérêt de la *modélisation* des processus cérébraux : il permet, à partir de données d'activation dont la définition temporelle est 40 secondes, de réaliser une interprétation en termes

de réponse neuronale, à une échelle de l'ordre de la milliseconde. De plus, il illustre également l'avantage que constitue notre approche *explicite* de modélisation. Les hypothèses sont clairement formulées, et modélisées de façon explicite. Par exemple, l'auto-inhibition de IGN est représentée par le cycle $IGN_C \rightarrow OGN_C \rightarrow IN_C \rightarrow IGN_C$, et l'habituation par la boucle thalamo-corticale $OGN_C \rightarrow IGN_T \rightarrow OGN_T \rightarrow FTN_C \rightarrow OGN_C$.

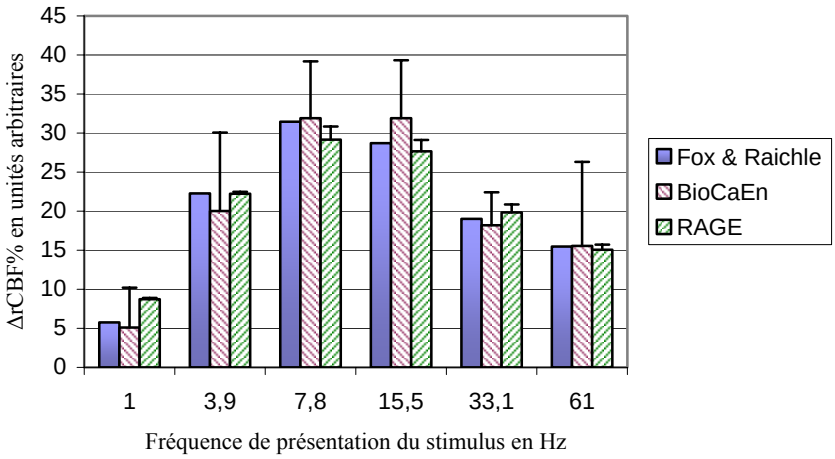

Figure VI.1.12 : résultats des simulations obtenues par BioCaEn et RAGE, comparés aux valeurs expérimentales de Fox & Raichle [Fox & Raichle '84, '85].

2. MODELE DE LA REPONSE DU GYRUS TEMPORAL SUPERIEUR DROIT LORS D'UNE TACHE PASSIVE DE CATEGORISATION

Le modèle présenté ici constitue un travail original, qui a fait l'objet de plusieurs publications [Labatut *et al.* '03a; Labatut *et al.* '03b]. Il repose sur une étude en activation menée par Serge Ruff et coauteurs [Ruff '00; Ruff *et al.* '01], dans notre groupe de recherche. L'étude cible une tâche de catégorisation phonémique passive. Le but est d'utiliser le mécanisme de traitement des types de RAGE pour représenter ce processus de catégorisation.

2.1. Etude en activation

A l'instar des expériences de Fox & Raichle, l'étude en activation de Ruff et coauteurs fait intervenir des processus automatiques. La tâche à effectuer est donc également passive, il s'agit d'entendre des stimuli auditifs. Le but de l'étude est de mettre en évidence des différences d'activation dans les aires corticales impliquées dans le traitement précoce des sons du langage, entre des sujets témoins et des sujets dyslexiques. L'hypothèse principale est que de telles différences pourraient refléter un dysfonctionnement dans ces aires, qui serait à la base des troubles de la perception catégorielle chez les dyslexiques [Ruff '00].

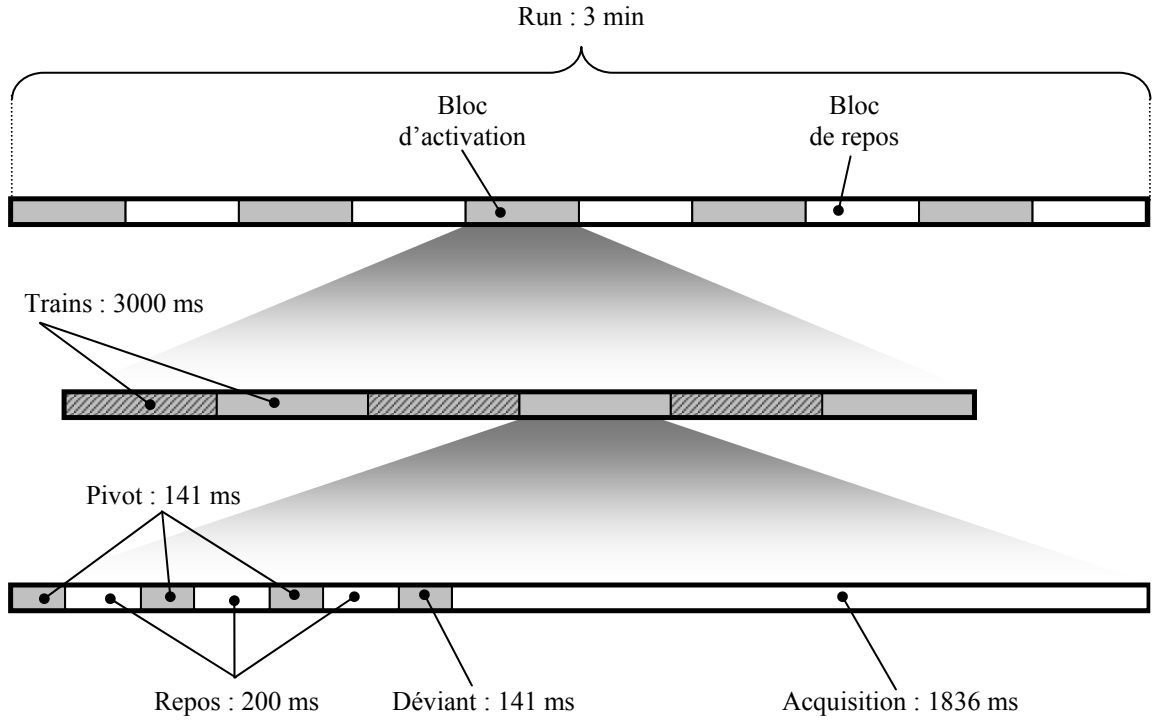

Figure VI.2.1 : organisation d'un run.

Il y a deux groupes de sujets, l'un composé de 6 individus dyslexiques et l'autres de 6 témoins. Chaque sujet est placé dans un imageur IRMf, et est soumis à une stimulation auditive délivrée par un dispositif électroacoustique spécifique, adapté à l'environnement magnétique propre à l'IRM. La durée totale du protocole est de 15 minutes, qui se décomposent en 5 *runs* de 3 minutes chacun (Figure VI.2.1). Un run est lui-même constitué de 10 *blocs*. On distingue deux types de blocs : les *blocs d'activation* et les *blocs de repos*, qui sont alternés au sein du run. Comme son nom l'indique, un bloc de repos est un laps de temps pendant lequel le sujet ne doit rien faire. Un bloc d'activation se compose de 6 *trains* de stimuli. Un train dure 3000 ms, il s'agit d'une série de 4 stimuli. Chaque

stimulus dure 141 ms et est séparé du suivant par une période de 200 ms, à l'exception du dernier stimulus, qui est suivi d'une période de 1836 ms. C'est pendant cette période que l'acquisition est effectuée, c'est-à-dire que c'est à ce moment-là que l'on mesure l'activité cérébrale du sujet.

Les stimuli sont des mixages de deux syllabes différentes, /pa/ et /ta/, sélectionnés dans un continuum de sons allant du /pa/ pur au /ta/ pur. Ces mixages sont arrangés de façon à avoir tous la même durée et la même intensité. Les syllabes /pa/ et /ta/ ont été choisies pour la netteté de l'effet de perception qu'elles provoquent. Chaque sujet a été testé avant l'étude en activation, afin de déterminer sa *frontière catégorielle*, c'est-à-dire le son qui marque la limite entre les sons catégorisés comme étant des /pa/, et ceux catégorisés comme étant des /ta/. Il faut souligner que la frontière catégorielle ne correspond pas à l'élément médian du continuum (50% /pa/ – 50% /ta/), mais qu'elle est située plus près d'un /pa/ pur que d'un /ta/ pur.

A partir de la frontière moyenne, on détermine un son appelé *pivot*, qui est nettement catégorisé comme un /ta/. Le pivot est suffisamment près (en termes de distance acoustique) de la frontière catégorielle, et en même temps suffisamment éloigné pour éviter qu'un effet d'habituation ne le fasse percevoir comme un /pa/ après plusieurs répétitions. A partir du pivot, on définit deux sons, appelés *déviants*, et situés à une égale distance d1 du pivot. L'un des déviants se rapproche plus d'un /ta/ pur, et l'autre est catégorisé comme un /pa/. Comme le premier est situé dans la même catégorie que le pivot, on parle de déviant cis-catégoriel, alors que le second est dit trans-catégoriel. On les note respectivement *cis-d1* et *trans-d1*. On réalise la même opération en prenant une distance d2 plus importante, et on obtient les déviants *cis-d2* et *trans-d2*.

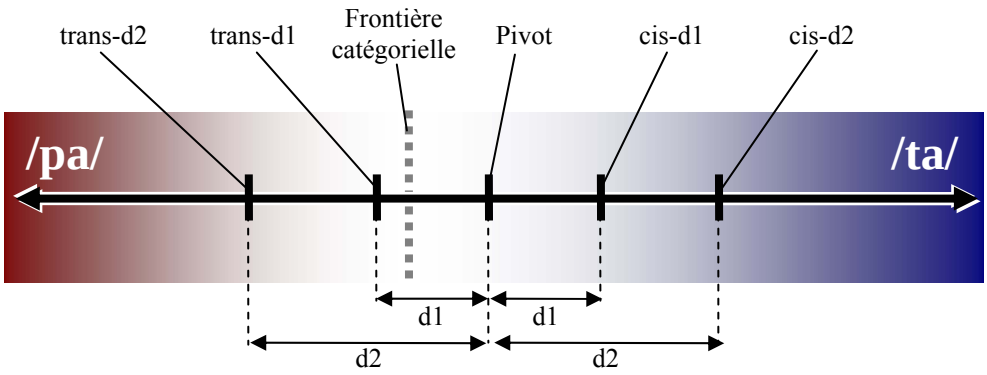

Figure VI.2.2 : disposition des différents mixages de syllabes utilisés pour la stimulation.

Chacun des 5 blocs d'activation d'un run est dédié à l'un des 5 sons décrits précédemment (le pivot et les 4 déviants). Dans le bloc standard, appelé *Dev0*, tous les trains sont composés uniquement du

pivot. Dans chacun des quatre autres blocs, le dernier son de 3 des 6 trains est remplacé par un déviant. On obtient les blocs *Dev2P*, *Dev1P*, *Dev1M* et *Dev2M*, respectivement, pour les déviants trans-d2, trans-d1, cis-d1 et cis-d2.

Les mesures d'IRMf donnent la variation du débit sanguin cérébral régional. On observe des activations temporales bilatérales (cortex auditifs primaire et associatif), avec une domination du côté droit. La comparaison des résultats des deux groupes de sujets révèle des différences d'activation entre les témoins et les sujets dyslexiques. On observe chez les témoins un effet catégoriel (influence de la catégorie sur l'activation) particulièrement marqué dans l'hémisphère droit, et inexistant chez les dyslexiques. De ce point de vue, le gyrus temporal supérieur droit fait partie des zones d'intérêt. Il s'agit d'une structure appartenant au cortex auditif associatif, impliquée dans le traitement précoce des stimuli verbaux. La Figure VI.2.3 présente les mesures d'activation obtenues pour cette région, exprimées en unités arbitraires en fonction de la condition (Dev2M, Dev1M, Dev0, Dev1P ou Dev2P). La droite horizontale en pointillés représente le niveau d'activation au repos.

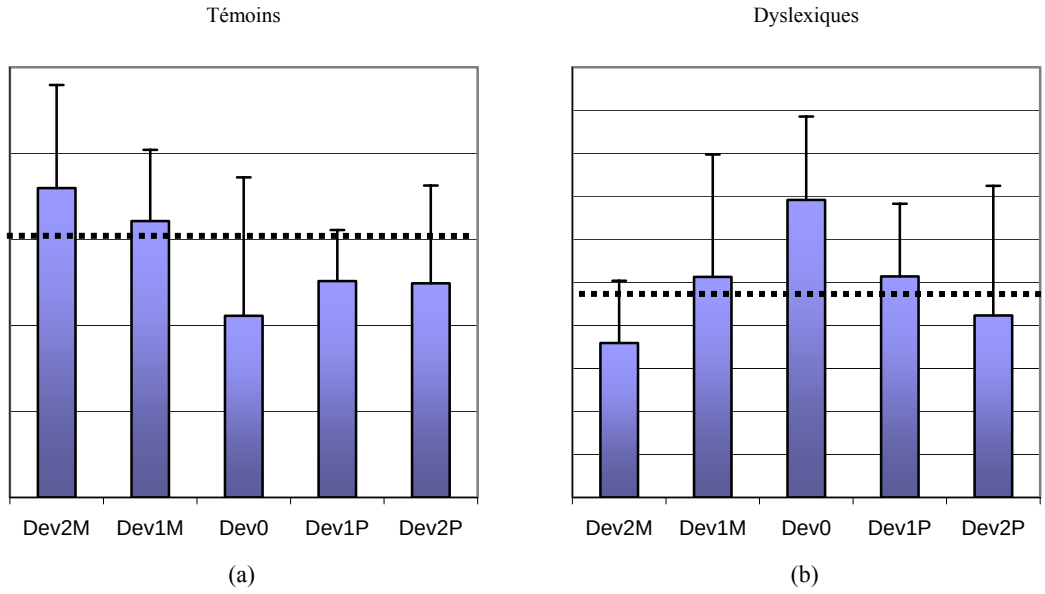

Figure VI.2.3 : activation moyenne du gyrus temporal supérieur droit pour chacun des 5 types de blocs [Ruff '00].
(a) : moyenne sur le groupe des témoins. (b) : moyenne sur le groupe des dyslexiques.

Ruff et coauteurs font remarquer que le groupe des témoins présente une hausse de l'activation par rapport au repos, pour les blocs dédiés aux déviants trans-catégoriels (Dev2M et Dev1M). Au contraire, on observe une nette baisse d'activation pour Dev0 et le blocs à déviant cis-catégoriel

(Dev1P et Dev2P). Dans le cas des dyslexiques, l'activation par rapport au repos est forte pour le pivot, puis elle est d'autant plus faible que le déviant employé est éloigné du pivot.

Ruff et coauteurs formulent plusieurs hypothèses pour expliquer leurs observations. Ils supposent, entre autres, que la faible activation chez les témoins en Dev0 serait due à un phénomène d'habituation : c'est toujours le même son (le pivot) qui est répété. En présence de déviants cis-catégoriel, on observe également une activation en dessous de celle mesurée au repos, mais plus élevée toutefois qu'en Dev0. Ceci ne serait le résultat de la sensibilité spécifique de la région à l'effet catégoriel, qui est également illustrée par la hausse d'activation observée pour les déviants trans-catégoriels. En ce qui concerne les sujets dyslexiques, l'absence d'habituation pour Dev0 laisse penser que les informations traitées par l'aire associative considérée lui apparaissent comme plus variées que ce n'est le cas pour les témoins. A cause de ce trouble de l'intégration, le traitement ne pourrait aboutir facilement au rapprochement entre un stimulus et un prototype mémorisé.

2.2. Modèle de gyrus temporal supérieur droit

L'étude corrobore l'hypothèse de départ, à savoir l'existence d'un dysfonctionnement des aires corticales impliquées dans le traitement précoce des stimuli auditifs, chez les dyslexiques. Elle montre également que le gyrus temporal supérieur droit (GTSD) était sensible aux différences catégorielles chez les témoins, mais pas chez les dyslexiques. On peut supposer que les différences observées entre les niveaux d'activation des témoins et des sujets dyslexiques résultent de différences dans les mécanismes de traitement de l'information cérébrale implémentés par le GTSD.

Le GTSD fait apparemment partie du réseau à grande échelle sollicité par la tâche étudiée, mais nous nous sommes concentrés uniquement sur cette région d'intérêt en particulier. Nous avons construit un modèle du fonctionnement du GTSD chez le sujet sain et chez le sujet dyslexique. Sa structure et son fonctionnement sont basés sur deux hypothèses principales. Nous supposons, tout d'abord, que le GTSD se compose de plusieurs processeurs d'information, dédiés chacun à un phonème spécifique, et entretenant des rapports de compétition avec les autres processeurs. Comme on ne peut pas les localiser, nous ne pouvons pas les représenter par des nœuds structurels, mais uniquement pas des nœuds fonctionnels. La deuxième hypothèse concerne la différence de traitement entre les sujets témoins et les dyslexiques : nous supposons que la seule différence de traitement réside dans l'absence de compétition entre les processeurs chez le dyslexique.

La Figure VI.2.4 décrit le modèle statique de GTSD. Les liens structurels sont représentés en gras. La définition des processeurs repose sur l'hypothèse de généricité, décrite dans le chapitre IV.1.1.3, qui stipule que des aires corticales peuvent partager certaines propriétés portant sur la structure du réseau cérébral sous-jacent et sur la nature fonctionnelle des nœuds qui le composent. A ce titre, chaque processeur est représenté par un réseau fonctionnel adapté du modèle de cortex visuel décrit et utilisé auparavant pour modéliser l'expérience de Fox & Raichel.

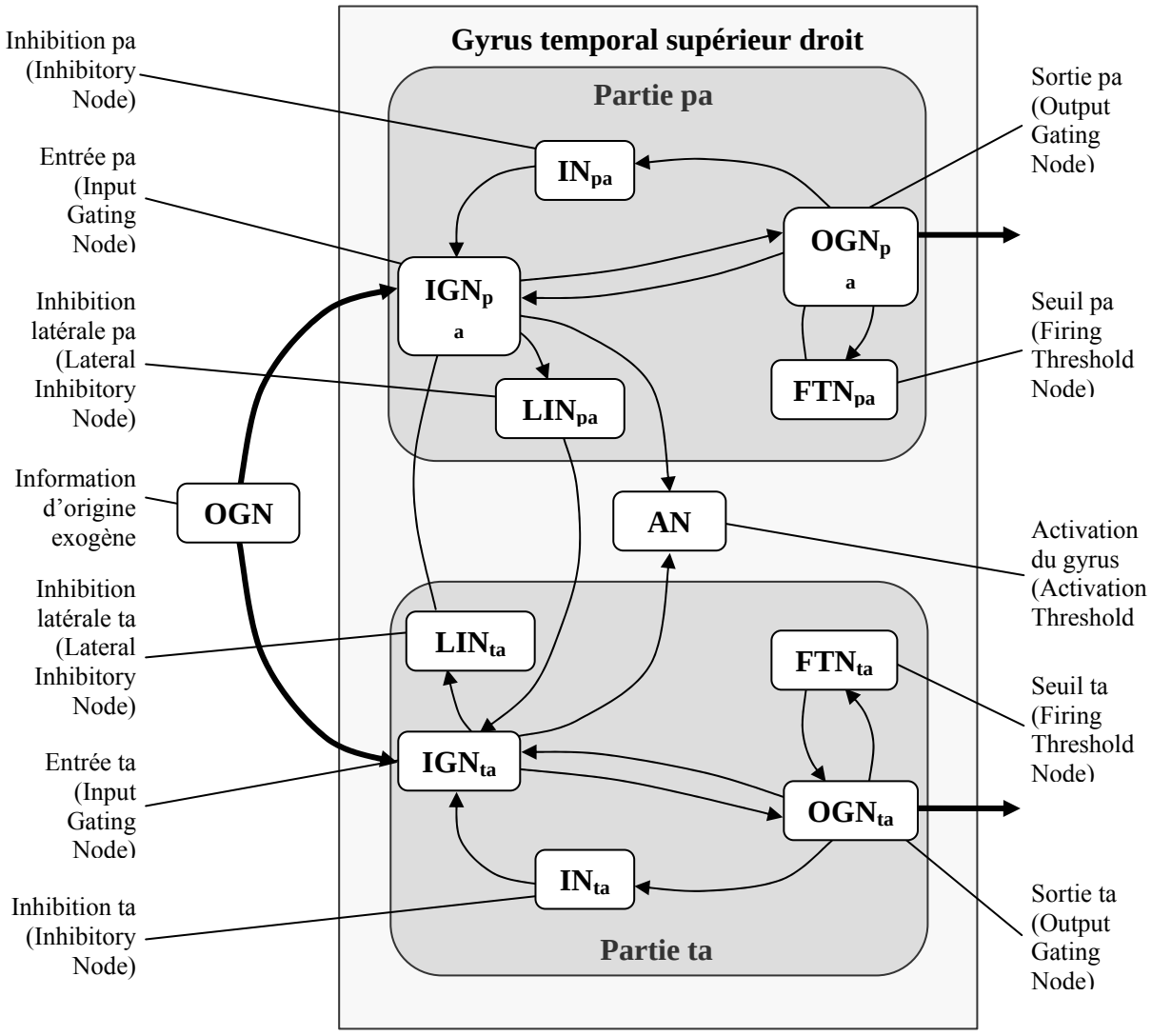

Figure VI.2.4 : modèle de gyrus temporal supérieur droit destiné à reproduire un processus de catégorisation automatique.

Comme l'expérience de Ruff et coauteurs n'implique que deux syllabes différentes, nous n'avons inclus que deux processeurs dans notre modèle, l'un est dédié à la syllabe /pa/, et l'autre à la syllabe /ta/. Les deux processeurs sont exactement identiques, à l'exception de leurs entrées, dont les sensibilités aux syllabes sont différentes. Pour cette raison, nous ne présentons que les équations du

processeur dédié à /pa/, celles du processeur dédié à /ta/ s'obtenant simplement en inversant les indices des nœuds dans les formules.

Il faut souligner que ce modèle a été créé à un stade où le développement de notre formalisme n'incluait pas encore de distinction entre les valeurs d'activation et les valeurs d'émission d'un nœud. Comme dans BioCaEn et dans le modèle de cortex visuel, nous ne présentons donc, pour chaque nœud, qu'une seule fonction pour modéliser le traitement et la propagation de l'information cérébrale.

L'information arrive d'une aire ayant déjà réalisé un prétraitement de l'information, et qui se trouve donc située en amont dans le réseau à grande échelle contenant le GTSD. C'est le nœud *OGN* (la sortie de l'aire en question) qui représente l'information entrant dans notre modèle. Comme dans le modèle de cortex visuel, la liaison entre OGN et les processeurs est de type structurel, ce qui implique un délai de transmission plus long que les autres relations (ici, deux fois plus long). La magnitude de OGN est fixée à 1 en présence d'un stimulus, et 0 en l'absence de stimulus. En ce qui concerne le type, il est défini sur un domaine simple composé de deux symboles, l'un représentant la syllabe pure /ta/ et l'autre modélisant un /pa/ pur :

$$D_{OGN} = \{pa; ta\} \qquad \text{(eq. VI.2.1)}$$

Les cinq différents mixages utilisés dans l'étude en activation sont obtenus en faisant varier les poids associés à ces symboles, comme décrit dans le Tableau VI.2.1.

Le nœud IGN_{pa} a le même rôle qu'IGN_C dans le modèle précédent, à savoir réaliser l'intégration des entrées de son processeur. L'équation permettant de calculer la magnitude a également une forme similaire :

$$\begin{aligned} M^{t}_{IGN_{pa}} = &\left(\sigma\left(M^{t-1}_{OGN_{pa}}\right) a^{(Stim)}_{IGN_{pa}} M^{t-2}_{OGN}\, f\left(simax\right)\right) + \left(a^{(IGN_{pa})}_{IGN_{pa}} M^{t-1}_{IGN_{pa}}\right) \\ &- \left(a^{(IN_{pa})}_{IGN_{pa}} M^{t-1}_{IN_{pa}}\right) - \left(a^{(LIN_{ta})}_{IGN_{pa}} M^{t-1}_{LIN_{ta}}\right) \end{aligned} \qquad \text{(eq. VI.2.2)}$$

La fonction sigmoïdale appliquée à la magnitude issue de OGN_{pa} permet toujours de modéliser la période réfractaire du processeur. IGN_{pa} subit une influence supplémentaire par rapport au modèle utilisé pour Fox et Raichle, il s'agit de celle, négative, du nœud LIN_{ta}, qui représente l'inhibition que le processeur dédié à /ta/ exerce sur son voisin. Au niveau du traitement des types, chaque IGN est

doté d'une TPT différente (tableau). Ce paramètre constitue la seule différence entre les deux sous-réseaux implémentant les processeurs.

nom	poids associé à *pa*	poids associé à *ta*
trans-d2	0,7	0,3
trans-d1	0,55	0,45
pivot	0,4	0,6
cis-d1	0,25	0,75
cis-d2	0,1	0,9

Tableau VI.2.1 : types utilisés pour modéliser les 5 différents stimuli.

nœud	archétype		préférence	symbole associé
	poids associé à *pa*	poids associé à *ta*		
IGN_{pa}	0,8	0,2	1	pa
IGN_{ta}	0,2	0,8	1	ta

Tableau VI.2.2 : TPT des nœuds d'entrée (IGN) des deux processeurs.

Chaque TPT ne contient qu'un seul archétype, défini sur le même domaine que le type provenant du nœud OGN. L'archétype d'IGN_{pa} est caractérisé par une forte proportion de symbole *pa*, alors que pour celui d'IGN_{ta}, c'est le poids associé à *ta* qui est le plus élevé. Comme il n'y a qu'un seul paramètre, la préférence associée est 1 (en raison de la contrainte de sommation des préférences, partie V.2.2.1 (eq. V.2.1)). Comme il n'y a qu'un seul stimulus en entrée, aucun problème de combinaison de types ne se pose. A noter que les types émis par les nœuds IGN possèdent le même domaine de définition que le type reçu de OGN : il n'y a pas combinaison de plusieurs informations hétérogènes, donc il est inutile de changer de domaine de définition. Dans ce modèle, l'intérêt de la partie qualitative de l'information cérébrale se limite au filtrage effectué par le nœud

d'entrée, le reste du traitement portant essentiellement sur la partie quantitative de l'information. Pour cette raison, nous ne présenterons par la suite que les équations destinées au traitement des magnitudes.

OGN_{pa} est la sortie du processeur dédié à /pa/. Dans un modèle englobant d'autres régions, il pourrait être relié à des nœuds situés en aval. Ce n'est pas le cas ici puisque notre modèle se limite au GTSD. Il a le même rôle que OGN_C dans le modèle précédent, et la fonction est similaire :

$$M_{OGN_{pa}}^{t} = \left(\sigma\left(M_{IGN_{pa}}^{t-1} - M_{FTN_{pa}}^{t-1}\right)a_{OGN_{pa}}^{(IGN_{pa})}M_{IGN_{pa}}^{t-1}\right) + \left(a_{OGN_{pa}}^{(OGN_{pa})}M_{OGN_{pa}}^{t-1}\right) \qquad \text{(eq. VI.2.3)}$$

Le nœud FTN a toujours un rôle de seuil dynamique. Toutefois, il n'y a pas de boucle thalamique dans le présent modèle, au contraire du modèle précédent. Le FTN modélise ici une population non-identifiée qui joue le même rôle que l'ensemble IGN_T-OGN_T-FTN_C du modèle précédent. Pour cette raison, le délai de transmission est plus long entre OGN_{pa} et FTN_{pa} (deux fois plus long que pour les autres liaisons) ce qui permet d'implémenter quand même un mécanisme d'habituation.

$$M_{FTN_{pa}}^{t} = 3 - \left(a_{FTN_{pa}}^{(FTN_{pa})}\left(3 - M_{FTN_{pa}}^{t-1}\right) + a_{FTN_{pa}}^{(OGN_{pa})}M_{OGN_{pa}}^{t-2}\right) \qquad \text{(eq. VI.2.4)}$$

Le nœud IN_{pa} est identique à son homologue IN_C :

$$M_{IN_{pa}}^{t} = a_{IN_{pa}}^{(OGN_{pa})}M_{OGN_{pa}}^{t-1} + a_{IN_{pa}}^{(IN_{pa})}M_{IN_{pa}}^{t-1} \qquad \text{(eq. VI.2.5)}$$

Le nœud d'inhibition latérale LIN_{pa} est nouveau par rapport au modèle précédent. Il est utilisé pour représenter explicitement le mécanisme de compétition entre les processeurs. Il s'agit d'un nœud influencé par IGN_{pa}, et qui a lui-même des répercussions négatives sur l'entrée LIN_{ta} du processeur concurrent. Le principe de calcul est le même que pour IN_{pa} :

$$M_{LIN_{pa}}^{t} = a_{LIN_{pa}}^{(IGN_{pa})}M_{IGN_{pa}}^{t-1} + a_{LIN_{pa}}^{(LIN_{pa})}M_{LIN_{pa}}^{t-1} \qquad \text{(eq. VI.2.6)}$$

Le nœud d'activation AN représente l'activation de tout le GTSD, telle qu'elle est mesurée par l'IRMf. Son échelle temporelle est donc de l'ordre de la seconde, alors que le reste du réseau fonctionne à l'échelle de la dynamique neuronale, de l'ordre de la milliseconde.

$$M^{t}_{AN_{pa}} = M^{t-1}_{IGN_{pa}} + M^{t-1}_{IGN_{ta}} + M^{t-1}_{AN_{pa}} \qquad \text{(eq. VI.2.7)}$$

A l'instar de la structure des réseaux implémentant les processeurs, les paramètres initiaux sont également issus du modèle de cortex visuel présenté précédemment. Toutefois, ce dernier présente quelques différences vis-à-vis du modèle de GTSD. Tout d'abord, le pas d'échantillonnage utilisé ici est de 2 ms (au lieu de 1 ms pour le modèle de cortex visuel). D'autre part, l'information traitée est de nature auditive, alors qu'il s'agissait de stimuli visuels dans le premier modèle. Enfin, la technique d'imagerie employée n'est pas la même : IRMf au lieu de TEP. Pour ces raisons, certains paramètres ont du être modifiés, au niveau du traitement de la magnitude comme à celui des types. Ces modifications ont été accomplies par affinement manuel des paramètres au moyen de simulations successives.. Les paramètres obtenus sont présentés dans les Tableau VI.2.3 et Tableau VI.2.4.

paramètre	équation	valeur
$a^{(Stim)}_{IGN_{pa}}$	(eq. VI.2.2)	0,6
$a^{(IGN_{pa})}_{IGN_{pa}}$		0,98
$a^{(IN_{pa})}_{IGN_{pa}}$		0,2
$a^{(IGN_{pa})}_{OGN_{pa}}$	(eq. VI.2.3)	0,4
$a^{(OGN_{pa})}_{OGN_{pa}}$		0,6

paramètre	équation	valeur
$a^{(IGN_{pa})}_{LIN_{pa}}$	(eq. VI.2.6)	0,8
$a^{(LIN_{pa})}_{LIN_{pa}}$		0,1
$a^{(FTN_{pa})}_{FTN_{pa}}$	(eq. VI.2.4)	0,995
$a^{(OGN_{pa})}_{FTN_{pa}}$		0,005

Tableau VI.2.3 : valeurs des paramètres identiques pour le modèle témoin et le modèle dyslexique.

Les modèles utilisés pour simuler un sujet sain et un sujet dyslexique sont exactement les mêmes en ce qui concerne la structure et les fonctions utilisées. La seule différence réside dans la valeur de certains paramètres, même si la plupart sont identiques dans les deux cas (Tableau VI.2.3). Trois seulement sont différents (Tableau VI.2.4), il s'agit du paramètre définissant la force de l'inhibition exercée par un processeur sur l'autre, et des deux paramètres utilisés pour calculer les valeurs des inhibitions internes des processeurs.

Nous avons vu précédemment que les inhibitions latérales représentaient les mécanismes de compétition entre processeurs. La différence au niveau du premier paramètre représente

explicitement notre hypothèse concernant l'absence de mécanisme compétitif chez le dyslexique, puisque l'influence des LIN sur les IGN est réduite à 0 (Tableau VI.2.4). La différence de valeur des deux autres paramètres est une répercussion de cette absence d'inhibition latérale, qui se traduit par une inhibition interne plus importante.

paramètre	équation	valeur	
		témoins	dyslexiques
$a_{IGN_{pa}}^{(LIN_{ta})}$	(eq. VI.2.2)	0,8	0
$a_{IN_{pa}}^{(OGN_{pa})}$	(eq. VI.2.5)	0,1	0,005
$a_{IN_{pa}}^{(IN_{pa})}$		0,8	0,995

Tableau VI.2.4 : valeurs des paramètres différents pour le modèle témoin et le modèle dyslexique.

2.3. Simulation et discussion

A partir du même réseau, nous avons deux ensembles de paramètres qui définissent un modèle de sujet témoin pour l'un et dyslexique pour l'autre. Sur ces deux modèles, nous avons réalisé des simulations en utilisant les 5 types (Tableau VI.2.1) correspondant aux différents mixages employés dans l'étude en activation. Au total, 10 simulations ont donc eu lieu. Comme les résultats de nos simulations sont destinés à être comparés à des valeurs moyennées à la fois sur des sujets et sur des blocs (c.f. la description du protocole, pour la définition d'un bloc), il n'est pas nécessaire de simuler un bloc entier (soit 6 trains) pour ensuite faire la moyenne. Pour chaque bloc, nous avons choisi de ne simuler qu'un seul train de syllabes : les trois pivots, suivis du son qui varie suivant le bloc (un des quatre déviants ou le pivot).

Le pas d'échantillonnage étant 2 ms, chaque syllabe du train est présentée pendant 70 instants (soit 140 ms) et est suivie de 100 itérations (soit 200 ms) de repos. La magnitude associée à un type de syllabe est constante, et de valeur 1. Pendant la période de repos, le type représente du bruit et la magnitude est nulle. Une simulation représente donc au total 680 itérations (1360 ms).

Comme précédemment (pour le modèle de cortex visuel), on ne peut pas directement comparer directement les valeurs obtenues par simulation aux données issues de l'IRMf. On ne peut pas non plus effectuer le même traitement, car on veut cette fois comparer l'activation correspondant à chaque bloc à l'activation au repos. Nous considérons que la valeur de la magnitude de AN à la fin de la simulation (instant 680) correspond à l'activation globale du GTSD pour le train simulé. Pour la condition • (i.e. Dev2M, Dev1M, Dev0, Dev1P ou Dev2P), on la note $Mag_{\bullet}$.

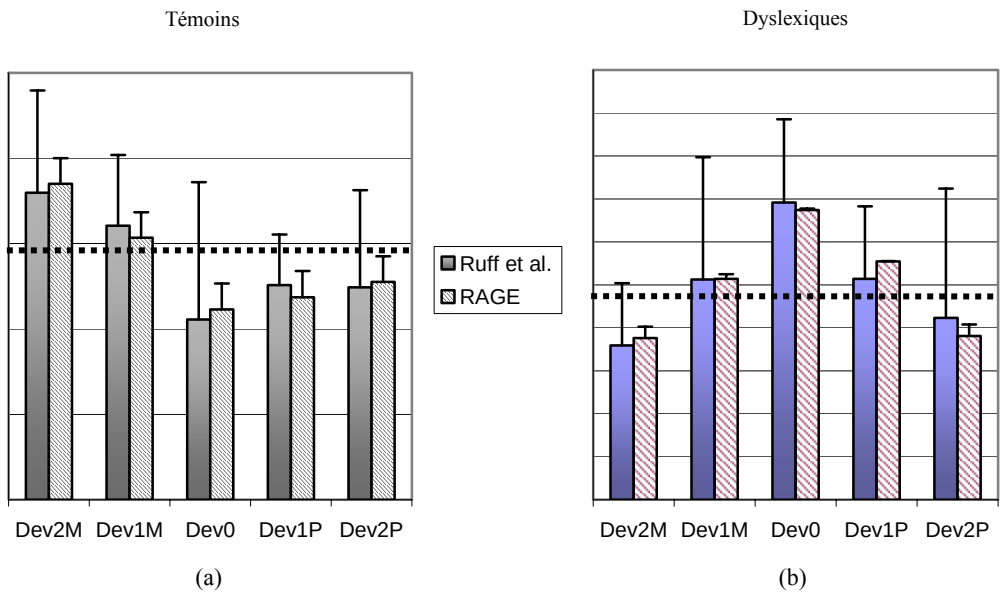

Figure VI.2.5 : comparaison des activations expérimentales et simulées.
(a) : pour les témoins. (b) : pour les dyslexiques.

La difficulté est que nous ne disposons d'une mesure d'activité du GTSD au repos que dans le cas expérimental. Il faut donc se ramener à des valeurs simulées comparables au repos expérimental. On calcule pour cela la moyenne et l'écart type, pour les cinq valeurs simulées $Mag_{\bullet}$ et pour les valeurs expérimentales $\Delta rCBF^{\bullet}_{exp}$. On obtient respectivement μ_{sim} et σ_{sim} dans le premier cas, et μ_{exp} et σ_{exp} dans le second. Puis on effectue l'opération suivante pour chaque valeur $Mag_{\bullet}$:

$$\Delta rCBF^{\bullet}_{sim} = \left(Mag_{\bullet} - \mu_{sim}\right)\frac{\sigma_{exp}}{\sigma_{sim}} + \mu_{exp} \qquad \text{(eq. VI.2.8)}$$

Grâce à cette opération, on obtient un ensemble de valeurs $\Delta rCBF^{\bullet}_{sim}$ de même moyenne et de même écart-type que l'ensemble de valeurs $\Delta rCBF^{\bullet}_{exp}$. En considérant ces valeurs relativement à

l'activation au repos expérimentale, on obtient la Figure VI.2.5, qui montre les niveaux d'activation moyens et les écarts types.

Les valeurs obtenues par simulation sont très proches des valeurs expérimentales, et qui plus est, elles suivent la même évolution. Dans le cas du modèle de témoin, on observe une hypoactivation en Dev0, une hypoactivation moins importante pour les déviants cis-catégoriels, et une hyperactivation pour les déviants trans-catégoriels : on retrouve donc le phénomène d'habituation en Dev0 et l'effet catégoriel. Dans le cas du modèle de dyslexique, l'activation décroît en fonction de la distance entre le quatrième son du train de stimuli et le pivot.

Nous avons vu que l'un des intérêts de la modélisation des processus cérébraux était d'étudier la dynamique neuronale. Dans le cas présent, nous pouvons considérer l'activation du nœud d'entrée en fonction du stimulus. La Figure VI.2.6.a décrit l'évolution du niveau d'activation des nœuds IGN_{pa} et IGN_{ta} dans le modèle témoin, avec un train de stimuli comportant un déviant trans-catégoriel en dernière position (condition Dev2M). On voit très bien que pour les 3 premiers stimuli (qui sont des pivots, donc plutôt des /ta/), l'entrée du processeur dédiée à /ta/ s'active nettement (pointillés), ce qui n'est pas le cas pour l'autre processeur (trait plein). Par contre, le dernier stimulus (un /pa/) déclenche une forte activation du processeur /pa/, alors que l'activation du processeur /ta/ est négligeable, car très rapidement écrasée par l'inhibition latérale. Pour la condition Dev1M (non représentée ici), l'évolution des activations est similaire, avec un niveau moins élevé, toutefois, pour le déviant trans-d1.

Dans la Figure VI.2.6.b, on utilise le train de stimuli défini par la condition Dev0, c'est-à-dire une séquence de 4 pivots. On peut dire, ici, que seul IGN_{ta} s'active. Il est possible d'observer l'effet du phénomène d'activation, puisque l'activation d'iGN_{ta} va en décroissant, ce qui explique l'activation globale au GTSD plus faible (en tenant compte de IGN_{pa} et IGN_{ta}) que celle observée dans la Figure VI.2.6. Pour les déviants Dev1P (non-représenté) et Dev2P (Figure VI.2.6.c), c'est-à-dire pour les déviants cis-catégoriels, l'activation correspondant à la dernière syllabe est légèrement plus élevée que ce n'est le cas pour Dev0. On peut supposer que, le dernier stimulus étant proche, mais différent, des trois premiers, l'effet d'habituation ne joue pas à plein.

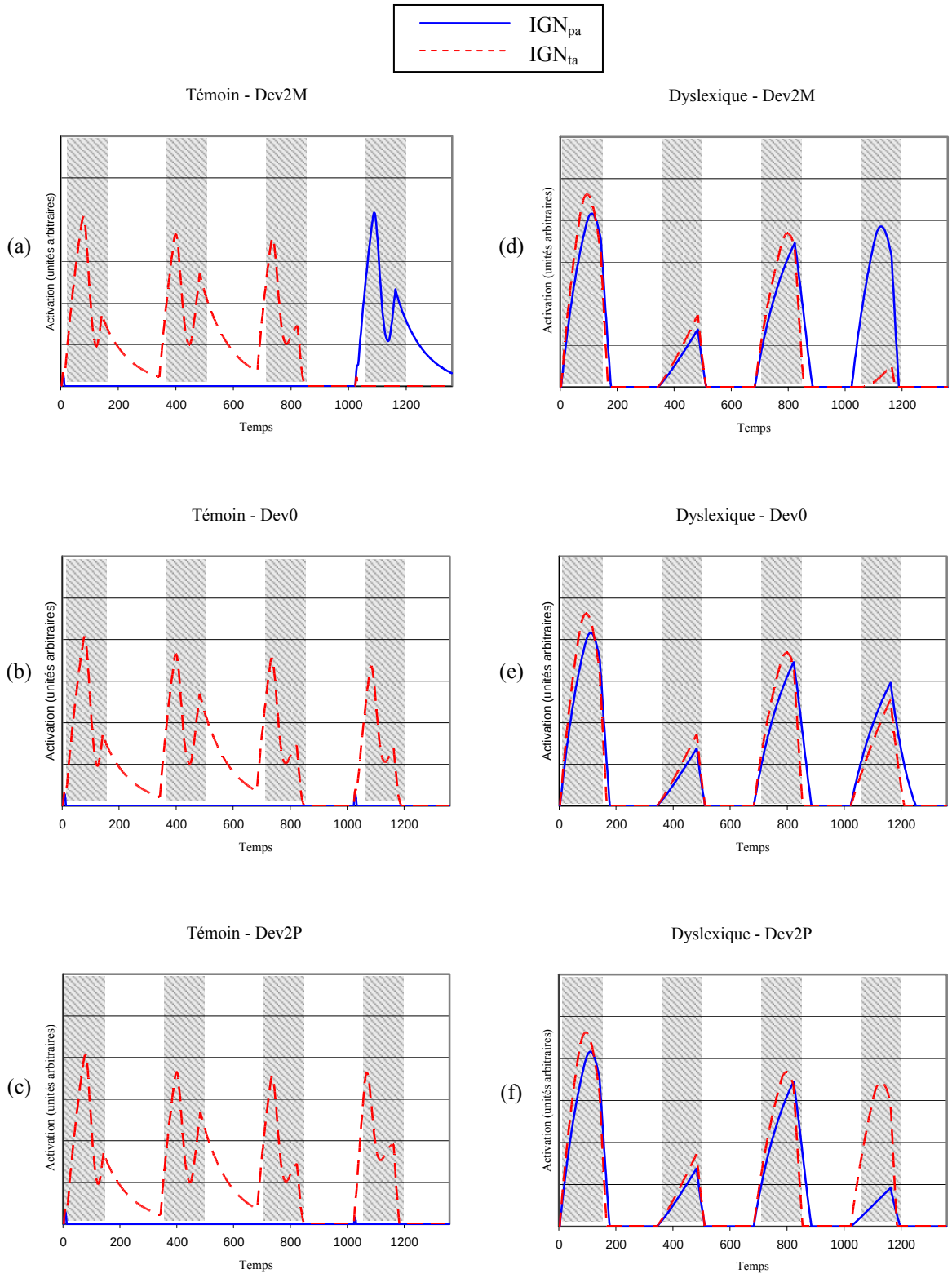

Figure VI.2.6 : évolution du niveau d'activation des nœuds d'entrée IGN_{pa} et IGN_{ta} dans les modèles témoin et dyslexique.
Les barres verticales grisées représentent les stimuli.

Intéressons-nous maintenant au modèle de dyslexique, pour lequel la Figure VI.2.6.d montre l'évolution des niveaux d'activation de IGN_{pa} et IGN_{ta} lorsque la condition est Dev2M. Cette fois, les deux entrées s'activent toutes les deux pendant toute la simulation. Quand c'est le pivot qui est présenté, on observe des activations presque identiques, celle d'IGN_{pa} étant légèrement inférieure. Ceci est du au fait que le pivot se rapproche plus d'un /ta/ pur que d'un /pa/ pur, et est donc mieux reconnu par le processeur dédié à /ta/. Lorsque c'est trans-d2 qui est présenté, on observe une franche domination de IGN_{pa}, pour les mêmes raisons. Dans le cas de l'utilisation du déviant trans-d1 (non représenté ici), le modèle se comporte de la même façon, avec une activation moins importante de IGN_{pa}, et légèrement plus importante de IGN_{ta}. C'est toutefois IGN_{pa} qui domine toujours largement.

La Figure VI.2.6.e montre les activations des nœuds d'entrée du modèle dyslexique soumis à la condition Dev0. Les activations des deux entrées suivent la même évolution. On observe une légère décroissance, comparable à celle que, dans la Figure VI.2.6.b, nous avons assimilée à la manifestation d'un mécanisme d'habituation.

La Figure VI.2.6.f correspond à la condition Dev2P (déviant cis-catégoriel). Elle présente une évolution symétrique à celle observée pour Dev2M, dans le sens où si les trois premières syllabes provoquent des activations identiques à celles de la Figure VI.2.6.d, en revanche, le déviant provoque ici une activation plus forte chez IGN_{ta} que chez IGN_{pa}.

L'interprétation plus générale que l'on peut faire de ces résultats est que dans le cas du témoin, la compétition entre les deux processeurs entraîne une baisse de l'activation globale du GTSD d'autant plus forte que le stimulus d'entrée est proche des archétypes caractérisant les deux TPT. C'est pourquoi l'activation la plus faible est observée pour Dev0. Au contraire, un stimulus distant ne va activer qu'un des deux processeurs, mais très fortement, et l'activation globale sera plus élevée que quand les deux processeurs s'activent faiblement. A ce mécanisme vient s'ajouter le phénomène d'habituation, dont l'influence pourrait se résumer par : plus la dernière syllabe est éloignée du pivot, plus l'effet de surprise est grand, brisant le phénomène d'habituation, et conduisant à une activation d'autant plus grande. Ceci s'explique par le fait que le phénomène d'habituation baisse progressivement l'activation d'IGN_{ta}, et donc du même coup l'inhibition que le processeur /ta/ exerce sur le processeur /pa/. Par conséquent, à la première apparition d'un stimulus propre à déclencher ce dernier, l'activation d'IGN_{pa} n'est que très peu réduite par l'inhibition latérale.

Dans le cas du dyslexique, l'absence de compétition entraîne systématiquement l'activation des deux processeurs. La domination de l'un sur l'autre dépend exclusivement du type de stimulation et de la TPT des nœuds d'entrée. Par conséquent, un type proche du pivot est susceptible de déclencher une activation plus élevée qu'un type proche d'une des deux syllabes pures. En d'autres termes, en l'absence de compétition, l'activation globale est plus importante quand les deux processeurs s'activent un peu, que quand un seul s'active beaucoup (à la différence de ce qui avait été observé pour le modèle témoin).

Ce modèle montre plusieurs points forts de RAGE. Tout d'abord, la possibilité de réutiliser des parties d'anciens modèles, et de les adapter pour en faire de nouveaux. Comme c'était déjà le cas pour le modèle de cortex visuel, ce modèle permet d'interpréter en termes de dynamique neuronale des résultats de neuroimagerie, d'échelles temporelles différentes. De plus, notre approche explicite de modélisation nous permet d'implémenter clairement nos hypothèses. Chaque processeur est modélisé par un sous-réseau bien distinct, et le mécanisme de compétition est représenté par les inhibitions latérales. Enfin, par rapport au modèle de cortex visuel, nous introduisons ici l'utilisation des types dans un but de catégorisation. Bien que simple (nous n'utilisons que 2 symboles différents, et les TPT ne contiennent qu'un seul archétype chacune), cette application des types montre le réel intérêt d'une telle représentation mixte de l'information cérébrale dans les réseaux cérébraux à grande échelle.

3. CONCLUSION

De façon générale, les deux applications décrites dans ce chapitre montrent certaines possibilités de RAGE quant à l'expression d'hypothèses concernant l'organisation fonctionnelle du cerveau. On a pu notamment remarquer l'intérêt d'une approche causale et explicite, à la fois pour la définition des modèles et pour l'interprétation de leurs résultats. Par exemple, dans l'étude originale réalisée avec BioCaEn, qui constitue la base de la première application, cette approche permet d'implémenter plusieurs hypothèses différentes expliquant un phénomène observé, puis de tester quelle hypothèse est la plus plausible. Cette première application montre également que la capacité d'expression de RAGE est au moins égale à celle de BioCaEn, et que RAGE offre en outre plusieurs avantages, en particulier l'utilisation de fonctions non-linéaires et un meilleur contrôle de la dispersion des valeurs numériques. Dans la deuxième application également, notre approche facilite l'implémentation des hypothèses.

La deuxième application présente en outre une application plus poussée de RAGE, puisque le rôle des types y est beaucoup plus important que dans le modèle de cortex visuel. Ceci démontre l'intérêt de notre représentation mixte de l'information, qui nous permet de réaliser des modèles à un haut niveau d'abstraction. Grâce à ce haut niveau d'abstraction, nous pouvons réaliser des modèles relativement simples, et donc plus facilement appréhendables par des utilisateurs, de mécanismes cérébraux qui sont relativement complexes si l'on se place à des niveaux inférieurs.

Enfin, en plus des possibilités expressives de RAGE, ces deux applications montrent surtout l'intérêt de notre approche de modélisation dans le cadre de l'interprétation de données de neuroimagerie. Nos modèles constituent, par définition, des simplifications des structures cérébrales ciblées. Toutefois, ils permettent d'en reproduire le comportement, en termes de l'activité observée du moins, et donnent accès aux détails de fonctionnement inaccessibles sinon, que ce soit au niveau temporel (accès à la dynamique à une échelle inférieure à celle de la technique de neuroimagerie employée) ou au niveau spatial (activité au niveau d'un nœud du réseau). Ainsi, les modèles constituent une forme de validation des hypothèses avancées lors de la modélisation. Ils permettent de raffiner ces hypothèses ou d'en dégager de nouvelles.

Tout le long de ce travail, notre objectif a été de construire un formalisme dédié à la modélisation cérébrale, dans un but d'interprétation de données de neuroimagerie, en termes de traitement de l'information cérébrale. Pour cela, nous nous sommes appuyés sur les travaux antérieurs de notre équipe, qui visaient des buts similaires. La définition de contraintes précises nous a permis de choisir dans un premier temps le formalisme de base le plus adapté à notre cas. Nous avons aménagé ce formalisme, les réseaux bayésiens dynamiques, afin de l'adapter plus parfaitement à nos contraintes. Ceci a nécessité la définition de plusieurs concepts concernant la représentation et le traitement de l'information cérébrale dans les réseaux de populations neuronales. Durant cette phase, nous avons eu pour objectif de définir des propriétés suffisamment générales pour rester ouvert aux évolutions dans le domaine des neurosciences. Par la suite, ces concepts ont été traduits en termes formels et unis au formalisme de départ (RBD), donnant ainsi naissance à RAGE (Réseaux Artificiels à Grande Echelle), notre outil de modélisation. Il est caractérisé, en particulier, par une approche causale, qui permet la formulation explicite d'hypothèses.

1. UN FORMALISME DEDIE A LA MODELISATION CEREBRALE A UN NIVEAU INTEGRE

Lorsque nous avons défini notre formalisme, notre but était de suivre le plus possible les contraintes définies au départ, à partir des propriétés du cerveau et de nos objectifs de modélisation. Le respect de ces contraintes transparaît à travers la description des principales caractéristiques de notre formalisme.

L'information manipulée par notre formalisme est duale. Elle comprend une partie numérique, la magnitude, représentant le niveau d'activation d'une population de neurones, ou d'un faisceau d'axones. C'est-à-dire que cette quantité tient à la fois du nombre de neurones activés et de la force leur activation. Elle se justifie donc d'un point de vue physiologique. De plus, même si le lien entre information cérébrale et mesures d'activation est mal connu, on peut rapprocher la magnitude des mesures d'activation obtenues par les techniques de neuroimagerie, qui restent le seul moyen

d'observation de l'activation cérébrale. L'utilisation d'une variable aléatoire pour représenter la magnitude nous permet d'utiliser les avantages qu'offre l'approche probabiliste, en ce qui concerne la manipulation de l'incertitude. L'autre partie de l'information, le type, est de nature symbolique. D'un point de vue physiologique, elle représente la configuration des neurones activés dans une population, ou d'axones transportant un influx nerveux dans un paquet d'axones. De manière plus abstraite, on considère qu'un type représente une catégorie d'information, ce qui permet de définir de manière simple des stimuli extérieurs au modèle. Le concept de type revêt une importance cruciale en ce qui concerne l'aspect explicite de notre formalisme. Sa nature symbolique rend notamment possible l'expression de mécanismes de catégorisation et d'apprentissage qui seraient difficiles à définir en n'utilisant que la partie numérique de l'information. En fait, le type permet de manipuler une information de haut niveau sémantique, ce qui est nécessaire pour rendre notre formalisme suffisamment explicite. Notre représentation de l'information illustre également notre volonté de définir un formalisme général. En effet, l'information peut être adaptée à n'importe quel domaine de modélisation cérébrale (modèle visuel, auditif, moteur, etc.), en modifiant simplement les domaines de définitions des types.

En ce qui concerne le traitement de l'information cérébrale, la caractéristique fondamentale de notre formalisme est sa nature parallèle et distribuée, grâce à l'utilisation d'un formalisme graphique (i.e. qui prend la forme d'un réseau). L'usage de graphes orientés est cohérent avec le traitement cérébral, autant physiologiquement que fonctionnellement. De plus, le formalisme utilisé est causal. Nous avons montré que le fonctionnement cérébral au niveau des réseaux d'aires répondait à notre définition de la causalité. L'utilisation de relations causales permet d'exprimer clairement les relations qui unissent les nœuds d'un modèle, et renforcent le côté explicite de notre modèle, notamment lors de l'expression d'hypothèses. L'utilisation des réseaux bayésiens dynamiques (RBD) nous permet, en outre, de représenter le temps explicitement, en le discrétisant de façon régulière. Cette représentation du temps est particulièrement adaptée à la comparaison entre nos simulations et les données issues de la neuroimagerie. De plus, nous avons vu que la nature non-linéaire des relations entre aires cérébrales est une hypothèse actuellement très plausible. Les RBD rendent possible l'inclusion de fonctions non-linéaires, sans restriction sur leur forme, dans nos modèles. Cette absence de restriction contribue grandement à la capacité qu'a notre formalisme de s'adapter aux évolutions des connaissances en neurosciences, notamment, ici, en ce qui concerne la propagation et le traitement de l'information.

Afin de remplir la contrainte de plasticité, nous proposons un ensemble de mécanismes d'apprentissage. Ils reposent sur plusieurs paramètres propres à certains nœuds. Le processus consiste à réaliser un apprentissage non-supervisé par modifications successives de ces paramètres d'apprentissage, c'est-à-dire que l'on adapte le traitement effectué par le nœud à ses entrées. Ce processus étant relativement indépendant des mécanismes de propagation de l'information, il est possible de l'adapter aux éventuelles évolutions des connaissances sans pour autant remettre le reste du formalisme en cause. Notre processus d'apprentissage repose en très grande partie sur le traitement des types. En fait, cet aspect, et, plus généralement, tout ce qui concerne la propagation et la combinaison des types, constituent un élément critiquable de notre formalisme. En effet, nous nous basons essentiellement sur des hypothèses pour définir nos mécanismes de traitement des types. Il faut souligner, à notre décharge, que les études en neuropsychologie et cognition menées jusqu'à aujourd'hui nous apportent plus des indices que des certitudes quant au lien entre le substrat cérébral et cet aspect catégoriel de l'information cérébrale, en particulier dans un contexte adaptatif. Par exemple, l'intégration d'informations multimodales au niveau des aires primaires fait actuellement l'objet d'études [Falchier et al. '02]. Au niveau applicatif, cela pose un problème de paramétrage des modèles, que nous allons aborder maintenant.

En ce qui concerne le traitement de la magnitude, il n'est pas possible d'utiliser les algorithmes de paramétrage dont disposent les RBD. En effet, ces algorithmes sont basés sur des estimations statistiques, réalisées à partir de données d'observations qui doivent être présentes en quantité suffisante. Or, comme l'a montré le chapitre dédié aux applications de RAGE, les mesures d'activation dont nous disposons ont, en règle générale, une définition temporelle largement supérieure à celle employée pour modéliser la dynamique neuronale dans le modèle. Ceci signifie que nous sommes loin de disposer d'observations pour chaque instant de simulation, et qu'il est donc impossible d'appliquer un de ces algorithmes. Le paramétrage repose donc en grande partie sur les hypothèses de modélisation, et sur les données de neuroanatomie et de neurophysiologie disponibles. Mais en ce qui concerne le traitement des types, il faut souligner que les informations apportées par les études expérimentales sont encore plus indirectes, ce qui implique que le paramétrage repose énormément sur les hypothèses de modélisation.

2. POSITION DE NOTRE APPROCHE

Notre démarche consiste à tenter d'expliquer les mécanismes cérébraux en termes de traitement de l'information, en s'appuyant sur des modèles biologiquement plausibles. On peut donc considérer que notre champ de recherche est celui des neurosciences computationnelles. Nous nous démarquons des modèles issus des sciences cognitives par notre volonté de fonder notre étude de la fonction sur un substrat biologique. De plus, à la différence des approches existant en neuroimagerie, nous nous intéressons à la fois aux manifestations externes de l'activité cérébrale, mais également aux mécanismes internes qui en sont la cause.

Au sein des neurosciences computationnelles, le formalisme de modélisation dominant est les réseaux de neurones formels (RNF) (chapitres II.3.1 et III.1.1). Nous pouvons d'ores et déjà écarter la comparaison entre notre approche et les modèles purement fonctionnels et purement structurels, puisque le but recherché n'a rien à voir avec le nôtre. Nous allons donc nous situer par rapport aux travaux intermédiaires. Ils ont en général en commun avec notre approche le fait de se placer à une grande échelle de modélisation, celle des réseaux de populations de neurones. Il s'agit également de modèles caractérisés par un traitement parallèle, distribué et non-supervisé, ce qui est aussi le cas de notre approche, à une moindre mesure toutefois en ce qui concerne l'aspect fonction et information réparties. Au chapitre des différences, il faut remarquer que les formalismes développés sur une base de RNF sont souvent dédiés à la modélisation d'une fonction bien spécifique, et visent certaines populations en particulier. De plus, ils reposent sur les propriétés d'émergence des RNF et sur le fait que fonction et information sont extrêmement réparties sur le réseau. Tout cela implique une spécification incomplète des propriétés fonctionnelles des modèles lors de leur définition, et peut entraîner des difficultés lorsque l'on tente de donner une interprétation fonctionnelle du réseau ou de certaines de ses parties. Ceci peut également entraîner une certaine rigidité du modèle, qui est défini pour implémenter une hypothèse donnée, et qu'il est difficile d'adapter à une évolution des connaissances, en raison de la nature très distribuée des RNF. A l'inverse, notre formalisme est résolument tourné vers une approche analytique. L'approche explicite et causale que nous utilisons exige de l'utilisateur qu'il fournisse une grande part de spécification. En contrepartie, l'interprétation des résultats est beaucoup plus aisée, et le modèle peut suivre l'évolution des connaissances. En ce qui concerne l'émergence dans notre formalisme, nous l'avons abordée dans notre travail, au moyen de quelques simulations simples, mais il reste encore beaucoup à faire dans ce domaine, notamment en étudiant le comportement d'un modèle sur des données plus réalistes.

En incluant une part de spécification sous la forme de schémas de connexion, les réseaux neuronaux à grande échelle (RNGE) (c.f. II.3.1.3) minimisent ces différences avec notre approche, et facilitent l'interprétation fonctionnelle des modèles. De plus, à l'instar de notre formalisme, cette approche des RNGE vise l'interprétation de données de neuroimagerie, ce qui n'était pas le cas des RNF décrits auparavant. Néanmoins, des différences subsistent avec notre approche, qui reste plus orientée vers la définition explicite des modèles, notamment en raison de la notion de relations causales. En ce qui concerne l'information manipulée, le formalisme que nous avons créé repose en grande partie sur les propriétés particulières de l'information cérébrale lorsqu'elle est considérée au niveau intégré des aires cérébrales. Ceci renvoie au concept de type, qui représente la composante qualitative de l'information cérébrale. On peut considérer que cette notion est également présente dans les RNGE, a travers les connexions existant entre les différents RNF qui constituent le modèle global : l'organisation des neurones activés dans la couche de sortie du RNF émetteur va influencer le comportement du RNF qui reçoit l'information. Toutefois, dans les RNGE, cette composante qualitative est complètement implicite, et d'autant plus difficile à interpréter, alors que notre approche à base de symboles bénéficie d'une grande transparence sémantique. Il faut néanmoins remarquer la plus grande plausibilité biologique des RNGE à ce niveau. Enfin, il existe une autre différence au niveau de la représentation de l'information : à travers l'utilisation de variables aléatoires, notre approche inclut explicitement une prise en compte de l'aspect indéterministe du traitement cérébral, ce qui n'est pas le cas des RNGE.

Le formalisme causal qualitatif BioCaEn est l'autre approche existant en neurosciences computationnelles, et il a permis d'initier notre travail. Notre but était, tout d'abord, de résoudre quelques problèmes rencontrés lors de l'utilisation de BioCaEn. Le premier, et le plus important des problèmes, est le fort accroissement de l'imprécision des valeurs numériques rencontré lors des simulations. Le second problème concerne la contrainte pesant sur la forme que les relations linéaires doivent prendre. Notre second objectif était d'étendre le pouvoir expressif du formalisme, qui présentait certaines limitations, comme l'absence de mécanismes d'apprentissage. L'apport principal de ce travail a été de substituer les techniques de modélisation bayésienne à la simulation semi-qualitative, dans le but de résoudre le problème d'imprécision excessive. L'utilisation du calcul des probabilités nous permet maintenant d'avoir une mesure de l'incertitude de nos résultats, alors que les méthodes semi-qualitatives gèrent l'imprécision de façon implicite. Ce changement de formalisme de base a été l'opportunité pour nous de préciser notre définition de la causalité. Nous avons également introduit plus de plausibilité biologique dans le formalisme. Cela s'est traduit par d'importantes modifications apportées au concept des types. Tout d'abord, en précisant

l'interprétation à donner aux différents symboles et poids associés, en termes de fonctionnement cérébral. Mais aussi, de façon plus formelle, en définissant de nouveaux opérateurs permettant de les combiner, et en tentant de les justifier par rapport à nos hypothèses sur le fonctionnement cérébral. Nous avons également adapté le concept de table de définition des types, afin de l'inclure dans un ensemble de mécanismes d'apprentissage originaux. Enfin, nous avons également redéfini les mécanismes de propagation de l'information. Pour cela, nous avons tout d'abord introduit une distinction entre l'activation d'une population neuronale et son émission. Ceci nous a permis, dans un second temps, de définir des interactions entre le traitement de la magnitude et le traitement des types, qui se faisaient uniquement en parallèle dans BioCaEn.

3. PERSPECTIVES

On peut distinguer deux types de perspectives d'évolution : celles qui concernent le formalisme lui-même, et celles qui se réfèrent à son implémentation. Pour cette dernière, il nous semble nécessaire d'unifier l'interface, qui permet de définir les modèles, et le moteur de simulation. Ceci peut être abordé de deux manières différentes (mais pas exclusives).

Tout d'abord, il est envisageable d'adapter un des algorithmes de filtrage non-linéaire mentionnés au chapitre III.2.1, afin de l'inclure dans l'outil de modélisation. Mais ceci implique, au préalable, de réaliser un tour d'horizon complet et une comparaison objective de ces différents moteurs, afin de retenir le plus adapté. Or, à l'heure actuelle, la communauté est particulièrement dynamique dans ce secteur du filtrage non-linéaire, et de nombreux nouveaux algorithmes et optimisation d'anciens algorithmes voient le jour assez régulièrement. Ces moteurs étant en général accessibles sous la forme de boîtes à outils Matlab, il s'avèrerait peut-être plus judicieux d'adapter à l'outil de modélisation une interface qui permettrait de traduire un modèle RAGE en un script Matlab intelligible par un de ces moteurs.

En ce qui concerne le formalisme proprement dit, il est possible d'aborder les perspectives d'un point de vue théorique et d'un point de vue applicatif. D'un point de vue applicatif, il est clair que nous n'avons pas encore fait le tour des possibilités de notre formalisme. Il est nécessaire de l'explorer plus à fond, en l'appliquant à la modélisation d'autres processus de catégorisation, mais aussi en le testant sur des modèles requerrant des capacités d'apprentissage. En effet, nous avons étudié le comportement adaptif d'un nœud vis-à-vis de données simples, mais il serait intéressant

d'explorer ce comportement dans contexte plus réaliste, c'est-à-dire en incluant plus de nœuds et des données plus complexes.

Lors de l'utilisation du formalisme comme outil de modélisation, un problème a été soulevé : il s'agit du paramétrage des modèles. La méthode actuellement utilisée consiste à combiner des données provenant de mesures de neurophysiologie et neuroanatomie à des valeurs basées sur des hypothèses de travail (notamment pour ce qui concerne le traitement de types), puis à réaliser un affinage manuel de ces valeurs par simulations successives. Pour que RAGE puisse être utilisé par un public de non-informaticiens/non-mathématiciens, il est nécessaire d'automatiser au moins en partie cette phase de développement d'un modèle. Une méthode pourrait être de faire spécifier par l'utilisateur le comportement attendu, concernant le modèle dans sa globalité ou seulement certains nœuds importants. A ce niveau, l'utilisation de la TMS (stimulation transcranienne, c.f. chapitre I.2.3) peut s'avérer utile pour déterminer quel type d'information est traitée par un nœud. Plus généralement, les données de neuroimagerie peuvent apporter la description globale d'un comportement qui peut être ensuite précisé par l'utilisateur. A partir de ces spécifications, une série de valeurs peut être générée, correspondant aux valeurs successives que le nœud doit prendre pour respecter le comportement défini par l'utilisateur. Il est également possible de modifier le modèle initial de manière à inclure des nœuds d'observation supplémentaires, associés aux nœuds dont le comportement a été prédéfini par l'utilisateur. En utilisant ce modèle modifié et les séries de données théoriques précédemment générées, on peut appliquer une méthode classique d'estimation des paramètres. La définition des distributions de probabilités initiales utilisées lors des simulations constitue un autre aspect important du paramétrage des modèles. Le filtre de Kalman non-linéaire que nous avons utilisé dans les simulations du chapitre VI imposant l'utilisation de v.a. normales, nous n'avons pas véritablement abordé cette question. Il s'agit néanmoins d'un sujet d'étude majeur en modélisation bayésienne [Bernardo & Smith '94; O'Hagan '94], qui devra être traité dans le cadre des RAGE.

La comparaison entre les valeurs obtenues par simulation et les données de neuroimagerie constitue une autre perspective de développement. Nos modèles représentent l'activité neuronale, mais les techniques tomographiques ne permettent que d'avoir des mesures indirectes de cette activité. Il n'est donc pas possible de les comparer directement. Dans les deux simulations que nous avons présentées, nous avons d'abord dû effectuer un traitement des valeurs simulées, avant de pouvoir effectuer cette comparaison. Or les méthodes que nous avons employées sont relativement approximatives, et, bien que suffisantes pour illustrer nos propos dans le contexte de ce travail, elles

ne constituent en aucun cas un outil définitif. Il apparaît nécessaire de développer des modèles d'interfaces, c'est-à-dire des modèles de nœuds purement fonctionnels, chargés de transformer l'activation neuronale calculée pour un nœud du modèle en une valeur comparable aux données issues d'une technique de neuroimagerie donnée. On se rapproche en cela des méthodes employées dans les réseaux de neurones à grande échelle (TEP et IRMf synthétique), qui emploient des sortes de filtres pour transformer l'activité d'un RNF en une valeur comparable à une mesure de neuroimagerie. L'intérêt du modèle d'interface est son indépendance au modèle : d'une part, il permettrait d'utiliser le même modèle avec différentes techniques d'imagerie, et d'autre part, une fois développé, il pourrait être inséré dans n'importe lequel de nos modèles. Toutefois, le développement de ce type d'interface sort du champ de recherche dans lequel nous nous plaçons, et nécessite une expertise relevant de la méthodologie en neuroimagerie.

En ce qui concerne l'aspect purement théorique du formalisme, les perspectives concernent essentiellement la manipulation des types. Par rapport à BioCaEn, nous avons fait évoluer la définition des types, et de l'information cérébrale en général, en la précisant et en renforçant sa plausibilité biologique. Cette définition, en raison de la sémantique que nous donnons aux poids associés aux symboles d'un type, semble compatible avec une approche possibiliste de la représentation et de la manipulation des types. Avec le développement des réseaux possibilistes dynamiques manipulant des symboles, l'expression du calcul des types en termes possibilistes permettrait d'utiliser une sorte de filtre de Kalman possibiliste pour le traitement de la partie du modèle concernant les types (c.f. Figure B.1.1). L'avantage serait de quantifier l'imprécision des types, et éventuellement de définir de nouveaux opérateurs basés sur la logique possibiliste. Par rapport à BioCaEn, nous avons également fait évoluer les mécanismes de propagation, et introduit des processus d'apprentissage, en nous appuyant, entre autres, sur notre nouvelle définition du type. A cette occasion, nous avons fait l'hypothèse, que pendant une simulation, la structure d'un modèle ne changeait pas. Cela se traduit notamment par le fait que les domaines de définition des types que nous manipulons sont fixes. Or, pour étendre l'expressivité de RAGE, il est nécessaire de définir des mécanismes permettant la modélisation dynamique des phénomènes de réorganisation du réseau à grande échelle. Ceci peut impliquer des changements dans les connexions entre populations neuronales, et donc la modification de la structure d'un modèle. La dernière perspective consiste donc à décrire un mécanisme chargé de modifier les domaines de définition des types d'émission, et par conséquent ceux des types d'activation et des archétypes des TPT.

Toutes les perspectives que nous avons décrites visent à amméliorer RAGE en tant qu'outil de modélisation cérébrale. On peut également s'intéresser à l'application de notre formalisme à la modélisation d'autres systèmes. En effet, il semble possible d'étendre la validité de nos principes à des systèmes complexes caractérisés par le traitement de flux d'information labellisée, dans un réseau orienté de nœuds fonctionnellement différenciés. Ceci pourrait par exemple s'appliquer à la modélisation de processus industriels, comme des chaînes de montage. Les différents composants sont transportés par une chaîne vers une machine chargée de les assembler, produisant un nouveau composant. On peut assimiler ceci à la propagation de différents types (les composants), qui sont combinés par un nœud (la machine), qui propage le résultat de cette combinaison sous la forme d'un nouveau type (le nouveau composant). On pourrait également appliquer notre formalisme à la modélisation d'écosystèmes, pour tracer des populations, ou bien des produits tels que des polluants. C'est en diversifiant les champs d'application qu'un formalisme s'enrichit de nouvelles fonctionnalités, et il est même envisageable que des concepts provenant de la modélisation de systèmes complètement différents du cerveau puissent s'appliquer à la modélisation cérébrale.

BIBLIOGRAPHIE

[Alary *et al.* '98] : F. Alary, B. Doyon, I. Loubinoux, C. Carel, K. Boulanouar, J.P. Ranjeva, P. Celsis & F. Chollet, Event-related potentials elicited by passive movements in humans: Characterization, source analysis, and comparison to fMRI, Neuroimage, 1998;8(4):377-90.

[Albert & Schnellenbach-Held '97] : A. Albert & M. Schnellenbach-Held, The fussy-sets-theory and its applications in structural engineering, Technical report, Technische Universitat, Darmstadt, 1997.

[Alexander *et al.* '92] : G.E. Alexander, M.R. Delong & M.D. Crutcher, Do cortical and basal ganglionic motor area use "motor programs" to control movement?, BBS, 1992;15:656-65.

[Andersen *et al.* '97] : R.A. Andersen, L.H. Snyder, D.C. Bradley & J. Xing, Multimodal representation of space in the posterior parietal cortex and its use in planning movements, Annu Rev Neurosci, 1997;20:303-30.

[Anderson & Moore '92] : B.D.O. Anderson & J.B. Moore, Kalman filtering: Whence, what and whither?, In: A. Antoulas, editor., Mathematical system theory: The influence of r.E. Kalman, Springer Verlag1992.

[Anderson & Rosenfeld '88] : J.A. Anderson & E. Rosenfeld, Neuro computing foundations of research, Cambridge: MIT PRESS, 1988.

[Anderson '88] : J.R. Anderson, A spreading activation theory of memory, In: A. Collins & E.E. Smith, editors., A perspective from psychology and artificial intelligence, Morgan Kaufmann: San Mateo (Ca), 1988.

[Anderson '89] : J.R. Anderson, A theory of the origins of human knowledge, Artificial Intelligence, 1989;40(3-4):313-51.

[Anderson *et al.* '02a] : J.R. Anderson, D. Bothell, M.D. Byrne & C. Lebiere, An integrated theory of the mind, Psychological Review (submitted), 2002a.

[Anderson *et al.* '02b] : J.R. Anderson, Y. Qin, M.-H. Sohn, V.A. Stenger & C.S. Carter, An information-processing model of the bold response in symbol manipulation tasks, Psychonomic Bulletin and Review, 2002b.

[Arbib '85a] : M.A. Arbib, Brain theory and cooperative computation, Human Neurobiol., 1985a;4:201-18.

[Arbib '85b] : M.A. Arbib, Brain theory and cooperative computation, Hum Neurobiol, 1985b;4(4):201-18.

[Arbib '95] : M.A. Arbib, The handbook of brain theory and neural networks, MIT Press: Cambridge, Mass., 1995.

[Arbib *et al.* '95] : M.A. Arbib, A. Bischoff, A.H. Fagg & S.T. Grafton, Synthetic PET: Analyzing large-scale properties of neural networks, Human Brain Mapping, 1995;2:225-33.

[Arbib *et al.* '98] : M.A. Arbib, P. Erdi & J. Szentágothai, Neural organization: Structure, function, and dynamics, MIT Press: Cambridge, Mass., 1998.

[Arbib *et al.* '00] : M.A. Arbib, A. Billard, M. Iacoboni & E. Oztop, Synthetic brain imaging: Grasping, mirror neurons and imitation, Neural Netw, 2000;13(8-9):975-97.

[Ardila '93] : A. Ardila, Toward a model of phoneme perception, Int J Neurosci, 1993;70(1-2):1-12.

[Arulampalam *et al.* '02] : M.S. Arulampalam, S. Maskell, N. Gordon & T. Clapp, A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking, Ieee Transactions on Signal Processing, 2002;50(2):174-88.

[Bélisle '02] : C. Bélisle, 08 - l'estimation des paramètres d'un modèle statistique, Probabilités pour ingénieurs (Course) STT-20694, Département de mathématiques et de statistique, Université Laval, Québec, Canada, http://www.mat.ulaval.ca/pages/belisle/STT-20694-A02/Probabilites-pour-ingenieurs.html, 2002.

[Benferhat *et al.* '99] : S. Benferhat, D. Dubois, L. Garcia & H. Prade, Réseaux possibilistes orientés et logique possibiliste, Technical Report Rapport IRIT/99-32-R, IRIT, Toulouse, 1999.

[Benferhat *et al.* '00] : S. Benferhat, D. Dubois & H. Prade, Kalman-like filtering in a qualitative setting, a preliminary draft, Technical report, 2000.

[Benferhat *et al.* '01] : S. Benferhat, D. Dubois, S. Kaci & H. Prade, Graphical readings of possibilistic logic bases, In: Proceedings of Conference Uncertainty in Artificial Intelligence, Morgan Kaufmann, 2001:24-31.

[Berger '29] : H. Berger, Über das elektroenkephalogramm des menschen, Archiv. für Psychiatrie und Nervenkrankheiten, 1929;87:527-70.

[Berleant & Kuipers '92] : D. Berleant & B. Kuipers, Qualitative-numeric simulation with q3, In: B. Faltings & P. Struss, editors., Recent advances in qualitative physics, MIT Press: Cambridge, 1992.

[Bernardo & Smith '94] : J.M. Bernardo & A.F.M. Smith, Bayesian theory, Wiley: Chichester, 1994.

[Binder *et al.* '97] : J. Binder, D. Koller, S. Russell & K. Kanazawa, Adaptive probabilistic networks with hidden variables, Machine Learning, 1997;29(2-3):213-44.

[Blakemore *et al.* '99] : S.J. Blakemore, C.D. Frith & D.M. Wolpert, Spatio-temporal prediction modulates the perception of self-produced stimuli, J Cogn Neurosci, 1999;11(5):551-9.

[Bonarini & Bontempi '94] : A. Bonarini & G. Bontempi, A qualitative simulation approach for fuzzy dynamic models, ACM Transactions on Modeling and Computer Simulation, 1994;4(4):285-313.

[Borgelt *et al.* '98] : C. Borgelt, J. Gebhardt & R. Krusse, Possibilistic graphical models, In: Proceedings of ISSEK, 1998.

[Borgelt & Gebhardt '99] : C. Borgelt & J. Gebhardt, A naive bayes style possibilistic classifier, In: Proceedings of European Congress on Intelligent Techniques and Soft Computing, Verlag Mainz, 1999.

[Bremmer *et al.* '01] : F. Bremmer, A. Schlack, N.J. Shah, O. Zafiris, M. Kubischik, K. Hoffmann, K. Zilles & G.R. Fink, Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys, Neuron, 2001;29(1):287-96.

[Breslow '96] : N.E. Breslow, Generalized linear models: Checking assumptions and strengthening conclusions, Statistica Applicata, 1996;8:23-41.

[Bressler '95] : S.L. Bressler, Large-scale cortical networks and cognition, Brain Res Rev, 1995;20(3):288-304.

[Broca '61] : P.P. Broca, Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole), Bulletin de la Société Anatomique, 1861;6:330-57.

[Büchel & Friston '97] : C. Büchel & K.J. Friston, Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI, Cereb Cortex, 1997;7(8):768-78.

[Büchel *et al.* '99] : C. Büchel, J.T. Coull & K.J. Friston, The predictive value of changes in effective connectivity for human learning, Science, 1999;283(5407):1538-41.

[Büchel & Friston '00] : C. Büchel & K. Friston, Assessing interactions among neuronal systems using functional neuroimaging, Neural Netw, 2000;13(8-9):871-82.

[Buntime '94] : W.L. Buntime, Operations for learning with graphical models, J. Artif. Intel. Res., 1994;2:159-225.

[Burnod '91] : Y. Burnod, Organizational levels of the cerebral cortex: An integrated model, Acta Biotheor, 1991;39(3-4):351-61.

[Burnod *et al.* '92] : Y. Burnod, P. Grandguillaume, I. Otto, S. Ferraina, P.B. Johnson & R. Caminiti, Visuomotor transformations underlying arm movements toward visual targets: A neural network model of cerebral cortical operations, J Neurosci, 1992;12(4):1435-53.

[Carlsson & Fullér '01] : C. Carlsson & R. Fullér, On possibilistic mean value and variance of fuzzy numbers, Fuzzy Sets and Systems, 2001;122:315-26.

[Carlsson *et al.* '02] : C. Carlsson, R. Fullér & P. Majlender, Some normative properties of possibility distributions, In: Proceedings of Symposium of Hungarian researchers on computational intelligence, 2002:61-71.

[Carlsson *et al.* '03] : C. Carlsson, R. Fullér & P. Majlender, Possibility distributions, a normative view, In: Proceedings of Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence, 2003:1-9.

[Carpenter & Grossberg '87a] : G. Carpenter & S. Grossberg, A massively parallel architecture for a self organizing neural pattern recognition machine, Computer Vision, Graphics, and Image Processing, 1987a;37:54-115.

[Carpenter & Grossberg '87b] : G. Carpenter & S. Grossberg, Art 2: Self-organization of stable category recognition codes for analog input patterns, Applied Optics, 1987b;26(23):4919-30.

[Carpenter & Grossberg '90] : G. Carpenter & S. Grossberg, Art3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures, Neural Networks, 1990;3:129-52.

[Carpenter *et al.* '91a] : G. Carpenter, S. Grossberg & J.H. Reynolds, Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, Neural Networks, 1991a;4:565-88.

[Carpenter *et al.* '91b] : G. Carpenter, S. Grossberg & D.B. Rosen, Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system, Neural Networks, 1991b;4:759-71.

[Carpenter *et al.* '92] : G. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds & D.B. Rosen, Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps, IEEE Transactions on Neural Networks, 1992;3(5):698-713.

[Carpenter & Grossberg '93] : G.A. Carpenter & S. Grossberg, Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions, Trends Neurosci, 1993;16(4):131-7.

[Carpenter & Grossberg '91] : G.A. Carpenter & S. Grossberg, Pattern recognition by self-organizing neural networks, Cambridge: MIT Press, 1991.

[Carter & Kohn '96] : C.K. Carter & R. Kohn, Markov chain monte carlo in conditionally gaussian state space models, Biometrika, 1996;83(3):589-601.

[Chollet '00] : F. Chollet, Plasticity of the adult human brain, In: A.W. Toga & J.C. Mazziotta, editors., Brain mapping: The systems, Academic Press: San Diego, 2000.

[Churchland & Sejnowski '92] : P.S. Churchland & T.J. Sejnowski, Neuroscience overview, The computational brain, MIT Press: Cambridge, Mass., 1992.

[Cohen '72] : D. Cohen, Magnetoencephalography: Evidence of magnetic fields produced by alpha rhythm currents, Science, 1972;161:664-6.

[Cohen *et al.* '90] : J.D. Cohen, K. Dunbar & J.L. McClelland, On the control of automatic processes: A parallel distributed processing account of the Stroop effect, Psychol Rev, 1990;97(3):332-61.

[Cohn '02] : S. Cohn, Householder triangularization, Course Math 347/847, Department of Mathematics and Statistics, University of Nebraska, Lincoln, http://www.math.unl.edu/~scohn/847s02/nla7.pdf, 2002.

[Cottier & Destrieux '01] : J.-P. Cottier & C. Destrieux, Imagerie fonctionnelle, métabolique et vasculaire du cerveau par résonance magnétique nucléaire, Course, Faculté de Médecine, Rennes, http://www.med.univ-rennes1.fr/etud/pharmaco/index4.htm, 2001.

[Crick '89] : F. Crick, The recent excitement about neural networks, Nature, 1989;337:129-32.

[Damasio '89] : A.R. Damasio, Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition, Cognition, 1989;33(1-2):25-62.

[de Cooman '97] : G. De Cooman, Possibility theory III: Possibilistic independence, International Journal of General Systems, 1997;25:353-71.

[de Kleer & Brown '86] : J. De Kleer & J.S. Brown, Theories of causal ordering, Artificial Intelligence, 1986;29:33-61.

[De Waegenaere & Wakker '97] : A. De Waegenaere & P. Wakker, Choquet integrals with respect to non-monotonic set functions, Discussion paper 44, Tilburg University, Center for Economic Research, http://netec.mcc.ac.uk/WoPEc/data/Papers/dgrkubcen199744.html, 1997.

[Defrise & Trebossen '02] : M. Defrise & R. Trebossen, La tomographie d'émission de positons, In: P. Grangeat, editor., La tomographie médicale : Imagerie morphologique et imagerie fonctionnelle, Hermès: Paris, 2002.

[Démonet *et al.* '94] : J.F. Démonet, C. Price, R. Wise & R.S. Frackowiak, A PET study of cognitive strategies in normal subjects during language tasks. Influence of phonetic ambiguity and sequence processing on phoneme monitoring, Brain, 1994;117 (Pt 4):671-82.

[Dominey & Arbib '92] : P.F. Dominey & M.A. Arbib, A cortico-subcortical model for generation of spatially accurate sequential saccades, Cereb Cortex, 1992;2(2):153-75.

[Drakopoulos '95] : J. Drakopoulos, Probabilities, possibilities, and fuzzy sets, International Journal of fuzzy sets and systems, 1995;75(1):1-15.

[Dreyfus '98] : G. Dreyfus, Les réseaux de neurones, Mécanique industrielle et matériaux, 1998;51.

[Dreyfus & Idan '98] : G. Dreyfus & Y. Idan, The canonical form of nonlinear discrete-time models, Neural Computation, 1998;10(1):133-64.

[Druzdzel & Simon '93] : M.J. Druzdzel & H.A. Simon, Causality in bayesian belief networks, In: Proceedings of UAI, 1993:3-11.

[Dubois & Prade '85] : D. Dubois & H. Prade, Fuzzy numbers: An overview, In: J.C. Bezdek, editor., Analysis of fuzzy information, CRC Press Inc.: Boca Raton, 1985.

[Dubois & Prade '93] : D. Dubois & H. Prade, Fuzzy sets and probability : Misunderstandings, bridges and gaps, In: Proceedings of IEEE Inter. Conf. on Fuzzy Systems, 1993;2:1059-68.

[Dubois *et al.* '94] : D. Dubois, F. Dupin De Saint-Cyr & H. Prade, Updating, transition constraints and possibilistic Markov chains, In: Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based Systems, 1994:826-31.

[Dubois & Prade '94] : D. Dubois & H. Prade, Ensembles flous et théorie des possibilités : Notions de base, Logique floue, Masson: Paris, 1994.

[Dubois & Prade '97] : D. Dubois & H. Prade, A fuzzy set approach to case-based decision, In: Proceedings of EFDAN, 1997.

[Duvernoy *et al.* '92] : H.M. Duvernoy, E.-A. Cabanis, M.-T. Iba-Zizen, J. Tamraz & J. Guyot, Principales fonctions corticales des lobes frontal, temporal, pariétal et occipital, In: H.M. Duvernoy, editor., Le cerveau humain : Surface, coupes sériées tridimensionnelles et irm, Springer-Verlag: Paris, 1992.

[Fahrmeir '98] : L. Fahrmeir, State space models: A brief history and some recent developments, Technical Report, University of Munich, Department of Statistics, 1998.

[Fahrmeir '99] : L. Fahrmeir, State space models: A brief history and some recent developments, In: Proceedings of International Statistical Institute, 1999.

[Fahrmeir & Knorr-Held '00] : L. Fahrmeir & L. Knorr-Held, Dynamic and semiparametric models, In: M. Schimek, editor., Smoothing and regression: Approaches, computation and application, Wiley: New York, 2000.

[Falchier *et al.* '02] : A. Falchier, S. Clavagnier, P. Barone & H. Kennedy, Anatomical evidence of multimodal integration in primate striate cortex, J Neurosci, 2002;22(13):5749-59.

[Felleman & Van Essen '91] : D.J. Felleman & D.C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex, Cereb Cortex, 1991;1(1):1-47.

[Fischbach '92] : G. Fischbach, Le cerveau et la pensée, Pour la science, 1992;181:28-38.

[Forbes *et al.* '95] : J. Forbes, T. Huang, K. Kanazawa & S. Russell, The batmobile: Towards a bayesian automated taxi, In: Proceedings of IJCAI, 1995:1878-85.

[Forbus '84] : K.D. Forbus, Qualitative process theory, Artificial Intelligence, 1984;24:85-168.

[Fourastié & Sahler '78] : J. Fourastié & B. Sahler, Probabilités et statistiques, Dunod, 1978.

[Fox & Raichle '84] : P.T. Fox & M.E. Raichle, Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography, J Neurophysiol, 1984;51(5):1109-20.

[Fox & Raichle '85] : P.T. Fox & M.E. Raichle, Stimulus rate determines regional brain blood flow in striate cortex, Ann Neurol, 1985;17(3):303-5.

[Franceries *et al.* '03] : X. Franceries, B. Doyon, N. Chauveau, B. Rigaud, P. Celsis & J.-P. Morucci, Solution of poisson's equation in

a volume conductor using resistor mesh models : Application to event related potential imaging, Journal of Applied physics, 2003;93(6):3578-88.

[Franceschini *et al.* '00] : M.A. Franceschini, V. Toronov, M.E. Filiaci, E. Gratton & S. Fantini, On-line optical imaging of the human brain with 160-ms temporal resolution, Optic express, 2000;6(3):49-57.

[Friedman *et al.* '98] : N. Friedman, M. Godszmidt & T.J. Lee, Bayesian network classification with continuous attributes: Getting the best of both discretization and parametric fitting, In: Proceedings of Fifteenth International Conference on Machine Learning (ICML), 1998.

[Friston '94] : K. Friston, Functional and effective connectivity: A synthesis, Hum. Brain Mapp., 1994;2:56-78.

[Friston *et al.* '95] : K.J. Friston, A.P. Holmes, K.J. Worsley, J.-P. Poline & R.S.J. Frackowiak, Statistical parametric maps in functional imaging: A general linear model approach, Hum. Brain Map., 1995;2:189-210.

[Friston *et al.* '98] : K.J. Friston, O. Josephs, G. Rees & R. Turner, Nonlinear event-related responses in fMRI, Mg. Res. Med., 1998;39:41-52.

[Fullér & Majlender '02] : R. Fullér & P. Majlender, On possibilistic dependencies, Technical report 477, Turku Centre for Computer Science, Abo, http://www.tucs.fi/Publications/insight.php?id=tFuMa02b&table=techreport, 2002.

[Gernero '01a] : L. Gernero, Les bases physiques et physiologiques de la magnétoencéphalographie et de l'electroencéphalographie, Technical report, CNRS-UPR640-LENA, http://www.ccr.jussieu.fr/meg-center/media/ecp2001/Meg11.pdf, 2001a.

[Gernero '01b] : L. Gernero, Localisation de sources en meg-EEG, Technical report, CNRS-UPR640-LENA, http://www.ccr.jussieu.fr/meg-center/media/ecp2001/Meg21.pdf, 2001b.

[Ghahramani '97] : Z. Ghahramani, Computational models of sensorimotor integration, In: P.G. Morasso & V. Sanguineti, editors., Self-organisation, computational maps and motor control, North-Holland: Amsterdam, 1997.

[Ghahramani & Jordan '97] : Z. Ghahramani & M.I. Jordan, Factorial hidden Markov models, Machine Learning, 1997;29(2-3):245-73.

[Ghahramani & Wolpert '97] : Z. Ghahramani & D.M. Wolpert, Modular decomposition in visuomotor learning, Nature, 1997;386(6623):392-5.

[Ghahramani '98] : Z. Ghahramani, Learning dynamic Bayesian networks, Adaptive Processing of Sequences and Data Structures, 1998;1387:168-97.

[Ghahramani & Hinton '98] : Z. Ghahramani & G. Hinton, Switching state-space models, Technical report CRG-TR-96-3, Department of Computer Science, University of Toronto, Toronto, http://citeseer.nj.nec.com/ghahramani96switching.html, 1998.

[Ghahramani '00a] : Z. Ghahramani, Bayesian learning of model structure, Course (transp), Gatsby Computational Neuroscience Unit, University College London, London, 2000a.

[Ghahramani '00b] : Z. Ghahramani, Computational neuroscience. Building blocks of movement, Nature, 2000b;407(6805):682-3.

[Ghahramani '01] : Z. Ghahramani, An introduction to hidden Markov models and Bayesian networks, International Journal of Pattern Recognition and Artificial Intelligence, 2001;15(1):9-42.

[Goldman-Rakic '88] : P.S. Goldman-Rakic, Topography of cognition: Parallel distributed networks in primate association cortex, Annu Rev Neurosci, 1988;11:137-56.

[Gössl *et al.* '00] : C. Gössl, D.P. Auer & L. Fahrmeir, Dynamic models in fMRI, Magn Reson Med, 2000;43(1):72-81.

[Gössl *et al.* '01] : C. Gössl, L. Fahrmeir & D.P. Auer, Bayesian modeling of the hemodynamic response function in bold fMRI, Neuroimage, 2001;14(1 Pt 1):140-8.

[Granger '69] : C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica, 1969;37:424-38.

[Grewal & Andrews '93] : M.S. Grewal & A.P. Andrews, Kalman filtering, theory and practice, Prentice Hall, 1993.

[Grossberg '76a] : S. Grossberg, Adaptive pattern classification and universal recoding I, Biol Cybern, 1976a;23:121-34.

[Grossberg '76b] : S. Grossberg, Adaptive pattern classification and universal recoding II, Biol Cybern, 1976b;23:187-202.

[Grossberg *et al.* '02] : S. Grossberg, S. Hwang & E. Mingolla, Thalamocortical dynamics of the McCollough effect: Boundary-surface alignment through perceptual learning, Vision Res, 2002;42(10):1259-86.

[Guigon *et al.* '94] : E. Guigon, P. Grandguillaume, I. Otto, L. Boutkhil & Y. Burnod, Neural network models of cortical functions based on the computational properties of the cerebral cortex, J Physiol Paris, 1994;88(5):291-308.

[Harthong '96] : J. Harthong, Probabilités et statistiques, de l'intuition aux applications, 1996.

[Harvey '89] : A.C. Harvey, Forecasting, structural time series models and the Kalman filter, Cambridge University Press: Cambridge (UK), 1989.

[Hayes '85] : P.J. Hayes, Naïve physics I: Ontology for liquids, In: J.R. Hobbs & R.C. Moore, editors., Formal theories of the common sense world, Ablex Publishing Corporation1985.

[Hebb '49] : D.O. Hebb, The organization of behavior, New York: Wiley, 1949.

[Heckerman & Breese '95] : D. Heckerman & J.S. Breese, Causal independance for probability assessment and inference using Bayesian networks, Technical Report MSR-TR-94-08, Microsoft Corporation, Redmond, 1995.

[Hedaen '72] : H. Hedaen, Introduction à la neuropsychologie, 1972.

[Herbster *et al.* '96] : A.N. Herbster, T. Nichols, M.B. Wiseman, M.A. Mintun, S.T. Dekosky & J.T. Becker, Functional connectivity in auditory-verbal short-term memory in alzheimer's disease, Neuroimage, 1996;4(2):67-77.

[Hintz-Madsen *et al.* '99] : M. Hintz-Madsen, L.K. Hansen, J. Larsen & K.T. Drzewiecki, A probabilistic neural network framework for detection of malignant melanoma, CRC Press1999.

[Hopfield '82] : J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, In: Proceedings of Nat'l Academy of Sciences, 1982:2554-8.

[Horwitz *et al.* '99] : B. Horwitz, M.A. Tagamets & A.R. McIntosh, Neural modeling, functional brain imaging, and cognition, Trends Cogn Sci, 1999;3(3):91-8.

[Horwitz *et al.* '00] : B. Horwitz, K.J. Friston & J.G. Taylor, Neural modeling and functional brain imaging: An overview, Neural Netw, 2000;13(8-9):829-46.

[Hume '00] : D. Hume, A treatise of human nature. Being an attempt to introduce the experimental method of reasoning into moral subjects, Oxford University Press: Oxford, 2000.

[Hurd '96] : W.J. Hurd, Optimum and practical noncausal smoothing filters for estimating carrier phase with phase process noise, Technical Report Code 314-30-61-02-02, NASA, 1996.

[Hush & Horne '93] : D.R. Hush & B.G. Horne, Progress in supervised neural networks, IEEE Signal Processing Magazine, 1993;10:1-38.

[Hyvönen '92] : E. Hyvönen, Constraint reasoning based on interval arithmetic. The tolerance propagation approach, Artificial Intelligence, 1992;58:71-112.

[Ilmoniemi '02] : R.J. Ilmoniemi, Transcranial magnetic stimulation - new modality in brain mapping,, Technical report, BioMag Laboratory, Helsinki University Central Hospital, Helsinki, 2002.

[Ingvar & Petersson '00] : M. Ingvar & K.M. Petersson, Functional maps and brain networks, In: J.C. Mazziotta, editor., Brain mapping: The systems, Academic Press: San Diego, 2000.

[Jani & Levine '00] : N.G. Jani & D.S. Levine, A neural network theory of proportional analogy-making, Neural Netw, 2000;13(2):149-83.

[Janssen *et al.* '96] : H.J. Janssen, G. De Cooman & E. Kerre, First results for a mathematical theory of possibilistic processes, In: Proceedings of European Meeting on Cybernetics and Systems Research, 1996:341-6.

[Jensen *et al.* '90] : F.V. Jensen, S.L. Lauritzen & K.G. Olesen, Bayesian updating in causal probabilistic networks by local computations, Computational Statistics Quarterly, 1990;4:269-82.

[Joliot *et al.* '98] : M. Joliot, F. Crivello, J.M. Badier, B. Diallo, N. Tzourio & B. Mazoyer, Anatomical congruence of metabolic and electromagnetic activation signals during a self-paced motor task: A combined PET-meg study, Neuroimage, 1998;7(4 Pt 1):337-51.

[Jordan & Jacobs '94] : M.I. Jordan & R.A. Jacobs, Hierarchical mixtures of experts and the EM algorithm, Neural Computation, 1994;6:181-214.

[Joublin *et al.* '96] : F. Joublin, F. Spengler, S. Wacquant & H.R. Dinse, A columnar model of somatosensory reorganizational plasticity based on hebbian and non-hebbian learning rules, Biol Cybern, 1996;74(3):275-86.

[Julier & Uhlmann '97] : S.J. Julier & J.K. Uhlmann, A new extension of the Kalman filter to nonlinear systems, In: Proceedings of Int. Symp. Aerospace/Defense Sensing, Simul. and Controls, SPIE, 1997.

[Julier & Uhlmann '02] : S.J. Julier & J.K. Uhlmann, Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations, In: Proceedings of IEEE American Control Conference, IEEE, 2002.

[Kalman '60] : R.E. Kalman, A new approach to linear filtering and prediction problems, J Basic Eng., Trans ASME, Series D, 1960;82:35-45.

[Kalman & Bucy '61] : R.E. Kalman & R.S. Bucy, New results in linear filtering and prediction theory, J Basic Eng., Trans ASME, Series D, 1961;83:95-108.

[Kaminski *et al.* '01] : M. Kaminski, M. Ding, W.A. Truccolo & S.L. Bressler, Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance, Biol Cybern, 2001;85(2):145-57.

[Kanazawa *et al.* '95] : K. Kanazawa, D. Koller & S.J. Russell, Stochastic simulation algorithms for dynamic probabilistic networks, In: Proceedings of UAI, 1995.

[Kant '00] : E. Kant, Critique de la faculté de juger, Flammarion: Paris, 2000.

[Kay *et al.* '00] : H. Kay, B. Rinner & B. Kuipers, Semi-quantitative system identification, Artificial Intelligence, 2000;119:103-40.

[Kennan *et al.* '02] : R.P. Kennan, D. Kim, A. Maki, H. Koizumi & R.T. Constable, Non-invasive assessment of language lateralization by transcranial near infrared optical topography and functional mri, Hum Brain Mapp, 2002;16(3):183-9.

[Kherif *et al.* '02] : F. Kherif, J.B. Poline, G. Flandin, H. Benali, O. Simon, S. Dehaene & K.J. Worsley, Multivariate model specification for fMRI data, Neuroimage, 2002;16(4):1068-83.

[Kieras & Meyer '95] : D. Kieras & D.E. Meyer, An overview of the epic architecture for cognition and performance with application to human-computer interaction, Technical report TR-95/ONR-EPIC-5, University of Michigan, Electrical Engineering and Computer Science Department, http://www.eecs.umich.edu/~kieras/epic.html, 1995.

[Kieras & Meyer '96] : D. Kieras & D.E. Meyer, The epic architecture: Principles of operation, On-line publication, http://www-personal.engin.umich.edu/~kieras/epic.html, 1996.

[Kjaerulff '92] : U. Kjaerulff, A computational scheme for reasoning in dynamic probabilistic networks, In: Proceedings of UAI, Morgan Kaufmann, 1992:121-9.

[Kjaerulff '93] : U. Kjaerulff, A computational scheme for dynamic Bayesian networks, Technical Report R93-2018, Institute for Electronic Systems, Aalborg, Denmark, 1993.

[Kjaerulff '95] : U. Kjaerulff, Dhugin: A computational system for dynamic time-sliced Bayesian networks, International Journal of Forecasting, Special Issue on Probability Forecasting, 1995;11:89-111.

[Koizumi *et al.* '99] : H. Koizumi, Y. Yamashita, A. Maki, T. Yamamoto, Y. Ito, H. Itagaki & R. Kennan, Higher-order brain function analysis by trans-cranial dynamic near-infrared spectroscopy imaging, Journal Biomed. Opt., 1999;4(4):403-13.

[Krause *et al.* '00] : J.B. Krause, J.G. Taylor, D. Schmidt, H. Hautzel, F.M. Mottaghy & H.W. Müller-Gärtner, Imaging and neural modelling in episodic and working memory processes, Neural Netw, 2000;13(8-9):847-59.

[Kuipers '01] : B. Kuipers, Qualitative simulation, In: R.A. Meyers, editor., Encyclopedia of

physical science and technology, Academic Press: New York, 2001.

[Kuipers '86] : B.J. Kuipers, Qualitative simulation, Artificial Intelligence, 1986;29:289-338.

[Labatut '00] : V. Labatut, Réseaux causaux : Quelle approche pour le cerveau ?, Mémoire de DEA, Formation Doctorale Representation de la Connaissance et Formalisation du Raisonnement, Université Toulouse III Paul Sabatier, Toulouse, 2000.

[Labatut & Pastor '01] : V. Labatut & J. Pastor, Bayesian modeling of cerebral information processing, In: Proceedings of AIME, Bayesian models in medicine Workshop, 2001:41-6.

[Labatut & Pastor '03a] : V. Labatut & J. Pastor, Modeling the cerebral activity with dynamic probabilistic networks, In: Proceedings of BioMedicine V, WIT Press, Southampton, 2003a:459-68.

[Labatut & Pastor '03b] : V. Labatut & J. Pastor, Dynamic Bayesian networks for integrated neural computation, In: Proceedings of IEEE-EMBS Neural Engineering, Omnipress, Madison, 2003b.

[Labatut *et al.* '03a] : V. Labatut, J. Pastor & S. Ruff, Dynamic Bayesian modeling of the cerebral activity, In: Proceedings of IJCAI, AAAI Press, 2003a.

[Labatut *et al.* '03b] : V. Labatut, J. Pastor, S. Ruff, J.-F. Démonet & P. Celsis, Cerebral modeling and dynamic Bayesian networks, Artificial Intelligence in Medecine, 2003b;(in press).

[Lacotte '96] : B. Lacotte, Modélisation par réseaux causaux qualitatifs de la structure fonctionnelle du cerveau humain, Mémoire de DEA, Formation Doctorale Representation de la Connaissance et Formalisation du Raisonnement, Université Toulouse III Paul Sabatier, Toulouse, 1996.

[Lafon *et al.* '97] : M. Lafon, L. Travé-Massuyès & J. Pastor, Biocaen: A causal model simulator for cerebral dynamics, In: Proceedings of Automation-2001, 1997:61-70.

[Lafon *et al.* '98] : M. Lafon, J. Pastor, L. Travé-Massuyès, B. Doyon, J.-F. Démonet & P. Celsis, Qualitative modeling of cerebral information propagation mechanisms, In: Proceedings of International conference on computational intelligence and neuroscience, 1998;2:21-3.

[Lafon *et al.* '99] : M. Lafon, L. Travé-Massuyès & J. Pastor, Hierarchical causal modeling of cerebral information propagation mechanisms, In: Proceedings of IJCAI, Qualitative and Model Based Reasoning for Complex Systems and their Control Workshop, 1999:26-32.

[Lafon '00] : M. Lafon, Modélisation de la propagation de l'information cérébrale par graphes causaux qualitatifs, Thèse de doctorat, Formation Doctorale Representation de la Connaissance et Formalisation du Raisonnement, Université Toulouse III Paul Sabatier, Toulouse, 2000.

[Laird *et al.* '86] : J.E. Laird, P.S. Rosenbloom & A. Newell, Chunking in soar: The anatomy of a general learning mechanism, Machine Learning, 1986;1:11-46.

[Laplane '94] : D. Laplane, Réflexions épistémologiques sur la question de l'organisation cérébrale, Revue de Neurologie (Paris), 1994;150(8-9):555-63.

[Lauritzen '95] : S.L. Lauritzen, The EM algorithm for graphical association models with missing data, Computational Statistics and Data Analysis, 1995;19:191-201.

[Lauritzen & Jensen '99] : S.L. Lauritzen & F. Jensen, Stable local computation with conditional gaussian distributions, Research Report R-99-2014, Department of Mathematical Sciences, Aalborg University, Aalborg, 1999.

[Lauterbur '73] : P.C. Lauterbur, Image formation by induced local interactions: Examples of employement of nmr, Nature, 1973;242:190-1.

[Le Bihan '00] : D. Le Bihan, What to expect from mri in the investigation of the central nervous system?, C R Acad Sci III, 2000;323(4):341-50.

[Lefebvre *et al.* '01] : T. Lefebvre, H. Bruyninckx & J. De Schutter, Kalman filters for nonlinear systems: A comparison of performance, Internal Report 01R033, Department of Mechanical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium, 2001.

[Leiner & Leiner '97] : H.C. Leiner & A.L. Leiner, How fibers subserve computing capabilities: Similarities between brains

and machines, Int Rev Neurobiol, 1997;41:535-53.

[Levine *et al.* '93] : D.S. Levine, R.W. Parks & P.S. Prueitt, Methodological and theoretical issues in neural network models of frontal cognitive functions, Int J Neurosci, 1993;72(3-4):209-33.

[Levine '00] : D.S. Levine, Coding and categorization, In: D.S. Levine, editor., Introduction to neural and cognitive modeling, Lawrence Erlbaum Associates: Mahwah, 2000.

[Lewis '01] : R.L. Lewis, Cognitive theory, soar, International encylopedia of the social and behavioral sciences, Pergamon (Elsevier Science): Amsterdam, 2001.

[Liang *et al.* '00] : H. Liang, M. Ding, R. Nakamura & S.L. Bressler, Causal influences in primate cerebral cortex during visual pattern discrimination, Neuroreport, 2000;11(13):2875-80.

[Loubinoux *et al.* '99] : I. Loubinoux, K. Boulanouar, J.P. Ranjeva, C. Carel, I. Berry, O. Rascol, P. Celsis & F. Chollet, Cerebral functional magnetic resonance imaging activation modulated by a single dose of the monoamine neurotransmission enhancers fluoxetine and fenozolone during hand sensorimotor tasks, J Cereb Blood Flow Metab, 1999;19(12):1365-75.

[Lu '00] : H. Lu, On stability of nonlinear continuous-time neural networks with delay, Neural Networks, 2000;13:1135-43.

[Lumer *et al.* '97] : E.D. Lumer, G.M. Edelman & G. Tononi, Neural dynamics in a model of the thalamocortical system. 1. Layers, loops and the emergence of fast synchronous rhythms, Cereb Cortex, 1997;7(3):207-27.

[Mallot & Giannakopoulos '96] : H.A. Mallot & F. Giannakopoulos, Population networks: A large-scale framework for modelling cortical neural networks, Biol Cybern, 1996;75(6):441-52.

[Marcos *et al.* '93] : S. Marcos, P. Roussel-Ragot, L. Personnaz, O. Nerrand, G. Dreyfus & C. Vignat, Réseaux de neurones pour le filtrage non-linéaire adaptatif, Traitement du Signal, 1993;8:409-22.

[Mazoyer & Belliveau '96] : B. Mazoyer & J.W. Belliveau, Les nouveaux progrès de l'imagerie, La Recherche, 1996;289:26-33.

[McClelland & Rumelhart '81] : J.L. McClelland & D.E. Rumelhart, An interactive activation model of context effects in letter perception: Part 1. An account of basic findings, Psychol Rev, 1981;88(5):375-407.

[McClelland *et al.* '86] : J.L. McClelland, D.E. Rumelhart & G.E. Hinton, The appeal of parallel distributed processing, In: D.E. Rumelhart, J.L. McClelland & T.P.R. Group, editors., Parallel distributed processing, MIT press1986.

[McClelland & Goddard '96] : J.L. McClelland & N.H. Goddard, Considerations arising from a complementary learning systems perspective on hippocampus and neocortex, Hippocampus, 1996;6(6):654-65.

[McCulloch & Pitts '43] : W.S. Mcculloch & W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, 1943;5:115-33.

[McIntosh '00] : A.R. McIntosh, Towards a network theory of cognition, Neural Netw, 2000;13(8-9):861-70.

[McNamee '02] : L. Mcnamee, Photogrammetric calibration of mobile robot kinematic models, PhD, University of Ottawa, Canada, http://www3.sympatico.ca/lou.ainley/thesis/thesis.htm, 2002.

[Mel '93] : B.W. Mel, Synaptic integration in an excitable dendritic tree, J Neurophysiol, 1993;70(3):1086-101.

[Mesulam '90] : M.M. Mesulam, Large-scale neurocognitive networks and distributed processing for attention, language, and memory, Ann Neurol, 1990;28(5):597-613.

[Meyer & Kieras '97] : D.E. Meyer & D.E. Kieras, Precis to a practical unified theory of cognition and action: Some lessons from epic computational models of human multiple-task performance, Technical Report TR-97/ONR-EPIC-8, University of Michigan, Psychology Department, http://www.eecs.umich.edu/~kieras/epic.html, 1997.

[Miller *et al.* '91] : E.K. Miller, L. Li & R. Desimone, A neural mechanism for working and recognition memory in inferior temporal cortex, Science, 1991;254(5036):1377-9.

[Minka '99] : T.P. Minka, From hidden markov models to linear dynamical systems, Technical Report 531, MIT, 1999.

[Minsky '75] : M. Minsky, A framework for representing knowledge, In: P. Winston, editor., The psychology of computer vision, McGraw-Hill: New-York (NY), 1975.

[Mitchell *et al.* '91] : I.J. Mitchell, J.M. Brotchie, G.D. Brown & A.R. Crossman, Modeling the functional organization of the basal ganglia. A parallel distributed processing approach, Mov Disord, 1991;6(3):189-204.

[Monchi *et al.* '00] : O. Monchi, J.G. Taylor & A. Dagher, A neural model of working memory processes in normal subjects, parkinson's disease and schizophrenia for fMRI design and predictions, Neural Netw, 2000;13(8-9):953-73.

[Moore '66] : R.E. Moore, Interval analysis, Prenctice-Hall, 1966.

[Murphy '98a] : K. Murphy, A brief introduction to graphical models and Bayesian networks, Course (net), http://www.ai.mit.edu/~murphyk/Bayes/bayes.html, 1998a.

[Murphy '98b] : K.P. Murphy, Switching Kalman filters, Technical Report, 1998b.

[Nerrand *et al.* '93] : O. Nerrand, L. Personnaz & G. Dreyfus, Non-linear recursive identification and control by neural networks : A general framework, In: Proceedings of European Control Conference, 1993.

[Nørgaard *et al.* '00] : M. Nørgaard, N.K. Poulsen & O. Ravn, Advances in derivative-free state estimation for nonlinear systems, Technical Repport IMM-REP-1998-15, Technical University of Danemark, Lyngby, http://citeseer.nj.nec.com/399660.html, 2000.

[O'Hagan '94] : A. O'hagan, Bayesian inference, Edward Arnold: London, 1994.

[Olesen '93] : K.G. Olesen, Causal probabilistic networks with both discrete and continuous variables, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993;15(3):275--9.

[Onla-or & Winstein '01] : S. Onla-Or & C.J. Winstein, Function of the 'direct' and 'indirect' pathways of the basal ganglia motor loop: Evidence from reciprocal aiming movements in parkinson's disease, Brain Res Cogn Brain Res, 2001;10(3):329-32.

[Opitz & Shavlik '93] : D.W. Opitz & J.W. Shavlik, Heuristically expanding knowledge-based neural networks, In: Proceedings of International Joint Conference on Artificial Intelligence, Morgan Kaufmann, 1993:1360-6.

[Oussar & Dreyfus '01] : Y. Oussar & G. Dreyfus, How to be a gray box: The art of dynamic semi-physical modeling, Neural Networks, 2001;14.

[Pariente *et al.* '01] : J. Pariente, I. Loubinoux, C. Carel, J.F. Albucher, A. Leger, C. Manelfe, O. Rascol & F. Chollet, Fluoxetine modulates motor performance and cerebral activation of patients recovering from stroke, Ann Neurol, 2001;50(6):718-29.

[Pascual-Leone *et al.* '95] : A. Pascual-Leone, D. Nguyet, L.G. Cohen, J.P. Brasil-Neto, A. Cammarota & M. Hallett, Modulation of muscle responses evoked by transcranial magnetic stimulation during the acquisition of new fine motor skills, J Neurophysiol, 1995;74(3):1037-45.

[Pastor *et al.* '97] : J. Pastor, L. Travé-Massuyès, J.-F. Démonet, B. Doyon & P. Celsis, Biocaen: A causal qualitative network for cerebral information propagation modeling, In: Proceedings of 11th International Workshop on Qualitative Reasoning, Istituto di Analisi Numerica CNR, 1997:305-15.

[Pastor *et al.* '00] : J. Pastor, M. Lafon, L. Travé-Massuyès, J.F. Démonet, B. Doyon & P. Celsis, Information processing in large-scale cerebral networks: The causal connectivity approach, Biol Cybern, 2000;82(1):49-59.

[Pearl '88] : J. Pearl, Probabilistic reasoning in intelligent systems: Networks of plausible inference, Morgan Kaufmann: San Fancisco, 1988.

[Pearl & Verma '91] : J. Pearl & T.S. Verma, A theory of inferred causation, In: Proceedings of Conference on Principles of Knowledge Representation and Reasoning, Morgan Kaufmann, 1991:441-52.

[Pearl '94] : J. Pearl, From Bayesian networks to causal networks, In: A. Gammerman,

editor., Bayesian networks and probabilistic reasoning, Walter, Alfred: London, UK, 1994.

[Pearl '96] : J. Pearl, Structural and probabilistic causality, In: D.R. Shanks, K.J. Holyoak & D.L. Medin, editors., Causal learning, Academic Press: San Diego, CA, 1996.

[Pearl '98] : J. Pearl, Graphs, causality, and structural equation models, Socioligical Methods and Research, 1998;27(2):226-84.

[Pearl '99] : J. Pearl, Bayesian networks, Mit encyclopedia of the cognitive sciences, MIT Press: Cambridge, MA, 1999.

[Pearl '00] : J. Pearl, Causality, Cambridge University Press: Cambridge, 2000.

[Pearlmutter '95] : B.A. Pearlmutter, Gradient calculations for dynamic recurrent neural networks: A survey, IEEE Tansactions on Neural Networks, 1995;6(5):1212-28.

[Peña *et al.* '00] : J.M. Peña, J.A. Lozano & P. Larrañaga, An improved Bayesian structural EM algorithm for learning Bayesian networks for clustering, Pattern Recognition Letters, 2000;21(8):779-86.

[Penhune & Doyon '02] : V.B. Penhune & J. Doyon, Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences, J Neurosci, 2002;22(4):1397-406.

[Pernier *et al.* '92] : J. Pernier, F. Perrin & O. Betrand, Identification des générateurs de l'activité électrique cérébrale, Le courrier du CNRS, 1992;79.

[Petersson *et al.* '97] : K.M. Petersson, C. Elfgren & M. Ingvar, A dynamic role of the medial temporal lobe during retrieval of declarative memory in man, Neuroimage, 1997;6(1):1-11.

[Posner *et al.* '88] : M.I. Posner, S.E. Petersen, P.T. Fox & M.E. Raichle, Localization of cognitive operations in the human brain, Science, 1988;240(4859):1627-31.

[Pradhan '93] : M. Pradhan, Belief networks, Australian Health Informatics Association Newsletter, 1993.

[Pulvermüller '96] : F. Pulvermüller, Hebb's concept of cell assemblies and the psychophysiology of word processing, Psychophysiology, 1996;33(4):317-33.

[Quillian '67] : M.R. Quillian, Word concepts: A theory and simulation of some basic semantic capabilities, Behav Sci, 1967;12(5):410-30.

[Radix '91] : J.-C. Radix, Pratique moderne des probabilités, Paris, France: Technique et documentation-Lavoisier, 1991: 437.

[Rae *et al.* '02] : C. Rae, J.A. Harasty, T.E. Dzendrowskyj, J.B. Talcott, J.M. Simpson, A.M. Blamire, R.M. Dixon, M.A. Lee, C.H. Thompson, P. Styles, A.J. Richardson & J.F. Stein, Cerebellar morphology in developmental dyslexia, Neuropsychologia, 2002;40(8):1285-92.

[Raichle '93] : M.E. Raichle, The scratchpad of the mind, Nature, 1993;363(6430):583-4.

[Raichle *et al.* '94] : M.E. Raichle, J.A. Fiez, T.O. Videen, A.M. Macleod, J.V. Pardo, P.T. Fox & S.E. Petersen, Practice-related changes in human brain functional anatomy during nonmotor learning, Cereb Cortex, 1994;4(1):8-26.

[Rankowitz '62] : A. Rankowitz, Positron scanner for locating brain tumors, IEEE Trans. Nucl. Sci., 1962;9:45-9.

[Reeke & Sporns '93] : G.N. Reeke, Jr. & O. Sporns, Behaviorally based modeling and computational approaches to neuroscience, Annu Rev Neurosci, 1993;16:597-623.

[Roland & Zilles '98] : P.E. Roland & K. Zilles, Structural divisions and functional fields in the human cerebral cortex, Brain Res Brain Res Rev, 1998;26(2-3):87-105.

[Rosenblatt '59] : F. Rosenblatt, Two theorems of statistical separability in the perceptron, In: Proceedings of Symposium on the Machinization of Thought Processes, 1959:421-56.

[Roweis & Ghahramani '99] : S. Roweis & Z. Ghahramani, A unifying review of linear gaussian models, Neural Computation, 1999;11(2):305-45.

[Roweis & Ghahramani '01] : S.T. Roweis & Z. Ghahramani, An EM algorithm for identification of nonlinear dynamical systems, In: S. Haykin, editor., Kalman filtering and neural networks, Wiley, John & Sons2001.

[Ruff '00] : S. Ruff, Perception catégorielle de la parole et dyslexie développementale : Étude en IRMf, mémoire de DEA, DEA National de Neuropsychologie, Université de Toulouse II - Le Mirail, Toulouse, 2000.

[Ruff *et al.* '01] : S. Ruff, K. Boulanouar, D. Cardebat, P. Celsis & J.F. Demonet, Brain correlates of impaired categorical phonetic perception in adult dyslexics, Neuroimage, 2001;13(6):S595-S.

[Ruff *et al.* '03] : S. Ruff, N. Marie, P. Celsis, D. Cardebat & J.-F. Démonet, Neural substrates of impaired categorical perception of phonemes in adult dyslexics: An fMRI study, Brain and Cognition, 2003;In Press.

[Rumelhart *et al.* '86] : D.E. Rumelhart, G.E. Hinton & R.J. Williams, Learning representations by back-propagating errors, Nature, 1986;323:533-6.

[Rumelhart & Mc Clelland '86] : D.E. Rumelhart & J.L. Mc Clelland, Parallel distributed processing: Explorations in the microstructures of cognition, MIT press: Cambridge, 1986.

[Russell *et al.* '94] : S.J. Russell, J. Binder & D. Koller, Adaptive probabilistic networks, Computer Science Division (EECS) University of California: Berkeley, Calif., 1994.

[Sadato *et al.* '96] : N. Sadato, A. Pascual-Leone, J. Grafman, V. Ibanez, M.P. Deiber, G. Dold & M. Hallett, Activation of the primary visual cortex by braille reading in blind subjects, Nature, 1996;380(6574):526-8.

[Sangüesa & Cortés '97] : R. Sangüesa & U. Cortés, Possibilistic conditional dependency, similarity and information measures: An application to causal network recovery, In: Proceedings of Fuzzy Theory and Technologies, 1997.

[Sarle '94] : W. Sarle, Neural networks and statistical models, In: Proceedings of SAS Users Group International Conference, SAS Institute, 1994:1538-50.

[Schank & Abelson '77] : R.C. Schank & R.P. Abelson, Scripts, plans, goals and understanding, Lawrence Erlbaum Associates: Hillsdale (NJ), 1977.

[Schank & Farrell '88] : R.C. Schank & R.G. Farrell, Memory, In: M.F. McTear, editor., Understanding cognitive science1988.

[Schei '97] : T.S. Schei, A finite-difference method for linearization in nonlinear estimation algorithms, Automatica, 1997;33(11):2053-8.

[Schiltz *et al.* '01] : C. Schiltz, J.M. Bodart, C. Michel & M. Crommelinck, A pet study of human skill learning: Changes in brain activity related to learning an orientation discrimination task, Cortex, 2001;37(2):243-65.

[Sergent '94] : J. Sergent, Brain-imaging studies of cognitive functions, Trends Neurosci, 1994;17(6):221-7.

[Shoham '88] : Y. Shoham, Reasoning about change: Time and causation from the standpoint of artificial intelligence, MIT Press, 1988.

[Skiena '90] : S. Skiena, Implementing discrete mathematics: Combinatorics and graph theory with mathematica, Addison-Wesley: Reading, 1990.

[Sporns *et al.* '00a] : O. Sporns, G. Tononi & G.M. Edelman, Connectivity and complexity: The relationship between neuroanatomy and brain dynamics, Neural Netw, 2000a;13(8-9):909-22.

[Sporns *et al.* '00b] : O. Sporns, G. Tononi & G.M. Edelman, Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices, Cereb Cortex, 2000b;10(2):127-41.

[Sporns *et al.* '02] : O. Sporns, G. Tononi & G.M. Edelman, Theoretical neuroanatomy and the connectivity of the cerebral cortex, Behav Brain Res, 2002;135(1-2):69-74.

[Suppes '70] : P. Suppes, A probabilistic theory of causation, North Holland: Amsterdam, 1970.

[Suri *et al.* '01] : R.E. Suri, J. Bargas & M.A. Arbib, Modeling functions of striatal dopamine modulation in learning and planning, Neuroscience, 2001;103(1):65-85.

[Svensén *et al.* '02] : M. Svensén, F. Kruggel & H. Benali, Ica of fMRI group study data, Neuroimage, 2002;16(3 Pt 1):551-63.

[Tagamets & Horwitz '00] : M.A. Tagamets & B. Horwitz, A model of working memory: Bridging the gap between electrophysiology and human brain imaging, Neural Netw, 2000;13(8-9):941-52.

[Taylor *et al.* '00] : J.G. Taylor, B. Krause, N.J. Shah, B. Horwitz & H.W. Mueller-Gaertner, On the relation between brain images and brain neural networks, Hum Brain Mapp, 2000;9(3):165-82.

[Thierry *et al.* '98] : G. Thierry, B. Doyon & J.F. Demonet, ERP mapping in phonological

and lexical semantic monitoring tasks: A study complementing previous PET results, Neuroimage, 1998;8(4):391-408.

[Tiesinga *et al.* '01] : P.H. Tiesinga, J.M. Fellous, J.V. Jose & T.J. Sejnowski, Computational model of carbachol-induced delta, theta, and gamma oscillations in the hippocampus, Hippocampus, 2001;11(3):251-74.

[Touzet '92] : C. Touzet, Les réseaux de neurones artificiels : Introduction au connexionisme, Course (transp), LERI, Nîmes, France, 1992.

[Travé-Massuyès *et al.* '93] : L. Travé-Massuyès, K. Bousson, J.-M. Evrard, F. Guerrin, B. Lucas, A. Missier, M. Tomasena & L. Zimmer, Non-causal versus causal qualitative modelling and simulation, Intelligent Systems Engineering Journal, 1993;2:159-82.

[Travé-Massuyès *et al.* '97] : L. Travé-Massuyès, P. Dague & F. Guerrin, Le raisonnement qualitatif, Hermès, 1997.

[van der Merwe & Wan '01] : R. Van Der Merwe & E.A. Wan, Efficient derivative-free Kalman filters for online learning, In: Proceedings of ESANN, D-Facto, 2001:205-10.

[van der Merwe & Wan '03a] : R. Van Der Merwe & E. Wan, Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models,, In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2003a.

[van der Merwe & Wan '03b] : R. Van Der Merwe & E. Wan, Sigma-point Kalman filters for probabilistic inference in dynamic state-space models, In: Proceedings of Workshop on Advances in Machine Learning, 2003b.

[van Mier '00] : H. Van Mier, Human learning, In: A.W. Toga & J.C. Mazziotta, editors., Brain mapping: The systems, Academic Press: San Diego, 2000.

[Wallace *et al.* '93] : J.G. Wallace, R.B. Silberstein, K. Bluff & A. Pipingas, Semantic transparency, brain monitoring and evaluation of hybrid cognitive architectures, Connection science, 1993;6(1):43-58.

[Walley & de Cooman '99] : P. Walley & G. De Cooman, Coherence of rules for defining conditional possibilities, International Journal of Approximate Reasoning, 1999;21:63-107.

[Wan & Nelson '96] : E.A. Wan & A.T. Nelson, Dual Kalman filtering methods for nonlinear prediction, estimation, and smoothing, Advances in Neural Information Processing Systems, 1996;9.

[Wang & Arbib '92] : D. Wang & M.A. Arbib, Modeling the dishabituation hierarchy: The role of the primordial hippocampus, Biol Cybern, 1992;67(6):535-44.

[Wang *et al.* '01] : S. Wang, S. Yan & Z. Dou, New type of neural fuzzy system and its application in automatic fire detection, In: Proceedings of International Conference on Automatic Fire Detection, National Institute of Standards and Technology, 2001:191-200.

[Wang & Buzsaki '96] : X.J. Wang & G. Buzsaki, Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model, J Neurosci, 1996;16(20):6402-13.

[Weiller *et al.* '93] : C. Weiller, S.C. Ramsay, R.J. Wise, K.J. Friston & R.S. Frackowiak, Individual patterns of functional reorganization in the human cerebral cortex after capsular infarction, Ann Neurol, 1993;33(2):181-9.

[Welch & Bishop '01] : G. Welch & G. Bishop, An introduction to the Kalman filter, Course, SIGGRAPH, Chapel Hill, NC, http://www.cs.unc.edu/~welch/kalman/kalman_filter/kalman.html, 2001.

[Wermter & Sun '00] : S. Wermter & R. Sun, An overview of hybrid neural systems, In: S. Wermter & R. Sun, editors., Hybrid neural systems, SpringerVerlag: Heidelberg, 2000.

[Wolpert *et al.* '95] : D.M. Wolpert, Z. Ghahramani & M.I. Jordan, An internal model for sensorimotor integration, Science, 1995;269(5232):1880-2.

[Wolpert & Ghahramani '00] : D.M. Wolpert & Z. Ghahramani, Computational principles of movement neuroscience, Nat Neurosci, 2000;3 Suppl:1212-7.

[Woodburn *et al.* '00] : R. Woodburn, A. Astaras, R. Dalzell, A.F. Murray & D.K. Mcneill, Computing with uncertainty in probabilistic neural networks on silicon, In: Proceedings of Symposium on neural computation, 2000.

[Zadeh '78] : L. Zadeh, Fuzzy sets as basis for a theory of possibility, Fuzzy Sets and Systems, 1978;1:3-28.

[Zadeh '65] : L.A. Zadeh, Fuzzy sets, Information and Control, 1965;8:338-53.

[Zilles '96] : K. Zilles, Pour un nouveau découpage du cortex, La Recherche, 1996;289:46-8.

A

ACI *Voir* Analyse en composantes indépendantes
ACP *Voir* Analyse en composantes principales
ACT *Voir* Adaptive control of though
Activation, 133, 156
Adaptatif, 76
Adaptive
  control of though, 55
  resonance theory, 80, 102, 112, 158, 185
Aire
  associative, 24
  cérébrale, 22, 31, 62
  de Broca, 28
  de Brodmann, 20, 22, 118
  pré-frontale, 25
  primaire, 23, 205
Algorithme, 103, 112, 292
Analyse
  en composantes indépendantes, 50
  en composantes principales, 50
Apprentissage, 28, 29, 44, 52, 61, 76, 93, 102, 112, 139, 167, 296
  semi-supervisé, 78
  supervisé, 77
Archétype, 140, 153, 162, 167, 169, 291
ART *Voir* Adaptive resonance theory
Attracteur, 113, 186
Axone, 18, 21, 40, 63, 118

B

BioCaEn, 12, 63, 65, 89, 122, 133, 184, 204, 227, 233
Blood oxygenation level dependant, 36, 56
Boîte
  grise, 82
  noire, 81
BOLD *Voir* Blood oxygenation level dependant

C

Calcul
  des intervalles, 63, 86, 281
  qualitatif, 278
Carte de Kohonen, 80
Causalité, 43, 51, 58, 64, 67, 85, 100, 123, 230
Cerveau, 15
  anatomie, 16
  fonction, 21
Chaîne de Markov, 97
Champ catégoriel, 127, 129, 136, 146
Coefficient
  de glissement, 169, 182
  de renforcement, 167, 181
Colonne corticale, 19, 21, 27, 30, 61, 62, 118
Combinaison de types, 148, 156
  linéaire, 149, 169, 294
  non-linéaire, 150, 294
Connectivité
  causale, 63
  effective, 52, 63
  fonctionnelle, 50, 63
Contrainte
  de modélisation, 40, 145, 184
  de sommation, 147, 153
Cortex, 17
Cytoarchitecture, 19, 22

D

Débit sanguin cérébral régional, 205, 215
Décomposition, 288
  fonctionnelle, 27, 119, 121
  structurelle, 119
Diaschisis, 29
Distance *Voir* Fonction de distance
Domaine
  de définition, 127, 290
  multiple, 129, 136, 146
  simple, 128, 146
Dopaminergique, 20
Dynamique, 75, 84

E

EEG *Voir* Electroencéphalographie
Electroencéphalographie, 33, 49, 51
Element de domaine, 148
Emission, 133
Ensemble flou, 269
EPIC Voir Executive process-interactive control
Etude en activation, 22, 39, 42, 204
Executive process-interactive control, 55

F

Facteur d'apprentissage, 77
Feed-forward, 74
Filtre
  à particules, 104
  à sigma-points, 105
  de Kalman, 49, 56, 98, 104, 185
  de Kalman étendu, 105
  de Kalman non-linéaire, 287
Fonction
  d'activation, 134, 156
  d'association, 154
  d'émission, 134, 157
  de distance, 160, 293
  de modulation, 161, 165, 167, 180, 291
  de préférence, 140, 153, 167, 169, 181
  de propagation, 165, 179, 291
  de similitude, 157, 178, 293
  sigmoïdale, 42, 137, 159, 163, 179, 218, 285
Fusion, 142, 171, 186

G

GABAergique, 20, 27, 60, 120, 166
Glissement, 28, 31, 142, 169, 186
Gyrus, 17, 118, 216

H

Habituation, 29, 44, 61
Hémisphère, 17
Hyperactivation, 49, 177
Hypoactivation, 49, 177, 224

I

IA *Voir* Intelligence artificielle
Imagerie par résonance magnétique, 36, 39, 48, 50, 53, 56, 178, 213
Implémentation, 298
Imprécision, 42, 48, 70, 90
Incertitude, 42, 70, 90
Inférence, 93, 103
Influence marginale, 86
Intégration
  spatiale, 135
  temporelle, 135
Intelligence artificielle, 47, 54, 73
Interface graphique, 298
Introduction, 141, 167
IRM *Voir* Imagerie par résonance magnétique

J

Java, 298

L

Liaison
  distale, 21, 118
  locale, 21, 118
Lobe, 17
Localisation
  Spatiale, 34, 48
  Temporelle, 48

M

Magnétoencéphalographie, 33, 49
Magnitude, 126, 146, 287
  d'activation, 134, 152, 168, 169
  d'émission, 134, 152, 165
Matlab, 302
Maximum de vraisemblance, 267
MEE *Voir* Modèle d'espace d'états
MEG *Voir* Magnétoencéphalographie
MES *Voir* Modèle d'équations structurelles
Mesure
  de possibilité, 274
  de probabilité, 257
  de similitude *Voir* Fonction de similitude
Méthode de Monte Carlo, 101
MMC *Voir* Modèle de Makov caché
Modèle
  cognitif, 54
  d'équations structurelles, 52
  d'espace d'états, 56, 98, 103, 287
  de Markov caché, 98
  générique, 121
  qualitatif causal, 63, 84, 184
  symbolique, 55, 60
Moindres carrés, 268

N

Neuroimagerie fonctionnelle, 32, 40, 47, 58, 65, 235
Neurone
  biologique, 17, 30, 37, 59, 76, 132
  formel, 59, 74, 284
Neurosciences, 145
  computationnelles, 58, 232
  intégratives, 15
NF *Voir* Neurone formel
Nœud
  fonctionnel, 119
  structurel, 118
Non-linéarité, 42, 64, 70, 103, 108, 150, 185

O

Oubli, 142, 167

P

Parallel Distributed Processing, 59
PDP *Voir* Parallel distributed processing
Perceptron, 79
Plasticité, 28, 44, 71, 121
Poids, 148, 186
Population de neurones, 19, 31, 32, 40, 118, 121, 126, 132, 138, 178
Préférence *Voir* Fonction de préférence
Primitive fonctionnelle, 27, 121

R

RAGE *Voir* Réseau artificiel à grande échelle
RBD *Voir* Réseau bayésien dynamique
RCGE *Voir* Réseau cérébral à grande échelle
Règle de Hebb, 77
Relation floue, 271
Renforcement, 78, 141, 167
Réorganisation, 28, 61
Représentation canonique, 75
Réseau
  artificiel à grande échelle, 13, 145, 177, 187, 204, 287
  bayésien, 56, 90, 103
  bayésien dynamique, 95, 103, 145, 184, 230, 287
  cérébral à grande échelle, 21, 26, 31, 118, 123, 131
  de neurones à grande échelle, 62, 65, 233
  de neurones formels, 58, 65, 74, 102, 131, 185, 232
  dynamique, 117, 123
  fonctionnel, 117, 120
  possibiliste, 93
  statique, 117, 121, 206
  structurel, 117
Résonance magnétique nucléaire, 36

Rétro-propagation, 78
RMN *Voir* Résonance magnétique nucléaire
RNF *Voir* Réseau de neurones formels
RNGE *Voir* Réseau de neurones à grande échelle

S

Sérotoninergique, 20
Seuil
  de fusion, 182
  d'oubli, 169, 181
Sigmoïde *Voir* Fonction sigmoïdale
Similitude *Voir* Fonction de similitude
Simulation, 187, 210, 222
SMT *Voir* Stimulation magnétique transcranienne
Soar, 55
Sous-apprentissage, 77
Statique, 75
Statistique, 48, 52, 93, 100, 257
Stimulation magnétique transcranienne, 38
Stimulus, 125
Structure
  de données, 289
  sous-corticale, 17, 20, 22, 25, 32, 33
Sur-apprentissage, 77
Symbole, 126, 128, 146, 154, 162, 289
Synapse, 18, 37

T

Table de préférence des types, 138, 145, 152, 167, 180, 185, 187, 208, 219, 287
Technique
  de surface, 32, 51
  tomographique, 32, 51
TEP *Voir* Tomographie par émission de positrons
Théorie des possibilités, 257, 269
Tomographie par émission de positrons, 35, 48, 53, 178, 204
TPT *Voir* Table de préférence des types
Type, 126, 127, 129, 146, 230, 233, 287
  bruit, 161
  d'activation, 134, 152, 168, 169
  d'émission, 134, 152, 161
  multiple, 129, 148
  simple, 128, 148

V

Variable
  aléatoire, 56, 91, 178
  floue, 94, 272
  normale, 92, 104
Vicariation, 28

ANNEXE A

RAPPELS ET NOTIONS THEORIQUES

1. PROBABILITES ET STATISTIQUES

Cette partie constitue un rappel sur les probabilités et les statistiques, englobant quelques notions et définitions essentielles à la compréhension de notre travail. Pour plus de détails, le lecteur se réfèrera à des ouvrages spécialisés [Fourastié & Sahler '78; Harthong '96; Radix '91].

1.1. Notions de probabilités

La théorie des probabilités permet de quantifier des phénomènes aléatoires. Intuitivement, la probabilité d'un évènement reproductible correspond à sa fréquence de réalisation : c'est ce que l'on appelle l'approche *fréquentiste* des probabilités. C'est-à-dire que si on considère une série de tentatives d'un même évènement, la probabilité que l'évènement se réalise correspond au rapport du nombre de réalisations par le nombre de tentatives.

1.1.1. Espace probabilisé

La théorie des probabilités a été axiomatisée par Andrei Kolmogoroff à partir de 3 objets $(\Omega, \mathcal{A}, P)$ appelés respectivement *espace des observables*, *tribu des évènements* et *loi de probabilité*.

Ω est un ensemble d'*éléments élémentaires* ω, également appelés *épreuves* ou *observables*.

$\mathcal{A}$, l'ensemble des évènements A, est une *tribu* (σ*-algèbre*) c'est-à-dire un sous-ensemble de l'ensemble des parties de Ω, stable pour la complémentation dans Ω et l'union finie ou dénombrable, et contenant l'ensemble vide.

P, la *loi de probabilité* sur $(\Omega, \mathcal{A})$, est une application dans $[0,1]$ telle que :

$P(\Omega) = 1$ (l'évènement *certain* est de probabilité 1) (eq. A.1.1)

et, pour tout ensemble dénombrable d'évènements *incompatibles* $\{A_i\}$ (i.e. disjoints deux à deux) :

$$P\left(\bigcup A_i\right) = \sum P(A_i) \quad \text{(eq. A.1.2)}$$

Le triplet $(\Omega, \mathcal{A}, P)$ est appelé *espace probabilisé*. On a les propriétés élémentaires suivantes :

$P(\varnothing) = 0$ (évènement *presque impossible* ou *P-négligeable*) (eq. A.1.3)

$$A \subset A' \Rightarrow P(A) \leq P(A') \quad \text{(eq. A.1.4)}$$

$$P(A) = 1 - P(\bar{A}) \quad \text{(eq. A.1.5)}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{(eq. A.1.6)}$$

où $\bar{A}$ désigne le complémentaire de A par rapport à Ω. On appelle $P(A \cap B)$ probabilité conjointe de A et B, et on la note également $P(A, B)$.

1.1.2. Probabilité conditionnelle et indépendance

Le concept de probabilité conditionnelle permet de s'intéresser à la probabilité d'un évènement A lorsque l'on sait qu'un autre évènement B est réalisé.

Considérons l'espace probabilisé $(\Omega, \mathcal{A}, P)$ et les évènements $A \in \mathcal{A}$ et $B \in \mathcal{A}$, avec $P(B) > 0$. Alors la *probabilité de* A *conditionnellement à* B, ou *probabilité de* A *sachant* B, est définie par :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \textit{(formule d'inversion)} \quad \text{(eq. A.1.7)}$$

D'où l'on tire la propriété, si A et B sont de probabilités non nulles :

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad \text{(eq. A.1.8)}$$

On définit un *système complet d'évènements* comme une suite finie ou dénombrable d'évènements $\{B_i\}$ constituant une *partition* de Ω, c'est-à-dire que les B_i sont disjoints et que leur union est égale à Ω.

Considérons un espace probabilisé $(\Omega, \mathcal{A}, P)$, un système complet d'évènements $\{B_i\}$ tel que pour tout B_i, $P(B_i) > 0$, et un évènement $A \in \mathcal{A}$ tel que $P(A) > 0$.Nous avons alors les propriétés suivantes :

$$P(A) = \sum P(A|B_i) P(B_i) \text{ (\textit{formule des probabilités totales})} \quad \text{(eq. A.1.9)}$$

$$P(B_k|A) = \frac{P(A|B_k) P(B_k)}{\sum P(A|B_i) P(B_i)} \text{ (\textit{loi de Bayes})} \quad \text{(eq. A.1.10)}$$

Dans un espace probabilisé $(\Omega, \mathcal{A}, P)$, l'*indépendance* entre deux évènements $A, B \in \mathcal{A}$ est définie par :

$$P(A \cap B) = P(A) P(B) \quad \text{(eq. A.1.11)}$$

On a la propriété suivante :

Si $P(A) > 0$, A et B sont *indépendants* si et seulement si $P(B|A) = P(B)$ (eq. A.1.12)

Il existe également une notion d'*indépendance conditionnelle*. On dit que A et B sont indépendants conditionnellement à C, avec $P(C) > 0$ si :

$$P(A \cap B|C) = P(A|C) P(B|C) \quad \text{(eq. A.1.13)}$$

On retrouve pour l'indépendance conditionnelle des propriétés équivalentes à celles de l'indépendance absolue.

1.1.3. Variables aléatoires

Considérons un espace probabilisé $(\Omega, \mathcal{A}, P)$ et $\mathcal{B}$, l'ensemble des boréliens de $\mathbb{R}$ (tribu engendrée par les intervalles ouverts de $\mathbb{R}$). Une *variable aléatoire réelle* (v.a.) X est une application de Ω dans $\mathbb{R}$ telle que :

$$\forall B \in \mathcal{B}: X^{-1}(B) \in \mathcal{A} \qquad \text{(eq. A.1.14)}$$

La *loi de probabilité* P_X d'une v.a. X est l'application de $\mathcal{B}$ dans $[0,1]$ suivante :

$$P_X(B) = P\left(X^{-1}(B)\right) \qquad \text{(eq. A.1.15)}$$

On parle de v.a. *discrète* si $\mathcal{B}$ est finie ou dénombrable. Si une v.a. n'est pas discrète, elle est dite *continue* et elle est caractérisée par une f fonction intégrable de $\mathbb{R}$ dans $[0,1]$ appelée *fonction de densité* :

$$\forall B \in \mathcal{B}: P_X(B) \triangleq \int_B f_X(x)dx \qquad \text{(eq. A.1.16)}$$

La probabilité qu'une v.a. continue prenne une valeur est nulle :

$$\forall x_0, P(X = x_0) = 0 \qquad \text{(eq. A.1.17)}$$

Par convention, les v.a. sont désignées par des majuscules, et les valeurs qu'elles peuvent prendre sont représentées par des minuscules.

On appelle *fonction de répartition* (ou *fonction cumulative*) d'une v.a. l'application de $\mathbb{R}$ dans $[0,1]$ telle que :

$$F_X(x) = P\left(\{\omega \in \Omega : X(\omega) \leq x\}\right) = P(X \leq x) \qquad \text{(eq. A.1.18)}$$

Pour une v.a. discrète pouvant prendre pour valeurs $\{x_1, \ldots, x_n\}$, avec les probabilités respectives $\{p_1, \ldots, p_n\}$, cette fonction est définie par :

$$\forall x, x_k \le x < x_{k+1} : F_X(x) = \sum_{i=1}^{k} p_i \qquad \text{(eq. A.1.19)}$$

Pour une v.a. continue, la fonction de répartition est définie grâce à la fonction de densité de la v.a. :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du \qquad \text{(eq. A.1.20)}$$

Dans le cas où on veut étudier la conjonction de deux (ou plus) v.a X et Y définies sur le même espace probabilisé $(\Omega, \mathcal{A}, P)$, on utilisera la *fonction de répartition conjointe* définie par :

$$F_{XY}(x, y) = P(X < x \wedge Y < y) \qquad \text{(eq. A.1.21)}$$

Dans le cas continu, la fonction de répartition conjointe est définie grâce à la *fonction de densité conjointe* :

$$F_{XY}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u, v)dudv \qquad \text{(eq. A.1.22)}$$

Deux v.a. X et Y sont *indépendantes* ssi :

$$F_{XY}(x, y) = F_X(x)F_Y(y) \qquad \text{(eq. A.1.23)}$$

Pour des v.a. continues, cela se traduit par :

$$f_{XY}(x, y) = f_X(x)\, f_Y(y) \qquad \text{(eq. A.1.24)}$$

La *loi de probabilité conditionnelle* $P(X|Y)$ de X au point x sachant que Y est au point y, est donnée par :

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \text{ si } X \text{ et } Y \text{ sont discrètes} \qquad \text{(eq. A.1.25)}$$

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \text{ si } X \text{ et } Y \text{ sont continues} \qquad \text{(eq. A.1.26)}$$

1.1.4. Moments d'une v.a.

Pour une v.a. discrète X, on définit la *moyenne* ou *espérance mathématique*, notée $E[X]$, comme :

$$E[X] = \sum x_i p_i \qquad \text{(eq. A.1.27)}$$

Pour une variable continue, la moyenne, quand elle existe (i.e. quand X est P-intégrable), est définie par :

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx \qquad \text{(eq. A.1.28)}$$

Considérons deux v.a. X et Y et deux constantes a et b. L'espérance mathématique les propriétés suivantes :

$$E[a] = a \qquad \text{(eq. A.1.29)}$$

$$E[aX] = aE[X] \qquad \text{(eq. A.1.30)}$$

$$E[X+Y] = E[X] + E[Y] \qquad \text{(eq. A.1.31)}$$

A partir de la moyenne, le concept de *moment* d'ordre n, noté α_n, est défini par :

$$\alpha_n = E\left[X^n\right] \qquad \text{(eq. A.1.32)}$$

et le *moment centré* d'ordre n, noté μ_n, est défini par :

$$\mu_n = E\left[\left(X - E[X]\right)^n\right] \qquad \text{(eq. A.1.33)}$$

On utilise surtout α_1, qui est la moyenne, et μ_2, appelée la variance, que l'on note σ^2 ou parfois $Var[X]$. La variance est la moyenne des écarts quadratiques entre les valeurs que peut prendre la v.a. et sa moyenne. Elle traduit la dispersion de la distribution autour de sa valeur moyenne. L'*écart-*

type, noté σ, est la racine carrée de la variance. La variance est aussi utilisée comme mesure d'incertitude : plus elle est faible, moins le résultat de l'expérience est incertain.

Pour une conjonction de v.a., on étend la notion de variance pour définir la *covariance*, qui est le moment centré conjoint d'ordre 2 des v.a.:

$$Cov[X,Y] = E\left[\left(X - E[X]\right)\left(Y - E[Y]\right)\right] \quad \text{(eq. A.1.34)}$$

On a les propriétés suivantes :

$$Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y] \quad \text{(eq. A.1.35)}$$

$$Var[aX+b] = a^2 Var[X] \quad \text{(eq. A.1.36)}$$

$$Cov[X,Y] = E[XY] - E[X]E[Y] \quad \text{(eq. A.1.37)}$$

$$Cov[aX+b, cY+d] = acCov[X,Y] \quad \text{(eq. A.1.38)}$$

$$Cov[X+Z,Y] = Cov[X,Y] + Cov[Z,Y] \quad \text{(eq. A.1.39)}$$

Si X et Y sont indépendantes, leur covariance est nulle (la réciproque n'est pas forcément vraie) Dans le cas où un vecteur X de n v.a. $X_1,\ldots,X_n$ est manipulé, la covariance est définie par la matrice :

$$C_X = E\left[\left(X - E[X]\right)\left(X - E[X]\right)^T\right] \quad \text{(eq. A.1.40)}$$

où M^T dénote la transposée de la matrice M. C_X est symétrique semi-définie positive. A partir de la covariance, on définit le coefficient de *corrélation de Pearson* de deux v.a. :

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]Var[Y]}} \quad \text{(eq. A.1.41)}$$

Le coefficient de corrélation est une mesure du degré de dépendance linéaire entre les deux v.a..

1.1.5. Loi normale

Les v.a. suivant une loi normale (Figure A.1.1) sont particulièrement utilisées en modélisation, et nous intéressent donc dans le cadre de ce travail. Ce type de v.a. offre de gros avantages calculatoires. Une v.a. X de distribution *normale*, également appelée distribution de *Laplace-Gauss* ou de *Gauss* (d'où le nom de *modèles gaussiens*), a pour fonction de densité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{(eq. A.1.42)}$$

où μ est la moyenne de X et σ son écart-type.

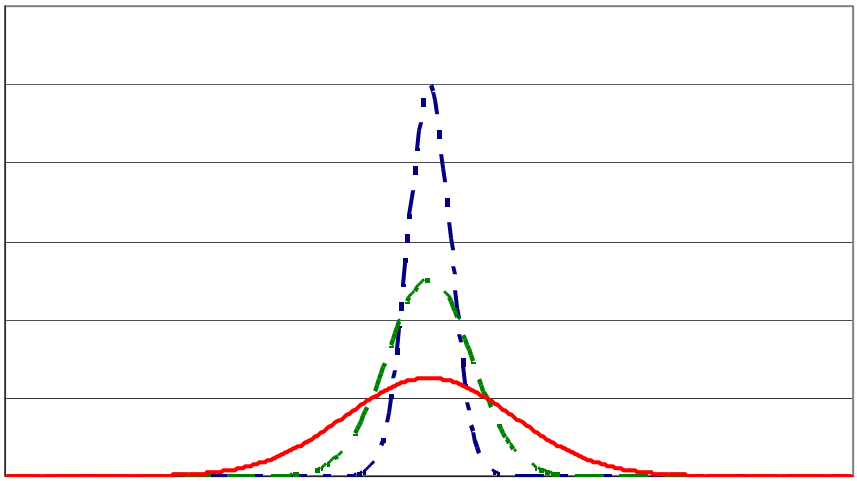

Figure A.1.1 : représentation graphique d'une fonction de densité normale (eq. A.1.42). Les trois courbes correspondent à une même moyenne, mais des variances différentes.

On note :

$$p(X) \sim N(\mu, \sigma^2) \text{ ou } X \sim N(\mu, \sigma^2) \qquad \text{(eq. A.1.43)}$$

Une v.a. normale est donc complètement caractérisée par sa moyenne et sa variance. Graphiquement, cette distribution se présente sous la forme d'une courbe en cloche dont le maximum est atteint en $x = \mu$ (Figure A.1.1).

Les v.a. normales sont très utilisées dans les sciences expérimentales, notamment les sciences de la vie, en raison du théorème central limite. D'après celui-ci, une somme d'un grand nombre de v.a. indépendantes suit une loi normale [Fourastié & Sahler '78]. De plus ce type de variables possède de bonnes propriétés mathématiques : une combinaison linéaire de variables normales est également une variable normale.

1.1.6. Mise à jour bayésienne

Les probabilités conditionnelles et la formule de Bayes permettent d'effectuer une mise à jour récursive de la loi d'une v.a. d'intérêt en fonction d'observations réalisées sur une autre variable. Soient X la v.a. d'intérêt mais dont on ne peut pas connaître la valeur (variable *cachée*) et Y la variable qu'il est possible d'observer (variable *observable*). Si ces deux variables sont dépendantes, d'après la loi de Bayes (eq. A.1.10), on a :

$$P\left(X = x \middle| Y = y\right) = \frac{P\left(Y = y \middle| X = x\right) P\left(X = x\right)}{\sum P\left(Y = y \middle| X = x_i\right) P\left(X = x_i\right)} \text{ (v.a. discrètes)} \qquad \text{(eq. A.1.44)}$$

$$f_{X|Y=y}\left(x\right) = \frac{f_{Y|X=x}\left(y\right) f_X\left(x\right)}{\int f_{Y|X=u}\left(y\right) f_X\left(u\right) du} \text{ (v.a. continues)} \qquad \text{(eq. A.1.45)}$$

Avec ces formules, il est possible de calculer la probabilité de X au vu d'observations effectuées sur Y, à condition de connaître $P\left(Y \middle| X\right)$ et $P\left(X\right)$. On appelle $P\left(X\right)$ *probabilité a priori*, c'est-à-dire avant la prise en compte de l'observation effectuée sur Y, alors que $P\left(X \middle| Y\right)$ est appelée *probabilité a posteriori*, i.e. après la prise en compte de l'observation. La probabilité *a posteriori*, une fois calculée, deviendra la nouvelle probabilité *a priori*, dans l'attente d'une nouvelle observation qui permettra d'effectuer une autre mise à jour.

La probabilité *a priori* est en général connue, soit par des calculs précédents, soit parce qu'elle a été déterminée de façon subjective (par un expert par exemple). C'est à cause de cette subjectivité que cette approche bayésienne des probabilités (ou approche subjective) est opposée à l'approche fréquentiste (ou objective). Dans l'interprétation bayésienne des statistiques, on considère qu'une probabilité dépend toujours d'un contexte, donc est toujours conditionnelle. C'est à dire que l'on va calculer des probabilités absolues (probabilité pour un événement ou une conjonction

d'évènements) à partir des probabilités conditionnelles. Dans la vision *fréquentiste,* on calcule les probabilités conditionnelles à partir des probabilités absolues [Pearl '00; Pradhan '93]. En d'autres termes, l'approche fréquentiste conçoit les probabilités comme l'information résultant de l'observation de grandes séquences d'évènements, alors que la conception bayésienne y voit plus une croyance personnelle en la réalisation d'un événement (d'où cet autre nom de « réseaux de croyances » pour les réseaux bayésiens).

1.2. Estimation en statistiques

1.2.1. Définitions

Soit un ensemble de données observées. On suppose qu'elles sont la réalisation d'un ensemble de v.a. $\{X_1,\ldots,X_n\}$ qu'on appelle *échantillon aléatoire*. Une *statistique* T est une v.a. définie en fonction de l'échantillon aléatoire (ex. : moyenne, variance, etc.) [Bélisle '02] :

$$T = h(X_1,\ldots,X_n) \qquad \text{(eq. A.1.46)}$$

On appelle *échantillon observé* un sous-ensemble des données observées correspondant à une réalisation des v.a. de l'échantillon aléatoire. Pour un échantillon observé $(x_1,\ldots,x_n)$ (x_i étant la valeur observée pour la v.a. X_i), il est donc possible de calculer la valeur observée de la statistique.

Supposons que les v.a. sont indépendantes et qu'elles suivent toutes la même loi $f_\theta(x)$, où θ est le paramètre (ou le vecteur de paramètres) de la fonction f. On note $\hat{\theta}$ un estimateur de θ. Il s'agit d'une statistique dépendant également de l'échantillon aléatoire, et dont la valeur est censée être une bonne approximation de celle de θ. On mesure sa précision en fonction de l'*erreur quadratique moyenne* (EQM), donnée par :

$$EQM(\hat{\theta}) = E\left[\left(\hat{\theta}-\theta\right)^2\right] \text{ (\textit{erreur quadratique moyenne})} \qquad \text{(eq. A.1.47)}$$

On appelle biais de l'estimateur la quantité suivante :

$$Biais\left(\hat{\theta},\theta\right) = E\left[\hat{\theta}\right] - \theta \quad (\textit{biais de l'estimateur}) \qquad \text{(eq. A.1.48)}$$

Un biais positif indique que l'estimateur a tendance à surestimer le paramètre, et inversement pour un biais négatif. Un bon estimateur est sans biais. On a la propriété suivante :

$$EQM\left(\hat{\theta}\right) = Var\left[\hat{\theta}\right] + Biais\left(\hat{\theta},\theta\right)^2 \qquad \text{(eq. A.1.49)}$$

Il existe un grand nombre de méthodes pour obtenir un estimateur, dont nous allons présenter ici les plus connues : *méthode du maximum de vraisemblance*, la *méthode des moments* et celle des moindres carrés.

1.2.2. Méthode du maximum de vraisemblance

On appelle *fonction de vraisemblance* la fonction suivante :

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_{\theta}(x_i) \quad (\textit{fonction de vraisemblance}) \qquad \text{(eq. A.1.50)}$$

Il s'agit en fait de la probabilité conjointe des v.a. $X_1,\ldots,X_n$ évaluées aux points $x_1,\ldots,x_n$. Si les $x_1,\ldots,x_n$ sont fixés, cette fonction exprime la probabilité d'obtenir les valeurs $x_1,\ldots,x_n$ quand le paramètre est θ. Elle quantifie donc la vraisemblance des données en fonction du paramètre. En maximisant cette fonction, on obtient par conséquent le paramètre pour lequel les données sont les plus vraisemblables :

$$\mathcal{L}\left(\hat{\theta}_{MAX}\right) = \max_{\theta} \mathcal{L}(\theta) \qquad \text{(eq. A.1.51)}$$

Parfois, il est plus pratique de travailler avec une fonction de la vraisemblance, plutôt qu'avec la vraisemblance elle-même. Il s'agit très souvent du logarithme de la vraisemblance (ou log-vraisemblance). L'estimateur obtenu par la méthode du maximum de vraisemblance est théoriquement le meilleur estimateur possible. Néanmoins, en pratique, il est parfois difficile de le calculer, d'où l'intérêt de la méthode des moments, qui est produit des estimateurs moins performants, mais plus aisés à calculer.

1.2.3. Méthode des moments

La *méthode des moments* est basée sur le fait que les moments d'une loi peuvent être calculés en fonction de ses paramètres, et inversement. On a donc :

$$\theta = g\left(\mu_1, \ldots, \mu_k\right) \qquad \text{(eq. A.1.52)}$$

Le principe consiste à choisir $\hat{\theta}$ tel qu'il fasse coïncider les moments empiriques (calculés sur les observations) avec les moments théoriques (eq. A.1.32). Le *moment empirique* d'ordre k d'un échantillon est défini par :

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n} X_i^k}{n} \quad (\textit{moment empirique}) \qquad \text{(eq. A.1.53)}$$

On a :

$$\hat{\theta}_{MOM} = g\left(\hat{\mu}_1, \ldots, \hat{\mu}_k\right) \qquad \text{(eq. A.1.54)}$$

Dans le cas où il n'y a qu'un seul paramètre, le moment utilisé est la moyenne. S'il y a plusieurs paramètres, on prend en général les moments d'ordre supérieur suivants.

1.2.4. Régression et méthode des moindres carrés

Jusqu'ici, on a considéré que les v.a. de l'échantillon aléatoire étaient indépendantes. Supposons maintenant que chaque v.a. X_i dépend d'un ensemble de variables déterministes $y_i^{(1)}, \ldots, y_i^{(k)}$, qu'on nomme variables explicatives. On a :

$$X_i = f_\theta\left(y_i^{(1)}, \ldots, y_i^{(k)}\right) + E_i \quad (\textit{modèle de régression}) \qquad \text{(eq. A.1.55)}$$

où la les E_i sont des va indépendantes et de même loi. Le but est ici d'étudier la nature du lien fonctionnel entre les X_i et les Y_i. L'équation (eq. A.1.55) est appelée *modèle de régression*. La fonction f dépend de un ou plusieurs paramètres inconnus θ, que l'on veut estimer. La *méthode des moindres carrés* consiste à utiliser pour cela l'erreur quadratique, que l'on va chercher à minimiser :

$$EQ\left(\hat{\theta}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - f_{\hat{\theta}}\left(y_i^{(1)},\ldots,y_i^{(k)}\right)\right)^2 \quad (\textit{erreur quadratique}) \qquad \text{(eq. A.1.56)}$$

Dans le cas où f est linéaire, on parle de modèle à *régression linéaire*. Il existe différents moyens d'estimation plus élaborés, adaptés à la nature de la fonction f utilisée dans (eq. A.1.55). Il existe également des méthodes pour estimer la fonction de façon *non-paramétrique*, par exemple en faisant apprendre f par un réseau de neurones formels à partir des valeurs observées.

2. ENSEMBLES FLOUS ET THEORIE DES POSSIBILITES

Cette partie est une rapide description de la théorie des ensembles flous et de la théorie des possibilités, qui découle de la première. Pour plus de détails, le lecteur devra se référer à des ouvrages spécialisés [Dubois & Prade '85].

2.1. Ensembles flous

2.1.1. Définitions

Les ensembles flous constituent une généralisation des ensembles classiques, définie par Lofti Zadeh [Zadeh '78] dans un objectif de modélisation du langage naturel. Le langage naturel est caractérisé par la manipulation de termes vagues, faisant référence à des classes d'objets ne possédant pas de critères d'appartenance précis, par exemple : chaud, grand, etc. Un *ensemble flou* F est défini, sur un *domaine* U, par une *fonction d'appartenance* μ_F telle que :

$$\forall u \in U : 0 \le \mu_F\left(u\right) \le 1 \qquad \text{(eq. A.2.1)}$$

On appelle la valeur $\mu_F\left(u\right)$ *degré d'appartenance* de l'élément u à l'ensemble flou F. Les degrés d'appartenance $\mu_F\left(u\right)=1$ et $\mu_F\left(u\right)=0$ signifient respectivement que u appartient et n'appartient pas à F. Dans le cas où les degrés d'appartenance sont toujours 0 ou 1 (valeur binaire) on se ramène à un ensemble classique. Si au contraire ils peuvent prendre des valeurs intermédiaires, l'ensemble est dit flou.

Prenons l'exemple de la phrase « de l'eau chaude ». Le mot « chaude » qualifie ici la température de l'eau, et pourrait être représenté par la fonction d'appartenance définie sur le domaine $[0,100]$ représentée dans la Figure A.2.1.

On définit l'α-coupe d'un ensemble flou, notée F_α, comme étant l'ensemble des éléments qui ont un degré d'appartenance d'au moins α :

$$F_\alpha = \{u \in U : \mu_F(u) \geq \alpha\} \qquad \text{(eq. A.2.2)}$$

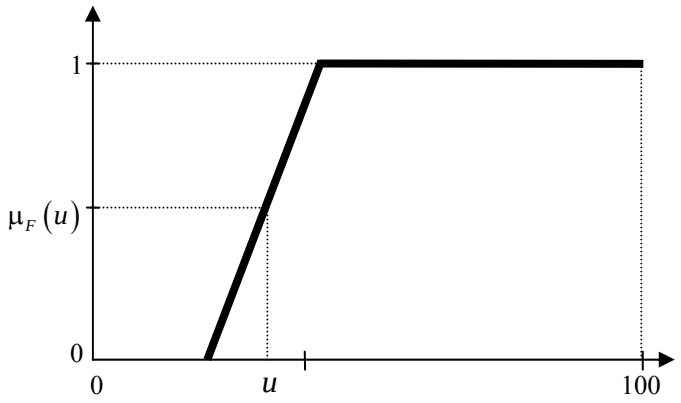

Figure A.2.1 : exemple de fonction d'appartenance d'un ensemble flou.

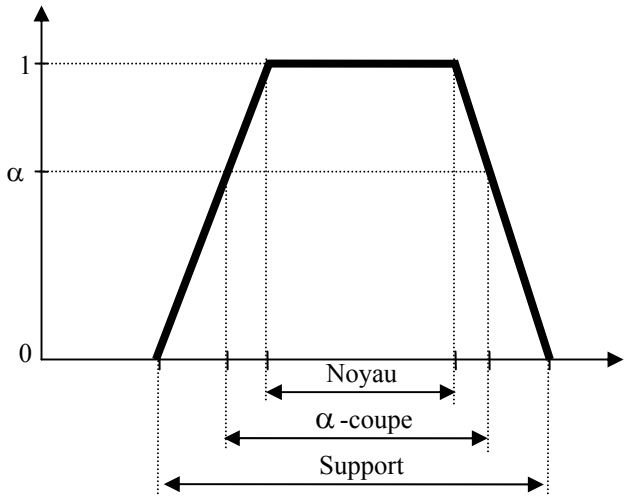

Figure A.2.2 : illustration des notions d'α-coupe, de support et de noyau d'un ensemble flou.

On distingue tout particulièrement le support $\underline{F} = \{u \in U : \mu_F(u) > 0\}$, contenant tous les éléments susceptibles d'appartenir un tant soit peu à F, et le noyau $\overline{F} = \{u \in U : \mu_F(u) = 1\}$, contenant tous les éléments appartenant complètement à F Figure A.2.2.

2.1.2. Opérations sur les ensembles flous

Il existe pour les ensembles flous des opérateurs équivalents à ceux des ensembles classiques : égalité, union, etc. Mais si la fonction d'appartenance d'un ensemble classique, qui est booléenne, ne laisse pas de choix possible pour la définition de ces opérateurs, ce n'est pas le cas pour les ensembles flous et leurs fonctions d'appartenance, qui renvoient des valeurs graduelles.

Les opérateurs sur des ensembles flous peuvent donc être définis de plusieurs façons. Par exemple, l'inclusion peut être vue comme le fait que tous les éléments un tant soit peu présents dans un ensemble le sont au moins autant dans l'ensemble contenant :

$$F \subseteq_1 G \Leftrightarrow \mu_F \leq \mu_G \qquad \text{(eq. A.2.3)}$$

Mais elle peut aussi être vue comme le fait que le noyau de l'ensemble contenu est inclus dans le support de l'ensemble contenant :

$$F \subseteq_1 G \Leftrightarrow \overline{F} \subseteq \underline{G} \qquad \text{(eq. A.2.4)}$$

Les définitions les plus courantes pour les opérateurs union et intersection, et qui seront utilisées par la suite, sont les suivantes [Dubois & Prade '94] :

$$\mu_{F \cup G}(u) = \max\left(\mu_F(u), \mu_G(u)\right) \qquad \text{(eq. A.2.5)}$$

$$\mu_{F \cap G}(u) = \min\left(\mu_F(u), \mu_G(u)\right) \qquad \text{(eq. A.2.6)}$$

2.1.3. Relations floues et variables floues

Une *relation floue* est un ensemble flou dont le domaine est un produit cartésien d'ensembles. Par exemple, considérons deux ensembles U et V. Alors l'ensemble flou R, défini sur $U \times V$ par μ_R

telle que $\forall(u,v) \in U \times V : 0 \leq \mu_R(u,v) \leq 1$, est une relation floue. Si U et V sont eux-mêmes des ensembles flous, on a un produit cartésien flou. C'est la conjonction $\min$ qui est utilisée pour définir sa fonction d'appartenance [Dubois & Prade '94] :

$$\mu_R = \min(\mu_U, \mu_V) \qquad \text{(eq. A.2.7)}$$

et la relation floue est dite *séparable*.

Soit un ensemble flou F disjonctif, c'est-à-dire contenant des valeurs mutuellement exclusives (à la manière d'un évènement en théorie des probabilités). Soit une variable x à valeurs dans U, et telle que la possibilité qu'elle prenne la valeur u est égale à μ_F. Alors x est une *variable floue*, et μ_F, la fonction d'appartenance de F, est appelée sa *distribution de possibilités*, et est notée π_x.

A la manière des distributions de probabilités conjointes, il existe des distributions de possibilités conjointes, qui sont définies comme des relations floues. Pour deux variables x et y de distributions π_x et π_y, la distribution conjointe se note $\pi_{x,y}$. Si cette dernière correspond à une relation floue séparable, on a donc d'après (eq. A.2.7) :

$$\pi_{x,y} = \min(\pi_x, \pi_y) \qquad \text{(eq. A.2.8)}$$

Ces deux variables sont alors dites *non-interactives*. Si on considère que la fonction $\min$ employée ici correspond au produit employé dans l'équation (eq. A.1.11), cette notion correspond formellement à l'indépendance probabiliste. Toutefois, elle en est sémantiquement différente : elle signifie une absence d'hypothèse sur la dépendance entre les variables [Dubois & Prade '94].

Différents opérateurs ont été définis pour les variables floues [Albert & Schnellenbach-Held '97]. Les premiers sont ceux que Lofti Zadeh présente dans son *principe d'extension sup-min* [Zadeh '65] :

$$\pi_{x \bullet y}(a) = \sup_{a=b \bullet c} \min(\pi_x(b), \pi_y(c)) \qquad \text{(eq. A.2.9)}$$

où $\bullet$ représente un des quatre opérateurs de base $\{+, -, *, /\}$, et x et y sont non-interactives.

2.1.4. Moments de variables floues

Il existe pour les variables floues des équivalents des moments des variables aléatoires (moyenne, variance, etc.). Mais, à l'instar des opérateurs sur les ensembles flous, il existe plusieurs définitions de ces *moments flous*. Nous présentons ici les versions les plus répandues, qui sont basées sur une décomposition de la fonction d'appartenance en α-coupes et utilisant des intégrales de Choquet [Dubois & Prade '94], c'est-à-dire qu'on réalise une intégration « horizontale » [De Waegenaere & Wakker '97] (alors qu'on pourrait qualifier l'intégrale au sens de Riemann de « verticale »).

L'α-coupe d'un ensemble flou A est un intervalle que l'on note $A_\alpha = \left[a_1(\alpha), a_2(\alpha) \right]$. L'intégrale de Choquet consiste à intégrer toutes les α-coupes de 0 à 1. Carlsson et Fullér [Carlsson & Fullér '01; Carlsson *et al.* '02; Carlsson *et al.* '03] définissent les moments d'une variable floue au moyen d'une *fonction de pondération* et du concept de *valeur centrale* d'une α-coupe. La fonction de pondération f doit être non négative, monotone, croissante et telle que :

$$\int_0^1 f(\alpha) d\alpha = 1 \qquad \text{(eq. A.2.10)}$$

La valeur centrale de l'α-coupe A_α est définie par :

$$C(A_\alpha) = \frac{1}{\int_{A_\alpha} dx} \int_{A_\alpha} x dx = \frac{a_1(\alpha) + a_2(\alpha)}{2} \qquad \text{(eq. A.2.11)}$$

L'espérance possibilistique d'une variable floue est alors donnée par :

$$E_f[A] = \int_0^1 C(A_\alpha) f(\alpha) d\alpha = \int_0^1 \frac{a_1(\alpha) + a_2(\alpha)}{2} f(\alpha) d\alpha \qquad \text{(eq. A.2.12)}$$

La notion de valeur centrale peut s'étendre à une α-coupe A_α résultant de l'application d'une fonction g à un ensemble de n α-coupes $(A_1)_\alpha, \ldots, (A_n)_\alpha$ [Fullér & Majlender '02] :

$$C\left(g(x_1, \ldots, x_n)\right) = \frac{1}{\int_{A_\alpha} dx} \int_{A_\alpha} g(x) dx \qquad \text{(eq. A.2.13)}$$

où $g(x) = g(x_1, \ldots, x_n)$ et A est la distribution de possibilités conjointes de $A_1, \ldots, A_n$. La variance et la covariance possibilistes sont alors respectivement définies par :

$$Cov_f[A,B] = \int_0^1 \left(C[AB_\alpha] - C[A_\alpha] C[B_\alpha] \right) f(\alpha) d\alpha \qquad \text{(eq. A.2.14)}$$

$$Var_f[A] = Cov_f[A,A] = \int_0^1 \left(\frac{a_2(\alpha) - a_1(\alpha)}{12} \right)^2 f(\alpha) d\alpha \qquad \text{(eq. A.2.15)}$$

Les propriétés des l'espérance possibiliste sont les mêmes que pour son homologue probabiliste, c'est-à-dire, entre autres :

$$E_f[A+B] = E_f[A] + E_f[B] \qquad \text{(eq. A.2.16)}$$

$$E_f[\lambda A] = \lambda E_f[A] \qquad \text{(eq. A.2.17)}$$

où B est un autre ensemble flou et λ est un réel. En ce qui concerne la variance et la covariance, on observe la même chose :

$$Var_f[\lambda A + \mu B] = \lambda^2 Var_f[A] + \mu^2 Var_f[B] + 2\lambda\mu Cov_f[A,B] \qquad \text{(eq. A.2.18)}$$

$$Cov_f[\lambda A + \mu B, C] = \lambda Cov_f[A,C] + \mu Cov_f[B,C] \qquad \text{(eq. A.2.19)}$$

Une mesure de corrélation possibiliste peut être définie, comme en théorie des probabilités :

$$\rho_f[A,B] = \frac{Cov_f[A,B]}{\sqrt{Var_f[A] Var_f[B]}} \qquad \text{(eq. A.2.20)}$$

2.2. Théorie des possibilités

2.2.1. Mesures possibilistes

La théorie des possibilités est basée sur le concept de variable floue. Elle permet de manipuler différentes mesures permettant de caractériser l'incertitude et l'imprécision d'évènements.

Considérons une variable floue x de domaine U et de distribution π_x. Soit A un évènement possibiliste, c'est-à-dire un sous-ensemble de U. On dit que A est réalisé quand la valeur de x appartient à A. La *mesure de possibilité* de A, notée $\Pi(A)$, représente la possibilité que A se réalise (Figure A.2.3). Elle est définie grâce à la distribution de possibilités de x :

$$\Pi(A) = \sup_{a \in A} \pi_x(a) \qquad \text{(eq. A.2.21)}$$

On voit tout de suite que si l'intersection de A et du support de π_x est vide (Figure A.2.3), alors l'évènement est impossible, et on a $\Pi(A) = 0$.

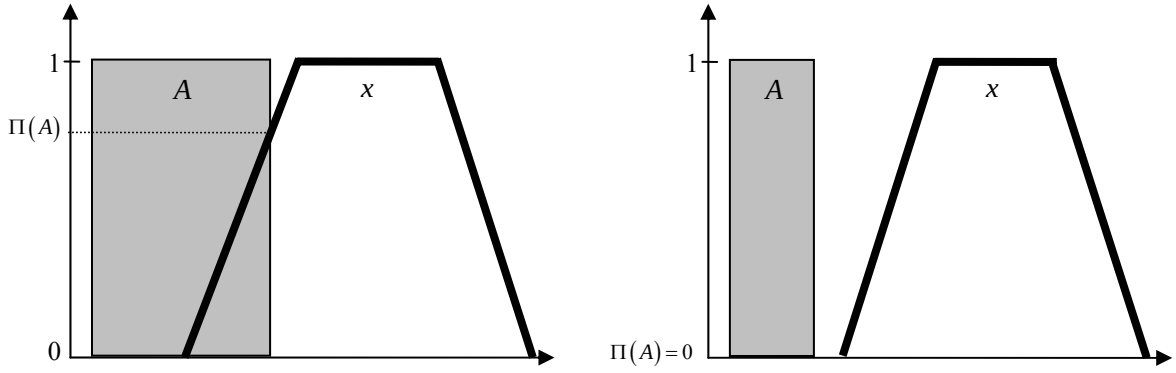

Figure A.2.3 : mesures de possibilité d'un évènement.

Une mesure de possibilités Π possède les propriétés suivantes :

$$\Pi(\varnothing) = 0 \qquad \text{(eq. A.2.22)}$$

$$\Pi\left(\bigcup A_i\right) = \max\left(\Pi(A_i)\right) \qquad \text{(eq. A.2.23)}$$

$$\Pi\left(\bigcap A_i\right) \leq \min\left(\Pi(A_i)\right) \qquad \text{(eq. A.2.24)}$$

où $\{A_i\}$ est un ensemble dénombrable d'évènements [Drakopoulos '95]. Comme en théorie des probabilités, on notera dorénavant $\Pi(A,B)$ la mesure de possibilité de $A \cap B$. Si de plus on a :

$$\Pi(U) = 1 \qquad \text{(eq. A.2.25)}$$

Alors on dit que la mesure de possibilités est *normalisée*, et on a la propriété :

$$\Pi\left(A \cup \bar{A}\right) = \max\left(\Pi(A), \Pi\left(\bar{A}\right)\right) = 1 \qquad \text{(eq. A.2.26)}$$

On associe à la mesure de possibilité $\Pi(A)$ une *mesure de nécessité* notée $N(A)$ et définie par :

$$N(A) = \min_{a \in A}\left[1 - \pi_x(a)\right] = 1 - \Pi\left(\bar{A}\right) \qquad \text{(eq. A.2.27)}$$

où $\bar{A}$ désigne le complémentaire de A dans U (Figure A.2.4). Par définition, on a donc :

$$N(A) > 0 \rightarrow \Pi(A) = 1 \qquad \text{(eq. A.2.28)}$$

On remarque que si A contient le support de π_x, alors $\Pi\left(\bar{A}\right) = 0$, d'où $N(A) = 1$. On dit que l'évènement est *certain*.

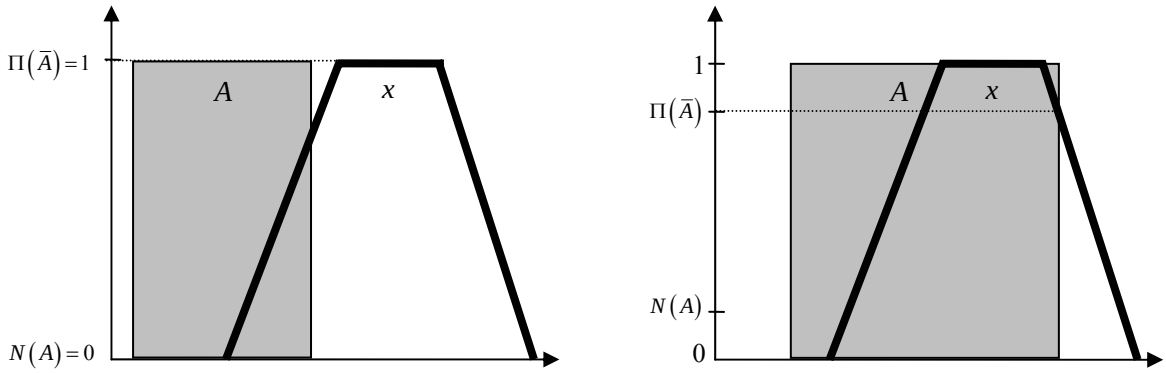

Figure A.2.4 : mesures de nécessité d'un évènement.

A noter qu'à la différence d'une distribution de probabilités, une distribution de possibilités n'est pas normalisée pour la somme. C'est-à-dire que pour une partition $\{B_i\}$ de Ω, on a $\sum P(B_i) = 1$, mais en général, ce n'est pas vrai pour les possibilités.

Du point de vue de l'interprétation, $\Pi(A)$ quantifie la possibilité que A se réalise. Si $\Pi(A) = 1$, cela signifie que l'évènement est possible sans restriction. Si $\Pi(A) = 0$, alors il est impossible que A arrive. Si $\Pi(A)$ prend une valeur intermédiaire, alors cela signifie que l'évènement est possible,

mais avec des restrictions. Cela permet de traduire le fait que l'on a connaissance d'informations contradictoires laissant penser à la fois que A peut et ne peut pas se réaliser.

2.2.2. Interprétation, possibilités conditionnelles & indépendance

L'interprétation donnée précédemment est très générale et très vague. A la manière des différentes approches (empirique et subjective) de la théorie des probabilités, la sémantique des possibilités a été raffinée en de nombreuses interprétations [Borgelt *et al.* '98; Dubois & Prade '93]. A ce niveau, il existe une différence importante entre la théorie des probabilités et celle des possibilités. En effet, dans cette dernière l'expression formelle des concepts de lien conditionnel et d'indépendance entre des évènements dépend de la sémantique des possibilités choisie. Il existe par exemple plus d'une dizaine de définitions de la possibilité conditionnelle [Walley & de Cooman '99].

Dubois et Prade distinguent deux grandes catégories d'interprétations : l'une ordinale et l'autre numérique [Dubois & Prade '94]. Dans le premier cas, la théorie des possibilités est vue comme un outil qualitatif qui permet d'ordonner différentes incertitudes, par opposition à la théorie des probabilités, qui manipule l'incertitude de façon quantitative. Dans la seconde approche, on envisage une mesure de possibilité comme la borne supérieure d'une mesure de probabilité mal connue. Dans le cas ordinal, la définition la plus générale de la possibilité conditionnelle est qu'elle doit être telle qu'elle vérifie l'équation suivante [Benferhat *et al.* '99]:

$$\Pi(A,B) = \min\left(\Pi(A|B), \Pi(B)\right) \qquad \text{(eq. A.2.29)}$$

On a donc deux cas possibles :

$$\text{Si } \Pi(A,B) < \Pi(B) \text{ alors } \Pi(A|B) = \Pi(A,B) \qquad \text{(eq. A.2.30)}$$

$$\text{Si } \Pi(A,B) = \Pi(B) \text{ alors } \Pi(A,B) \leq \Pi(A|B) \leq 1 \qquad \text{(eq. A.2.31)}$$

Dans le cas numérique, la définition est similaire à celle utilisée pour les probabilités :

$$\Pi(A|B) = \frac{\Pi(A,B)}{\Pi(B)} \text{ avec } \Pi(B) \neq 0 \qquad \text{(eq. A.2.32)}$$

Cette approche quantitative est la plus intéressante dans le contexte qui est le nôtre, c'est-à-dire la modélisation d'un système via des variables numériques continues, pour lesquelles on veut pouvoir quantifier l'imprécision ou l'incertitude.

De la même façon qu'il existe plusieurs définitions de possibilité conditionnelle, il existe évidemment un grand nombre de définitions pour l'indépendance possibiliste entre des évènements [de Cooman '97], puisque celle-ci est exprimée en général en utilisant des possibilités conditionnelles. Dans le cadre d'une approche quantitative, on peut utiliser une définition analogue à celle de la théorie des probabilités , c'est-à-dire que A et B sont indépendants si et seulement si [Benferhat *et al.* '01] :

$$\Pi\left(A|B\right)=\Pi\left(A\right) \text{ ou bien } \Pi\left(B|A\right)=\Pi\left(B\right) \qquad \text{(eq. A.2.33)}$$

3. CALCULS QUALITATIF ET SEMI-QUALITATIF

3.1. Calcul qualitatif

Le principe du calcul (ou raisonnement) qualitatif est de ne traiter que certains aspects caractéristiques de d'information, et pas forcément l'information elle-même. Cela permet d'effectuer des calculs sans connaître précisément cette information. Ces aspects caractéristiques sont appelés *qualités*, et peuvent être assimilés à des symboles plus ou moins ordonnés. Des *algèbres qualitatives* permettent de manipuler les qualités comme des entités algébriques. L'ensemble des qualités décrivant une variable est appelé *espace des qualités*. En calcul qualitatif, l'imprécision est traitée de manière *implicite*, elle dépend des choix que l'on fait quand on détermine les aspects qualitatifs de l'information qui vont être retenus. Pour reprendre un exemple de Marc Lafon [Lafon '00], considérons une variable représentant la température de l'eau. Cette variable peut être caractérisée par l'état de l'eau: glacé, liquide ou gazeux. Mais on peut aussi choisir d'être plus précis en utilisant un espace des qualités plus important : glacé, froid, frais, tiède, chaud, bouillant, gazeux…

3.1.1. Algèbre des signes

L'algèbre des signes [Travé-Massuyès *et al.* '97] est le plus simple formalisme qualitatif existant. Elle utilise des catégories décrivant les réels selon leur signe (positif, négatif, nul, indéterminé). Le plus souvent, elle s'applique aux modèles dans lesquels l'évolution d'un système est décrite par les tendances des variables d'état du système. L'espace des qualités S est constitué de quatre symboles :

$$S = \{-, 0, +, ?\} \qquad \text{(eq. A.3.1)}$$

Lorsque ces symboles désignent des variations, ils représentent une évolution décroissante, stationnaire, croissante, et enfin indéterminée. On peut définir un ordre sur les trois permiers symboles : $- < 0 < +$. Deux opérations sont définies dans cette algèbre : l'addition qualitative $\oplus$, et la multiplication qualitative $\otimes$. Le symbole ? permet de désigner le résultat d'opérations telles que $+ \oplus -$ ou $+ \otimes ?$.

On définit également une égalité (bien qu'elle ne définisse pas une classe d'équivalence) qualitative sur S, noté $\approx$. Pour tout symbole s de l'ensemble des qualités, cette relation vérifie $s \approx s$, $s \approx ?$, et $? \approx s$.

Le nombre restreint de symboles manipulés réduit de manière évidente la quantité de calculs nécessaire au traitement des données du modèle. Toutefois, cette algèbre possède un défaut important, puisque l'information manipulée est grossière, ce qui entraîne de nombreuses indéterminations, et rend difficile son utilisation dans le cadre de systèmes complexes.

3.1.2. Ordres de grandeur

Pour gagner en précision, le calcul qualitatif utilisant les ordres de grandeurs permet d'accroître le nombre de qualités utilisées pour décrire l'information. Cette amélioration de la précision se fait au détriment de la complexité des calculs, c'est pourquoi, suivant le système modélisé, un compromis doit être trouvé. On distingue deux types d'ordres de grandeur : ils peuvent être absolus ou relatifs.

Avec les ordres de grandeurs absolus, l'ensemble des qualités S est une partition de l'ensemble des réels $\mathbb{R}$, plus fine que celle utilisée pour l'algèbre des signes. Les symboles ont la propriété d'être naturellement ordonnés par $<$. Par exemple, prenons : $S = \{NG, NP, 0, PP, PG\}$, les symboles

représentant respectivement les grands nombres négatifs, les petits nombres négatifs, zéro, les petits nombres positifs et les grands nombres positifs. On a alors $NG < NP < 0 < PP < PG$.

A partir des symboles de S, il est possible de construire d'autres symboles. Soient $s_1, s_2 \in S$ tels que $s_1 < s_2$. Le nouveau symbole $[s_1, s_2]$ est constitué de tous les éléments constituant s_1, s_2, ainsi que de tous les éléments qui séparent s_1 et s_2. L'ensemble, noté S^*, constitué de l'union de S et des $[s_1, s_2]$ est appelé *univers de description*. Pour notre exemple, S^* contient donc, en plus de $\{NG, NP, 0, PP, PG\}$, l'ensemble des nombres négatifs $[NG, NP]$, l'ensemble des petits nombres $[NP, PP]$, l'ensemble de tous les nombres $[NG, PG]$, etc…¨

On définit une relation d'ordre sur S^*, notée $\prec$, telle que pour $s_1, s_2 \in S^*$:

$$s_1 \prec s_2 \text{ ssi } \forall x \in \mathbb{R} : x \in s_1 \rightarrow x \in s_2 \qquad \text{(eq. A.3.2)}$$

Dans notre exemple, on a ainsi $[NP, 0] \prec [NP, PP]$. On définit également une relation d'égalité :

$$s_1 \approx s_2 \text{ ssi } \exists s_3 \in S^* : s_3 \in s_1 \wedge s_3 \in s_2 \qquad \text{(eq. A.3.3)}$$

A l'aide de ces relations, on peut construire des abstractions des fonctions à valeurs dans $\mathbb{R}$, appelées *fonctions qualitatives associées*.

Dans les ordres de grandeurs relatifs, on ne réalise pas une partition *absolue* de $\mathbb{R}$. La partition se fait *relativement* à une qualité considérée. Par exemple, pour une qualité donnée s, on va partager $\mathbb{R}$ en les ensembles suivants : valeurs très inférieures à s, valeurs modérément inférieures à s, valeurs légèrement inférieures à s, etc. Pour réaliser ces comparaisons, il est nécessaire de définir des opérateurs sur les ordres de grandeurs. Il existe plusieurs formalismes, qui diffèrent dans les opérateurs employés, et sont plus ou moins adaptés suivant le contexte de modélisation.

3.2. Calcul semi-qualitatif

3.2.1. Arithmétique des intervalles

Le raisonnement sur les intervalles constitue une généralisation du raisonnement sur les ordres de grandeur, permettant de fournir des résultats plus précis. Considérons l'espace des qualités constitué de tous les réels, et tel que chacun de ses éléments est un réel (le nombre de qualités est infini). L'univers de description correspond alors à l'ensemble des intervalles fermés réels. Donc, raisonner sur les intervalles, revient à raisonner sur une algèbre des ordres de grandeurs avec la partition la plus fine possible [Travé-Massuyès *et al.* '97]. Cette approche permettant de faire cohabiter des quantités (les réels) et des qualités (les intervalles) constitue donc un pont entre le raisonnement qualitatif et le raisonnement quantitatif.

Une caractéristique essentielle de cette approche est qu'il n'est pas nécessaire de spécifier une partition *a priori* de $\mathbb{R}$. Pour chaque variable, on définit non plus un ensemble fini de qualités, mais un *domaine*, c'est à dire un intervalle réel qui contient l'ensemble des valeurs que la variable peut prendre.

De plus, dans l'arithmétique des intervalles, les variations des valeurs des variables se font progressivement, en raison du nombre infini de qualités, et non plus par à coup, comme c'est le cas pour une variable qui passerait de PP à PG dans le cadre du raisonnement sur les ordres de grandeur absolus. Cette continuité dans les variations constitue un atout pour la modélisation dynamique, puisqu'il est de ce fait possible de traiter des états intermédiaires dans les évènements qui affectent les variables.

Quand on utilise cette arithmétique pour effectuer les calculs dans un modèle qualitatif, on ne détermine pas le comportement de chaque variable lors de la simulation, comme c'est le cas avec des algèbres purement qualitatives, mais un ensemble contenant la valeur de chaque variable à chaque instant, sous la forme d'un intervalle. C'est pourquoi on qualifie généralement ce traitement par intervalles d'approche *semi-qualitative*.

Le calcul des intervalles a été formalisé, notamment, par Moore [Moore '66]. A la manière de l'égalité définie pour les ordres de grandeurs (eq. A.3.3), deux intervalles sont qualitativement égaux si leur intersection est non-vide. De ce fait, cette relation d'égalité est intransitive :

$$[a,b] \approx [c,d] \text{ ssi } \exists x \in \mathbb{R} : x \in [a,b] \wedge x \in [c,d] \quad \text{(eq. A.3.4)}$$

Les opérations d'addition et de multiplication qualitatives sont définies par :

$$I + J \triangleq \{x + y : x \in I \wedge y \in J\} \quad \text{(eq. A.3.5)}$$

$$I \times J \triangleq \{x \times y : x \in I \wedge y \in J\} \quad \text{(eq. A.3.6)}$$

où I et J sont des intervalles, et x et y des réels. On définit également l'opposé d'un intervalle I :

$$-I \triangleq \{-x : x \in I\} \quad \text{(eq. A.3.7)}$$

L'inverse n'est défini que pour les intervalles ne contenant pas 0 :

$$I^{-1} \triangleq \{x^{-1} : x \in I\} \quad \text{(eq. A.3.8)}$$

Les fonctions exponentielles, logarithmiques et trigonométriques sont définies de la même façon [Travé-Massuyès *et al.* '97]. Le calcul peut se faire directement sur les bornes des intervalles [Bonarini & Bontempi '94]:

$$[a,b] + [c,d] \triangleq [a + c, b + d] \quad \text{(eq. A.3.9)}$$

$$[a,b] - [c,d] \triangleq [a - d, b - c] \quad \text{(eq. A.3.10)}$$

$$[a,b] \times [c,d] \triangleq \left[\min(a \times c, a \times d, b \times c, b \times d), \max(a \times c, a \times d, b \times c, b \times d)\right] \quad \text{(eq. A.3.11)}$$

$$\frac{[a,b]}{[c,d]} \triangleq \left[\min\left(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right), \max\left(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right)\right] \text{ (avec } 0 \notin [c,d]) \quad \text{(eq. A.3.12)}$$

3.2.2. Propriétés

Les opérateurs sur les intervalles ont les propriétés suivantes :

Commutativité : $I + J = J + I$ et $I \times J = J \times I$ (eq. A.3.13)

Associativité : $(I + J) + K = I + (J + K)$ et $(I \times J) \times K = I \times (J \times K)$ (eq. A.3.14)

La *distributivité* n'est pas toujours vraie :

$$\exists I, J, K : I \times (J + K) \neq I \times J + I \times K \quad \text{(eq. A.3.15)}$$

Par exemple, prenons :

$$[a,b] \times (1-1) = 0 \quad \text{(eq. A.3.16)}$$

Pour $a \neq b$, on obtient en distribuant $[a,b]$:

$$[a,b] - [a,b] = [a,b] + [-b,-a] = [a-b, b-a] \neq [0,0] \quad \text{(eq. A.3.17)}$$

De plus, le résultat du produit d'un intervalle et de son inverse n'est pas l'unité. Pour un intervalle $[a,b]$ ne contenant pas 0, on a :

$$\forall a \neq b : \frac{[a,b]}{[a,b]} = \left[\min\left(\frac{a}{a}, \frac{a}{b}, \frac{b}{a}, \frac{b}{b}\right), \max\left(\frac{a}{a}, \frac{a}{b}, \frac{b}{a}, \frac{b}{b}\right) \right] \neq [1,1] \quad \text{(eq. A.3.18)}$$

De façon plus générale, l'arithmétique des intervalles ne tient pas compte des dépendances entre les variables. En d'autres termes, si la même variable apparaît plusieurs fois dans une fonction, on considère ses occurrences comme autant de valeurs différentes [Travé-Massuyès *et al.* '97]. A cause de cette restriction, le résultat de l'évaluation de la fonction en utilisant le calcul des intervalles peut être différent de l'ensemble des valeurs que peut prendre la fonction évaluée sur les réels. Toutefois, les solutions réelles sont toujours contenues dans le résultat du calcul sur les intervalles :

$$f(I_1, \ldots, I_n) \supseteq \{ f(x_1, \ldots, x_n) : x_i \in I_i \} \quad \text{(eq. A.3.19)}$$

4. NEURONE FORMEL

Un *neurone formel* (NF) (formel par opposition au neurone biologique) est une fonction bornée, généralement non-linéaire, et dépendant de paramètres [Dreyfus '98]. Soit le NF constitué de la fonction $f(x_1,\ldots,x_n)$ et ayant $w_1,\ldots,w_n$ pour paramètres. La Figure A.4.1 est une représentation graphique de ce NF.

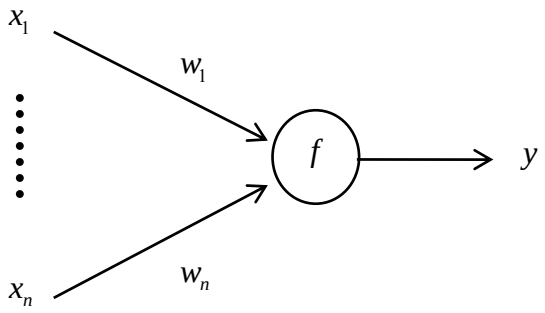

Figure A.4.1 : représentation graphique d'un neurone formel.

On appelle $x_1,\ldots,x_n$ les entrées du neurone, et $y = f(x_1,\ldots,x_n)$ est sa sortie. La fonction f est en fait la composition de deux fonctions : une *fonction d'activation* f_a qui calcule la *valeur d'activation*, à laquelle est appliquée une *fonction de seuil* f_s, ou *fonction de transfert*, qui calcule la *valeur d'émission* :

$$f(x_1,\ldots,x_n) = f_s\left(f_a(x_1,\ldots,x_n)\right) \qquad \text{(eq. A.4.1)}$$

La fonction d'activation est souvent définie comme la somme des entrées pondérées par les paramètres du neurone :

$$f_a(x_1,\ldots,x_n) = \sum w_i x_i \qquad \text{(eq. A.4.2)}$$

La fonction de seuil, elle, est en général une fonction non-linéaire dans $[0,1]$. Cette décomposition est due au fait que la fonction f modélise un neurone biologique (NB). Les entrées du NF correspondent aux *potentiels post-synaptiques* (PPS) (c.f. la description du neurone naturel, chapitre I.1.1.2) qui arrivent au NB. Dans le NB, les *dendrites* transmettent ces PPS, qui peuvent être excitateurs ou inhibiteurs. Ce mécanisme est représenté par les paramètres, ou *poids*, du NF : un

poids positif représente une connexion excitatrice et un poids négatif une connexion inhibitrice. Puis les PPS du NB sont intégrés au niveau du *soma*, ce qui correspond à la somme effectuée dans le NF par f_a. Suivant la valeur du potentiel intégré, le *cône d'emboîtement* du NB cause alors la décharge neuronale, ou pas. Cela correspond dans le NF à l'application de la fonction f_s. Enfin, *l'axone* du NB transmet le potentiel d'action vers les NB afférents, ce qui revient à propager la valeur y dans le NF.

On distingue différents types de neurones, suivant la nature du signal qui est traité, et suivant les fonctions f_a et f_s utilisées. Par exemple, dans le modèle initial, de Mc Culloch et Pitts [McCulloch & Pitts '43] les valeurs sont booléennes, f_a est la somme pondérée (eq. A.4.2), et f_s est la fonction de *Heavyside*, c'est-à-dire que le NF ne décharge (il n'émet un signal) que si sa valeur d'activation dépasse un certain seuil θ :

$$f_s(x) = \begin{cases} 0 \ si \ x \le \theta \\ 1 \ si \ x > \theta \end{cases} \qquad \text{(eq. A.4.3)}$$

C'est souvent la fonction *tangente hyperbolique* (ou *sigmoïde*) qui est utilisée pour f_s :

$$f_s(x) = \frac{1}{1+e^{-k(x-\theta)}} \qquad \text{(eq. A.4.4)}$$

où k est un paramètre de la fonction déterminant son écrasement.

ANNEXE B

ALGORITHMES ET IMPLEMENTATION

RAGE s'appuie en grande partie sur les réseaux bayésiens dynamiques, en particulier les modèles d'espaces d'états, pour tout ce qui concerne la représentation et la propagation de la magnitude. Mais toute la partie du modèle dédiée au calcul des types et de la TPT nécessite un traitement particulier, puisque cet aspect est propre à notre formalisme. Dans cette annexe, nous présentons les structures de données que nous utilisons pour représenter notre information duale, ainsi que les algorithmes qui permettent de traiter cette information et d'appliquer ainsi les principes de propagation et d'apprentissage décrits dans les chapitres précédents. Puis, nous décrivons comment nous avons implémenté ces structures de données et algorithmes pour réaliser une version utilisable de notre outil de modélisation.

1. ORGANISATION GENERALE

De façon générale, la décomposition temporelle employée lors de la définition d'un modèle RAGE fait que celui-ci prend la forme d'un réseau simplement connecté. L'information qui est manipulée est duale, et ses deux composantes ne s'influencent qu'en des points très précis (lors du calcul des valeurs d'émission). En fait, on peut voir un modèle comme deux réseaux parallèles, dédiés au traitement de la magnitude pour l'un et des types pour l'autre (Figure B.1.1).

Ces deux réseaux peuvent néanmoins avoir des contacts, puisque nous avons vu que le type d'activation et la TPT peuvent influencer la magnitude d'émission (pour la même population neuronale), et de même pour la magnitude d'activation, qui peut influer sur le type d'émission et la TPT. Dans le premier cas, la communication se fait en considérant le type d'activation comme une entrée de l'état (c.f. la variable U_k dans la Figure III.1.8) définissant la magnitude d'émission. Dans l'autre sens, nous nous utilisons une méthode issue de la littérature. Elle est employée notamment lors de l'estimation simultanée de l'état et des paramètres dans des filtres de Kalman non-linéaires

par les algorithmes dits d'*estimation double* [Wan & Nelson '96]. Dans ce type d'algorithmes, deux filtres sont utilisés en même temps. L'un calcule les probabilités *a posteriori* de l'état conditionnellement aux observations et à une estimation précédente des paramètres. Le second calcule les probabilités *a posteriori* des paramètres conditionnellement aux observations et à une estimation précédente de l'état. La communication entre les deux filtres se fait via l'échange des estimations qu'ils ont calculées, c'est-à-dire des moyennes *a posteriori*. Ce type d'estimation est caractérisé par une bonne stabilité, mais il a toutefois le défaut d'ignorer l'incertitude lors de l'échange d'information, ce qui peut aboutir au calcul de matrices d'erreurs d'estimation plus resserrées qu'elles ne le sont vraiment [Roweis & Ghahramani '01]. A l'instar de l'échange d'information entre les deux RBD, qui se fait via les moyennes, nous utilisons la moyenne de la magnitude d'activation lors du processus de calcul du type d'activation.

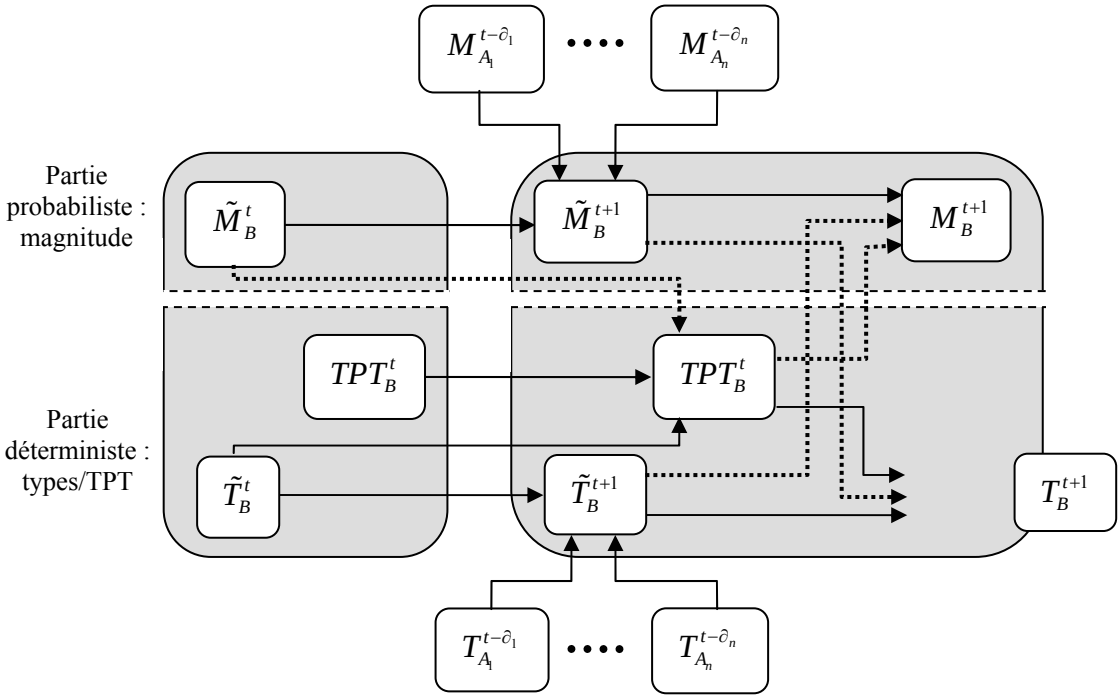

Figure B.1.1 : décomposition de la Figure IV.5.1 en deux sous-réseaux, l'un dédié à la magnitude et l'autre aux types et TPT.
Les relations inter-réseaux sont représentées en pointillés.

Nous avons utilisé cette propriété de décomposition de nos modèles en deux sous-réseaux pour profiter des travaux réalisés dans le domaine des filtres de Kalman non-linéaires. En effet, plusieurs algorithmes de propagation dédiées aux filtres de Kalman non-linéaires ont fait leur apparition [Nørgaard *et al.* '00; van der Merwe & Wan '03b]. La communauté travaillant sur les filtres de

Kalman non-linéaires est importante, et les algorithmes mis à notre disposition sont efficaces. Pour ces raisons, nous avons décidé d'abandonner l'idée de réaliser complètement l'algorithme de propagation de l'information dans le réseau. Nous plutôt choisi d'utiliser un des algorithmes dédiés aux filtres de Kalman non-linéaires existants pour traiter la partie magnitude de l'information, et d'y adjoindre un algorithme de traitement des types.

2. REPRESENTATION DES DONNEES

2.1. Information

La magnitude étant une v.a. de distribution normale, sa moyenne et sa variance suffisent à la caractériser. Par conséquent, nous représentons une magnitude par deux réels. Dans l'algorithme de filtrage de Kalman non-linéaire donné dans le chapitre III.2.1.3, toutes les v.a. faisant partie du vecteur d'état sont regroupées, et représentées sous la forme d'un vecteur de moyennes et d'une matrice de covariance.

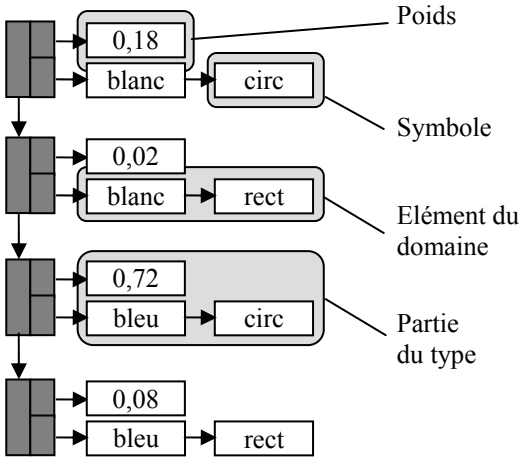

Figure B.2.1 : exemple d'utilisation de la structure de données utilisée pour représenter un type.

Le type est une structure de données plus complexe. Il s'agit d'une liste de couples associant un élément d du domaine de définition D, et un réel représentant le poids p associé à d. Un élément du domaine de définition d est lui-même représenté par la liste des n symboles $(s_1,\ldots,s_n)$ qui le composent. Un symbole est une simple chaîne de caractères. On ne représente

pas tous les éléments du domaine de définition du type, mais seulement ceux auxquels est associé un poids non-nul.

La Figure B.2.1 donne un exemple d'utilisation de cette structure de données pour représenter le type T_g de l'exemple du chapitre V.1.2.3.b (eq. V.1.30). Les flèches représentent les liens formant les listes. On appelle *partie* du type la paire (élément, poids). Par la suite, dans le paragraphe décrivant les algorithmes, l'élément de domaine appartenant à une partie L sera noté `L.élément`, et le poids associé à cet élément sera noté `L.poids`.

2.2. TPT

La TPT est représentée par une liste de triplet associant un type (tel qu'il est décrit plus haut), un réel, et un symbole.

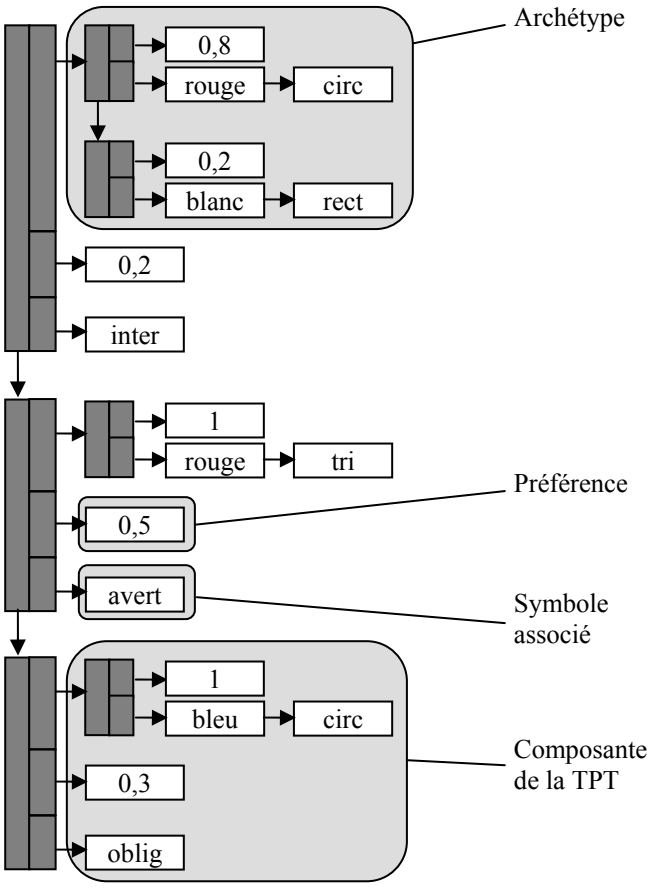

Figure B.2.2 : exemple d'utilisation de la structure de données utilisée pour représenter une TPT.

Le type correspond à un archétype, le réel est la préférence correspondante, c'est-à-dire la valeur renvoyée par la fonction *pref* , et le symbole est la valeur renvoyée par la fonction *symb* , c'est-à-dire le symbole associé à l'archétype lors de la construction du type d'émission.

La Figure B.2.2 illustre cette description par un exemple. On appelle *composante* de la TPT le triplet (archétype, préférence, symbole associé). Par la suite, dans la partie décrivant les algorithmes, l'archétype d'une composante K sera noté `K.archétype`, la préférence de l'archétype sera notée `K.préférence`, et son symbole associé `K.symb`.

2.3. Fonctions pour le traitement et l'apprentissage

En ce qui concerne le traitement de l'information implémenté par les nœuds dynamiques, on peut faire une distinction entre les mécanismes laissant beaucoup de liberté à l'utilisateur et ceux qui sont très contraints.

Le premier cas regroupe les fonctions $\tilde{f}_{M_X}$ (magnitude d'activation (eq. V.2.14)), $\tilde{f}_{T_X}$ (type d'activation (eq. V.2.15)), $f_{T_X}^{(1)}$ (type d'émission (eq. V.2.27)), $f_{M_X}^{(1)}$ et $f_{M_X}^{(2)}$ (magnitude d'émission (eq. V.2.37)), et $f_{TPT_X}^{(1)}$ et $f_{TPT_X}^{(2)}$ (apprentissage (eq. V.3.1), (eq. V.3.2) et (eq. V.3.6)). L'utilisateur est libre de définir les fonctions remplissant ces rôles, tout en respectant les propriétés de notre formalisme. Ainsi, $\tilde{f}_{T_X}$ ne doit être définie qu'en utilisant des $\oplus$ et $\otimes$, et les fonctions $f_{T_X}^{(1)}$, $f_{M_X}^{(1)}$, et $f_{TPT_X}^{(1)}$ doivent respecter certaines contraintes (croissance, domaine de définition, c.f. chapitre V).

Le second cas désigne la fonction $f_{T_X}^{(2)}$ dans le calcul du type d'émission (eq. V.2.27), qui n'est pas une fonction directement définie par l'utilisateur, puisqu'elle dépend uniquement de la TPT.

Les premières fonctions doivent être représentées de façon à pouvoir être facilement définies et modifiées, puisqu'elles dépendent de l'utilisateur. De plus, leur représentation doit également permettre de vérifier leur validité, par rapport aux contraintes portant sur les fonctions définies dans RAGE (par exemple, ce qui concerne la combinaison des types). Pour cela, nous utilisons une décomposition en fonctions et opérateurs élémentaires, qui prend l'aspect d'un graphe. Par

exemple, les Figure B.2.3.a et Figure B.2.3.b sont les représentations respectives des deux fonctions suivantes :

$$T = \left(T_1 \oplus T_2\right) \otimes T_3 \tag{B.2.1}$$

$$M = M_{\ 1}\left(aM_2 + bM_3\right) + cM_4 \tag{B.2.2}$$

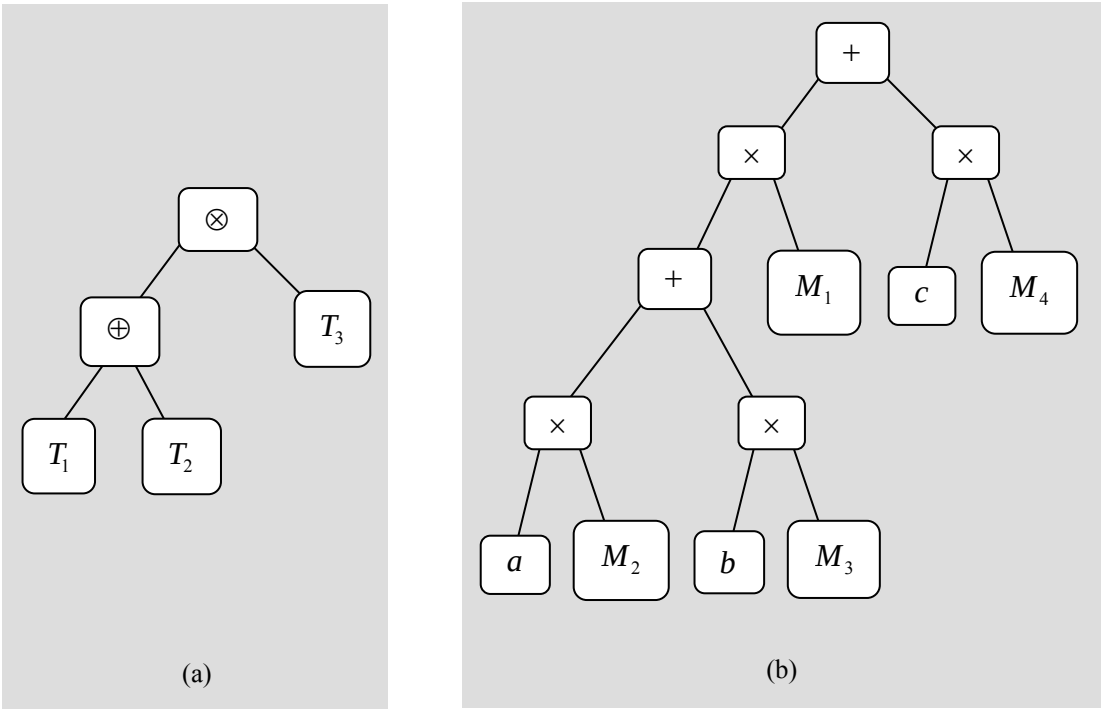

Figure B.2.3 : exemple d'utilisation de la structure de données utilisée pour représenter une fonction.
(a) : fonction exprimant un type. (b) : fonction exprimant une magnitude.

3. ALGORITHMES

Les algorithmes les plus remarquables se divisent en deux : un ensemble d'outils utilisés à différentes étapes de la simulation, et les algorithmes de simulation à proprement parler. Les premiers portent essentiellement sur la manipulation des types. Dans les seconds, on trouve l'algorithme de propagation, qui permet de calculer les magnitudes et types d'activation et d'émission, et l'algorithme d'apprentissage, qui permet de faire évoluer la TPT.

3.1. Fonctions auxiliaires

3.1.1. Distance & similitude

Dans le cas des fonctions de similitude linéaire et sigmoïdale, le calcul de la similitude entre de deux types passe d'abord par le calcul de la distance entre les deux types, grâce à l'algorithme suivant :

```
Etant donnés :
T1, T2 : les deux types à comparer

Soient :
dist    : distance entre les deux types
L       : partie d'un type
D       : élément d'un domaine de définition
p1, p2  : poids

dist←0

Pour toute partie L de T1
  p1 ← L.poids
  d ← L.élément
  Si d est présent dans T2
    p2 ← poids associé à d dans T2
  Sinon
    p2 ← 0
  Fin
  dist ← dist + 0.5 * abs(p1 - p2)
Fin
Pour toute partie L de T2
  p2 ← L.poids
  d ← L.élément
  Si d est absent de T1
    dist ← dist + 0.5 * p2
  Fin
Fin
```

Il est inutile de détailler un algorithme pour le calcul de la similitude, puisque celle-ci s'obtient grâce à des fonctions très simples de la distance dans le cas des fonctions de similitude sigmoïdale (eq. V.2.22) ou linéaire (eq. V.2.20). En ce qui concerne la similitude employant le cosinus, là aussi, le calcul ne nécessite pas d'algorithme particulier, puisqu'il suffit de calculer de façon classique le produit scalaire et la norme des vecteurs formés par les poids des types (eq. V.2.18).

3.1.2. Combinaison de types

a. Combinaison linéaire

La combinaison linéaire de types repose sur les principes décrits dans le chapitre V.1.2.3.a. Les types doivent être de même domaine, et le résultat est également de même domaine. Le calcul porte donc uniquement sur les poids :

```
Etant donnés :
V   : tableau des types à combiner
C   : tableau des coefficients associés

Soient :
Tres      : résultat
Targ      : type inclus dans V
Carg      : coefficient de C associé à ce type
Larg, Lres : parties d'un type
d         : élément d'un domaine de définition

Tres ← initialisation
total ← somme des coefficients de C

Pour tout type Targ de V
  Pour toute partie Larg de Targ
    d ← Larg.élément
    Si d est absent de Tres
      introduction d'une nouvelle partie Lres dans Tres
      Lres.élément ← d
      Lres.poids ← 0
    Fin
    Lres.poids ← Lres.poids + (Carg/total)* Larg.poids
  Fin
Fin
```

b. Combinaison non-linéaire

Les principes de la combinaison non-linéaire de types ont été décrits dans le chapitre V.1.2.3.b. Les types à combiner ne sont pas forcément définis sur le même domaine. Cette fois, le traitement a lieu sur les poids et sur les symboles :

```
Etant donné :
V   : tableau des types à combiner

Soient :
Tres      : résultat
Targ      : type inclus dans V
T         : type auxiliaire
Larg, Lres : parties d'un type
```

```
L        : partie auxiliaire
d_arg, d_res : éléments d'un domaine de définition

T_res ← initialisation

Pour tout type T_arg de V
  Si T_res est vide
    T_res ← T_arg                                    //première itération
  Sinon
    T ← initialisation
    Pour toute partie L_arg de T_arg
      d_arg ← L_arg.élément
      Pour toute partie L_res de T_res
        d_res ← L_res.élément
        introduction d'une nouvelle partie L dans T
        L.élément ← symboles de d_arg et de d_res
        L_res.poids ← L_arg.poids * L_res.poids
      Fin
    Fin
    T_res ← T
  Fin
Fin
```

3.2. Traitement principal

3.2.1. Propagation

Le calcul du type d'activation repose uniquement sur l'utilisation des combinaisons linéaire et non-linéaire de types, dont les algorithmes ont été donnés précédemment. Les magnitudes d'activation et d'émission sont déterminées en appliquant aux fonctions définies par l'utilisateur l'algorithme du filtre de Kalman non-linéaire décrit dans le chapitre III.2.1.3.

Le type d'émission est beaucoup plus contraint par le formalisme. Voici l'algorithme permettant de le calculer à partir des magnitude et type d'activation, et de la TPT :

```
Etant donnés :
TPT      : table de préférence des types
T_act    : type d'activation
M_act    : magnitude d'activation

Soient :
T_bruit  : type représentant du bruit
T_em     : type d'émission
T        : type auxiliaire
K        : composante de TPT
L        : partie de type
total    : poids total
sim      : valeur de similitude
```

```
total ← 0

Pour toute composante K de la TPT
  T ← K.archetype
  sim ← similitude entre T et T_act
  total ← total + sim * K.preference
Fin

T_em ← initialisation

Pour toute composante K de la TPT
  T ← K.archetype
  sim ← similitude entre T et T_act
  poids ← (sim * K.preference)/total
  introduction d'une nouvelle partie L dans T_em
  L.élément ← K.symbole
  L.poids ← poids
Fin
T_em ← ⊕(T_em, T_bruit, f(M_act), 1-f(M_act))
```

où `f(M_act)` représente le résultat de $f_{T_X}^{(1)}\left(\overline{\tilde{M}_X^t}\right)$, dans l'équation (eq. V.2.27).

3.2.2. Apprentissage

Les mécanismes d'apprentissage sont complexes, et essentiellement basés sur les types. Voici l'algorithme qui permet de réviser la TPT en fonction de l'ancienne TPT et des magnitude et type d'activation :

```
Etant donnés :
TPT      : table de préférence des types
T_act    : type d'activation
M_act    : magnitude d'activation

Soient :
c_r      : coefficient de renforcement
s_o      : seuil d'oubli
c_g      : coefficient de glissement
s_f      : seuil de fusion
T, U     : types auxiliaires
K, J     : composantes de TPT
L        : partie de type
sim      : valeur de similitude
simax    : similitude maximale
varPref  : variation de préférence
totalPref                                  : somme des préférences
varGliss : variation des poids
pref_act : préférence de l'éventuel nouvel archétype

simax ← 0
totalPref ← 0
```

```
Pour toute composante K de la TPT
  T ← K.archetype
  sim ← similitude entre T et T_act
  Si sim > simax
    simax ← sim
  Fin
  varPref ← c_r * f(M_act) * sim
  K.preference ← K.preference + varPref            // renforcement
  totalPref ← totalPref + K.preference
  Si K.preference < s_o                                  // oubli
    supprimer K de la TPT
  Sinon
    varGli ← c_g * f(M_act) * sim                    // glissement
    K.archetype ← ⊕(T, T_act , 1- varGli, varGli)
  Fin
Fin

pref_act ← c_r * f(M_act) * f(simax)                //introduction
Si pref_act >= s_o
  introduction d'une nouvelle composante K dans TPT
  K.preference ← pref_act
  K.archetype ← T_act
  K.symbole ← nouveau symbole
  totalPref ← totalPref + pref_act
Fin

Pour toute composante K de la TPT                          // fusion
  trouvé ← vrai
  supprimé ← faux
  Tant que (trouvé = vrai et supprimé = faux)
    trouvé ← faux
    T ← K.archetype
    Tant que (il reste une composante J de la TPT (sauf K)
              et trouvé = faux)
      U ← J.archetype
      sim = similitude entre T et U
      Si sim < s_f                          // normalisation et fusion
        Si K.preference >= J.preference
          K.preference ← (K.preference + J.preference)/totalPref
          supprimer J de la TPT
        Sinon
          J.preference ← (K.preference + J.preference)/totalPref
          supprimer K de la TPT
          supprimé ← vrai
        Fin
        trouvé ← vrai
      Fin
    Fin
  Fin
Fin
```

où `f(M_act)` représente le résultat de $f^{(1)}_{TPT_X}\left(\overline{\tilde{M}^t_X}\right)$, dans les équations (eq. V.3.1), (eq. V.3.2), et (eq. V.3.6), tandis que `f(M_act)` représente le résultat de $f^{(2)}_{TPT_X}\left(simax^t_X\right)$, dans l'équation (eq. V.3.2).

4. IMPLEMENTATION

On peut distinguer deux phases dans l'utilisation de notre outil : la spécification d'un modèle, et la simulation. La première phase consiste à définir le modèle (structure, paramètres, fonctions), mais aussi les entrées qui vont l'alimenter lors de la simulation. Lors de la deuxième phase, la simulation est réalisée grâce au modèle, aux données précédemment définis, et à un moteur de simulation implémentant les algorithmes décrits précédemment. La simulation produit un ensemble de valeurs représentant l'évolution des variables qui composent le modèle au cours de la simulation.

La première version de notre outil a été réalisée en Java. Elle comprenait, dans un même logiciel, l'interface graphique permettant de concevoir les modèles, et le moteur permettant de réaliser les simulations. Mais, rapidement, de profonds changements ont été apportés au formalisme, ce qui s'est traduit par d'importantes modifications à apporter à notre outil, essentiellement au niveau de la représentation des données et du moteur de simulation. Ces changements permanents et essais successifs nous ont amené à laisser momentanément de côté le développement de l'interface, jusqu'à la stabilisation du formalisme et la finalisation d'une version plus définitive du moteur. Pour ces raisons, l'interface présentée ici n'est pas parfaitement adaptée à notre formalisme et au moteur de simulation qui en découle. Néanmoins, dans les grandes lignes, elle est respecte nos contraintes de modélisation. De plus, elle possède certaines propriétés, au niveau de son ergonomie, qui la rendent indispensable à la finalisation d'un outil destiné à des utilisateurs qui n'ont pas de notions en programmation informatique.

4.1. Interface

L'interface graphique (Figure B.4.1) permet de contrôler la création du modèle au fur et à mesure de sa définition par l'utilisateur, et évite ainsi d'éventuelles erreurs. Elle épargne une phase de vérification du modèle destinée par exemple à s'assurer que ce dernier ne contient pas de relations instantanées, qui engendreraient des cycles orientés dans le réseau dynamique, etc..

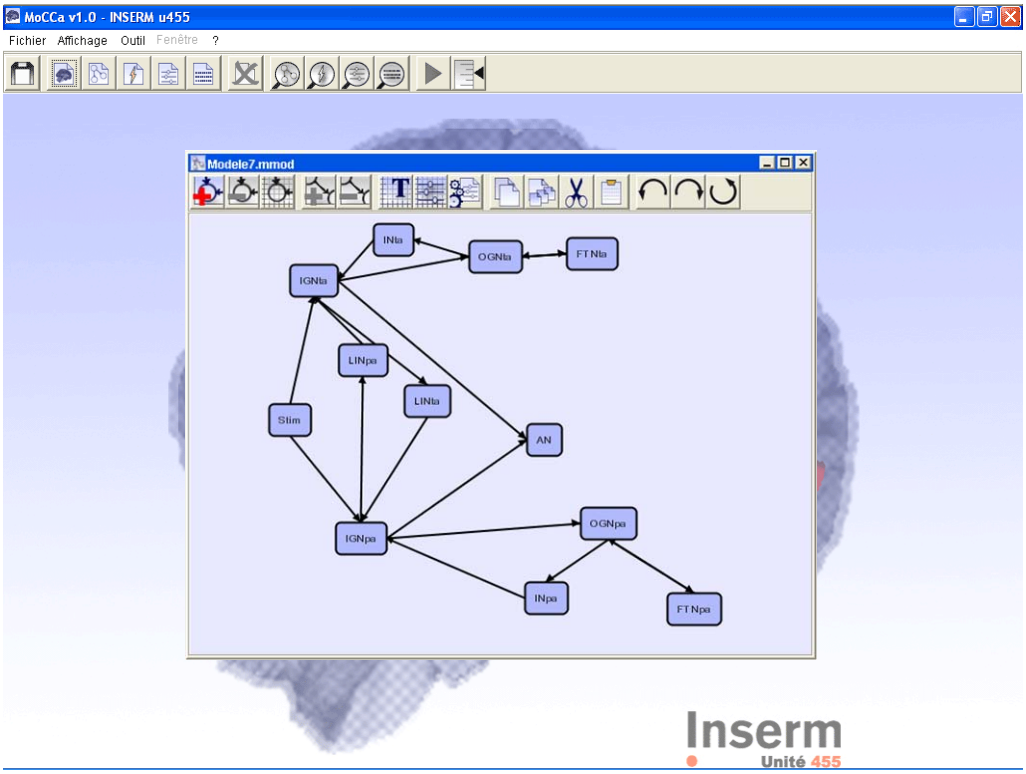

Figure B.4.1 : définition du réseau statique.

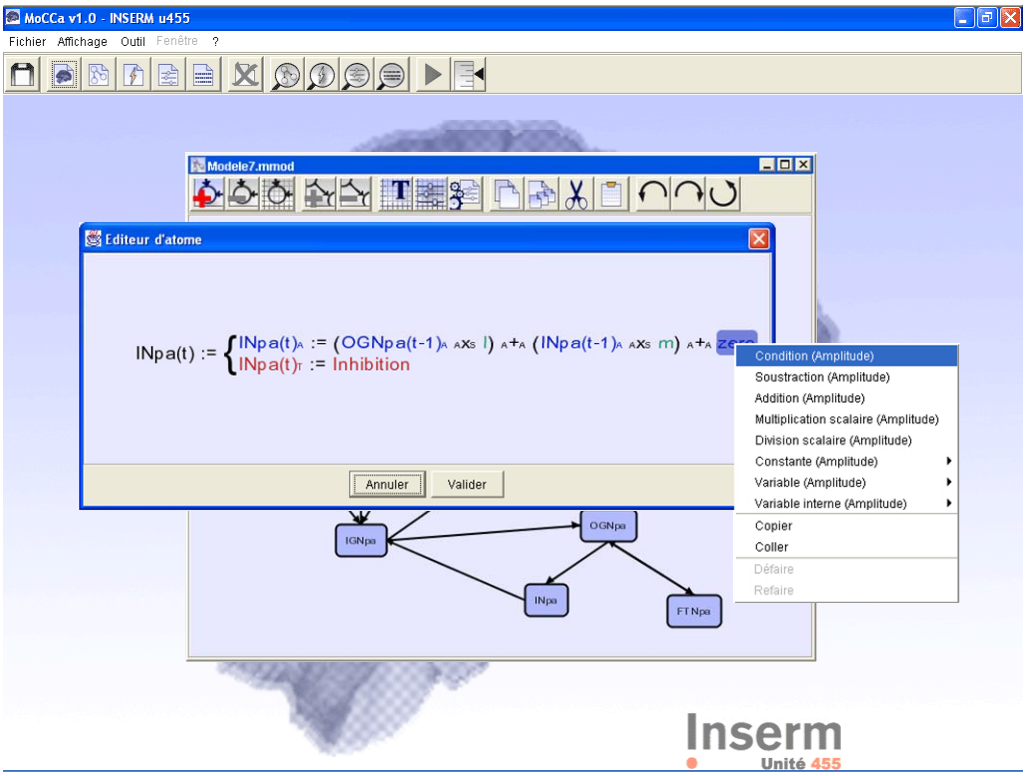

Figure B.4.2 : définition des équations associées aux noeuds.

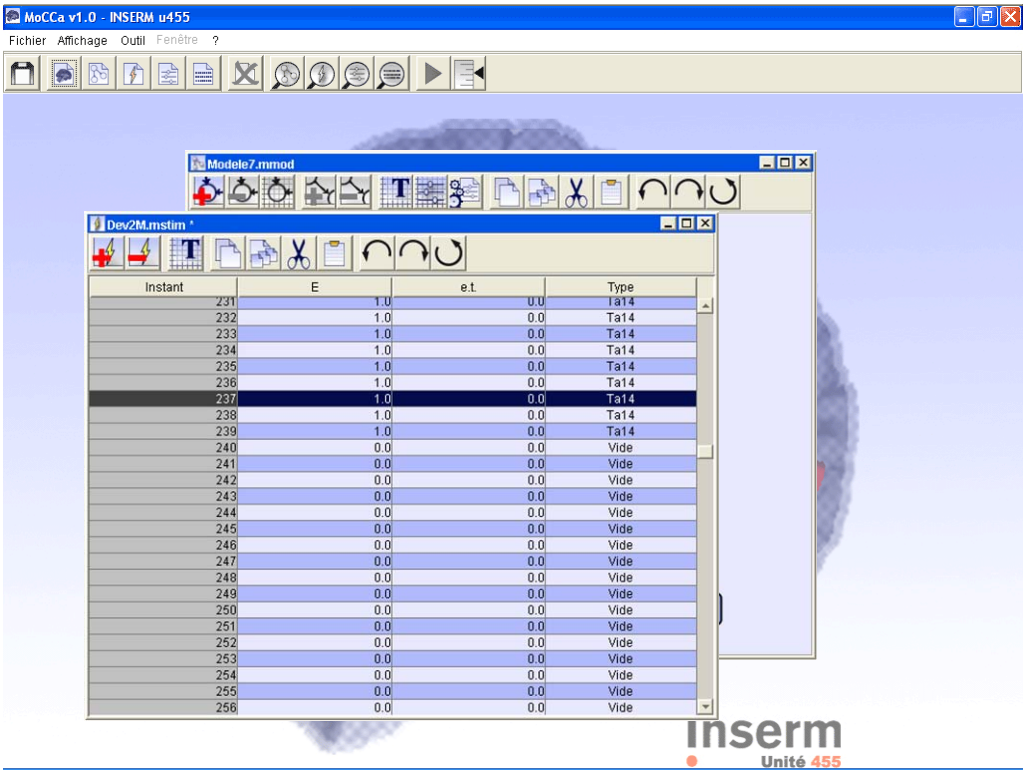

Figure B.4.3 : définition des stimuli.

Un éditeur graphique permet tout d'abord de définir la structure du réseau statique sous la forme d'un réseau orienté de nœuds (Figure B.4.1). Puis, pour chaque nœud, les différentes fonctions d'activation et d'émission doivent être exprimées. Un éditeur d'équation permet de définir ces fonctions, tout en contrôlant leur validité (Figure B.4.2). La construction se fait uniquement au moyen de menus. Ceci permet, d'une part, de ne proposer à l'utilisateur que des valeurs utilisables dans la fonction en cours de définition. Par exemple, les seules valeurs proposées par l'interface pour définir les fonctions d'activation dépendent des nœuds qui ont été connectés au nœud courrant lors de la spécification du réseau statique (i.e. on se restreint aux seules entrées de ce nœud). D'autre part, les opérateurs seuls proposés pour définit les fonctions sont ceux qui respectent nos contraintes. Par exemple, lors de la définition de la fonction permettant de calculer le type d'activation, l'opérateur de combinaison proposé (linéaire ou non-linéaire) dépend des domaines de définition des types déjà présents dans l'équation. De même, pour un opérateur déjà présent dans la fonction, l'interface ne proposera que des types adéquats. C'est lors de la définition de ces fonctions que l'on associe un délai aux relations entre les différents nœuds du réseau statique, et que l'on passe ainsi au réseau dynamique. La encore, lors de la définition d'un délai, l'interface ne

propose d'utiliser que des délais compatibles avec les délais déjà définis pour les autres nœuds, dans le but d'éviter la création d'un cycle orienté dans le réseau dynamique.

Toutes ces vérifications sont grandement facilitées par l'utilisation d'un langage objet et la représentation des fonctions définies par l'utilisateur sous la forme d'un graphe dont les nœuds sont des objets correspondant à des valeurs, des variables, des fonctions, ou des opérateurs. En effet, on peut alors définir de façon très simple les compatibilités entre opérateurs/fonctions et variables/valeurs.

Une fois le modèle défini, il est temps de préciser les valeurs du ou des stimuli qui viendront l'alimenter en information durant la simulation Figure B.4.3. Là encore, la compatibilité entre les types de stimuli et les propriété des nœuds qui les reçoivent en entrée est automatiquement vérifiée. Il est également possible de définir plusieurs paramétrages différents pour un même modèle, dans le but d'enchainer plusieurs simulations (Figure B.4.4). Cela permet, en comparant les résultats de ces simulations aux données réelles, d'estimer les valeurs de certains paramètres par essais successifs. Pour cela, on joue sur les valeurs de ces paramètres, en retenant les valeurs donnant les meilleurs résultats.

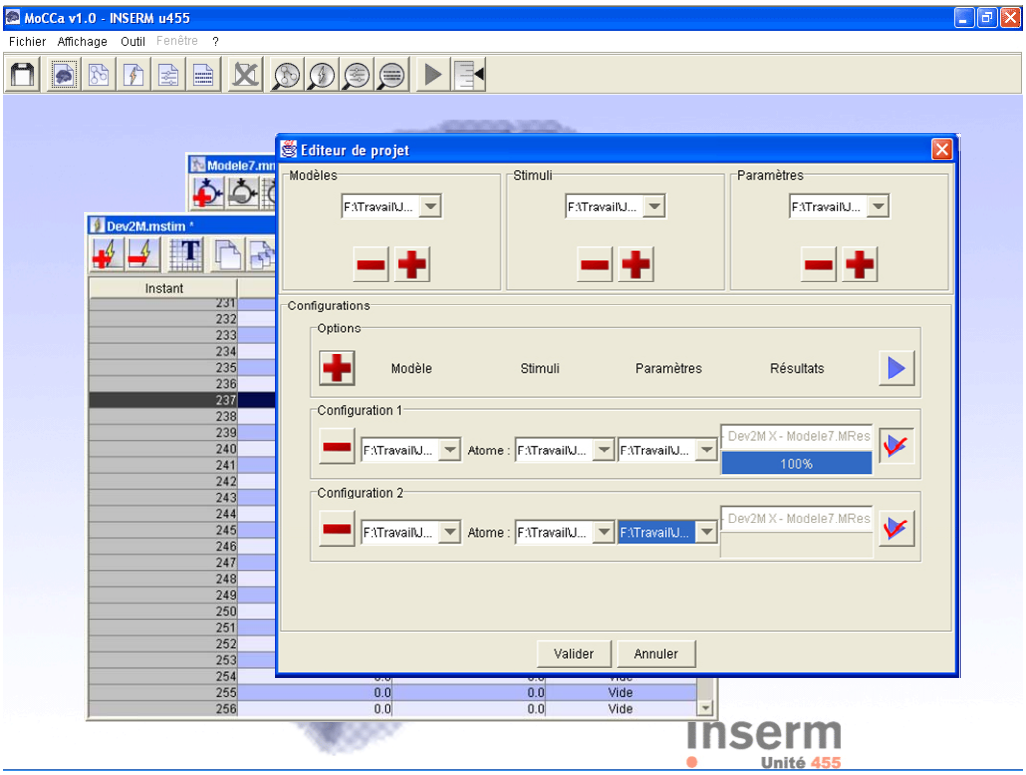

Figure B.4.4 : mise en œuvre de la simulation.

Une fois les simulations terminées, les résultats sont présentés sous forme de tableaux (Figure B.4.5), qui peuvent être sauvegardés sous des formats compatibles avec des tableurs, afin de permettre des traitements ultérieurs de nos données.

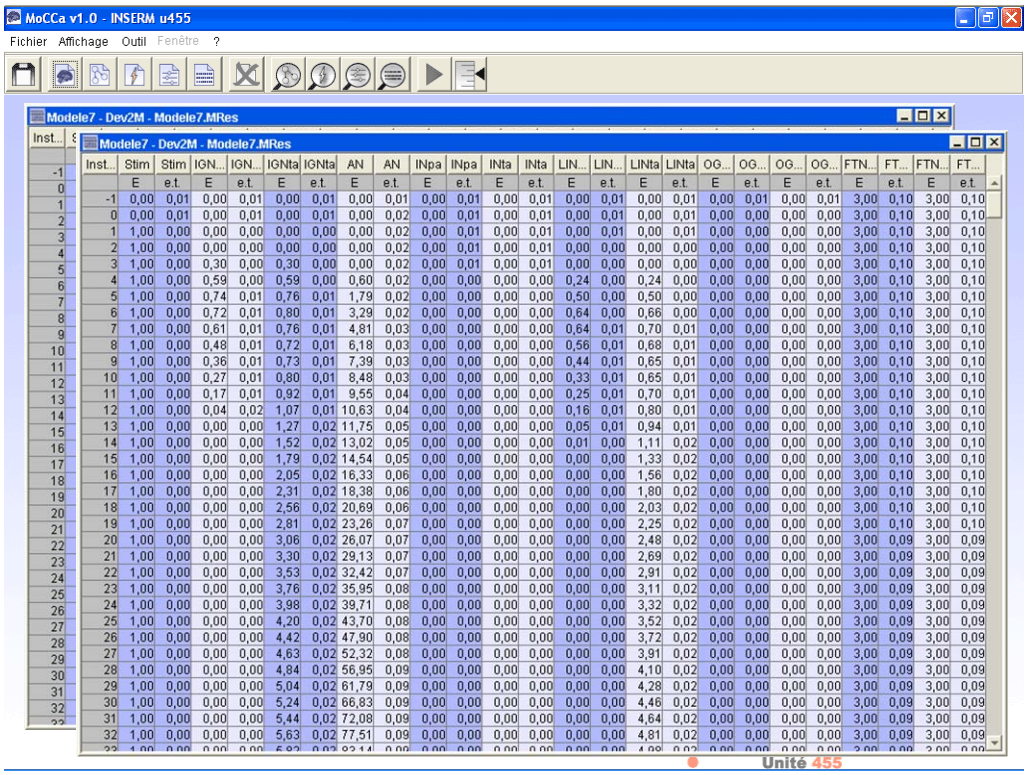

Figure B.4.5 : résultats de la simulation.

4.2. Moteur

Après avoir développé une première version complète de notre moteur en Java, nous avons décidé de séparer moteur et interface. De plus, nous avons décomposé le processus de simulation en deux : un côté type et un côté magnitude. Le but était de profiter des résultats mis à notre disposition par la communauté spécialisée dans les algorithmes de filtrage non-linéaires.

En effet, plusieurs boîtes à outils Matlab dédiées aux filtres de Kalman non-linéaires ont fait leur apparition, notamment Kalmtool (*Kal*man *Tool*box) [Nørgaard *et al.* '00] et ReBel (*Re*cursive *B*ayesian *e*stimation *l*ibrary) [van der Merwe & Wan '03b]. La première est une implémentation des filtres à différences finies d'ordre 1 et d'ordre 2. La dernière, qui est la plus récente, offre l'avantage

de proposer un vaste choix d'algorithmes d'inférence, applicables à un modèle sans avoir à changer l'implémentation de celui-ci (filtre unscented, à différences finies, à particules, etc.).

Nous avons donc choisi d'utiliser ces outils pour traiter la partie magnitude de l'information. La propagation des types et l'apprentissage sont effectués par l'implémentation des algorithmes décrits précédemment. Par souci de compatibilité avec les filtres non-linéaires, cette implémentation a été effectuée sous la forme de scripts Matlab. Par la suite, nous projetons d'adapter en langage Java un des algorithmes de filtrage non-linéaire, afin de profiter de l'interface déjà réalisée (en Java). Ceci nécessitera également une modification de l'interface, afin de l'adapter aux dernières évolutions de notre formalisme. Nous comptons également offrir la possibilité de concevoir un modèle grâce à notre interface, puis de l'exporter sous une forme compatible avec les boîtes à outils Matlab citées auparavant. L'intérêt de cette démarche est de profiter de l'évolution des algorithmes de filtrage non-linéaire.

Vincent Labatut

LARGE-SCALE PROBABILISTIC CAUSAL NETWORKS: A NEW FORMALISM FOR CEREBRAL INFORMATION PROCESSING MODELING

Ph.D. Advisor: Josette Pastor
INSERM U455

- Abstract -

The understanding and the prediction of the clinical outcomes of focal or degenerative cerebral lesions, as well as the assessment of rehabilitation procedures, necessitate knowing the cerebral substratum of cognitive or sensorimotor functions. This is achieved by activation studies, where subjects are asked to perform a specific task while data of their brain functioning are obtained through functional neuroimaging techniques. Such studies, as well as animal experiments, have shown that sensorimotor or cognitive functions are the offspring of the activity of large-scale networks of anatomically connected cerebral regions. However, no one-to-one correspondence between activated networks and functions can be found. Furthermore, apparently conflicting activation data can only be explained by understanding how the activation of large-scale networks derives from cerebral information processing mechanisms.

At this level, cerebral mechanisms are the synthesis of more basic neurobiological, neurophysiological or neuropsychological processes. They can only be approached with the help of explicit computational models, based on the knowledge of more basic processes.

In this work, we aim at defining a new and flexible formalism allowing building such models, for a better interpretation of cerebral functional images. It is based on dynamic Bayesian networks, a causal, probabilistic and graphical modelling approach. The brain is viewed as a network of anatomically connected functional areas. Each area is a processing unit which is represented by a set of nodes in the dynamic Bayesian network. In the formalism, the information processed by one node is the abstraction of the cerebral activity in the corresponding neuronal population. We use a two-sided information representation dedicated to the representation of integrated cerebral information. It is made of a numerical value corresponding to the activation level, and a symbolic value representing the pattern of activated neurons.

This work is a part of a project which is dedicated to the definition of an original causal approach for the modelling of cerebral information processing in large-scale networks and the interpretation of neuroimaging data. Causality allows us to express explicitly our hypotheses on cerebral mechanisms. Our contribution can be described in two points. In an artificial intelligence perspective, the use of labelled random variables in dynamic Bayesian networks allows us to define original non-supervised learning mechanisms. From the computational neuroscience viewpoint, we propose a new causal formalism, which is more adapted than formal neural networks to model cerebral processing at the global level of cerebral areas networks. Furthermore, our formalism is more biologically plausible than other causal approach like causal qualitative networks.

Keywords: computational neuroscience – dynamic Bayesian networks – Kalman filter – functional neuroimaging – large-scale cerebral network – non-supervised adaptive learning.

INSERM U455
Pavillon Riser, CHU Purpan, F-31059 Toulouse Cedex 3

RESEAUX CAUSAUX PROBABILISTES A GRANDE ECHELLE :
UN NOUVEAU FORMALISME POUR LA MODELISATION DU TRAITEMENT DE L'INFORMATION CEREBRALE

Thèse pour obtenir le grade de docteur en informatique de l'université Toulouse III – Paul Sabatier

Présentée et soutenue le 18 décembre 2003 par Vincent Labatut

Directrice de recherche : Josette Pastor

RESUME : La compréhension du fonctionnement cérébral passe par l'étude des relations entre les structures cérébrales et les fonctions cognitives qu'elles implémentent. Les études en activation, qui permettent d'obtenir, grâce aux techniques de neuroimagerie fonctionnelle, des données sur l'activité cérébrale pendant l'accomplissement d'une tâche cognitive, visent à étudier ces liens. Ces études, ainsi que de nombreux travaux chez l'animal, suggèrent que le support neurologique des fonctions cognitives est constitué de réseaux à grande échelle d'aires corticales et de régions sous-corticales interconnectées. Cependant, la mise en correspondance simple entre réseaux activés et tâche accomplie est insuffisante pour comprendre comment l'activation découle du traitement de l'information par le cerveau. De plus, le traitement cérébral est très complexe, et les mesures fournies par la neuroimagerie sont incomplètes, indirectes, et de natures différentes, ce qui complique grandement l'interprétation des données obtenues. Un outil de modélisation explicite des mécanismes de traitement et de propagation de l'information cérébrale dans les réseaux à grande échelle est nécessaire pour palier ces défauts et permettre l'interprétation des mesures de l'activité cérébrale en termes de traitement de l'information.

Nous proposons ici un formalisme original répondant à ces objectifs et aux contraintes imposées par le système à modéliser, le cerveau. Il est basé sur une approche graphique causale et probabiliste, les réseaux bayésiens dynamiques, et sur une représentation duale de l'information. Nous considérons le cerveau comme un ensemble de régions fonctionnelles anatomiquement interconnectées, chaque région étant un centre de traitement de l'information qui peut être modélisé par un noeud du réseau bayésien. L'information manipulée dans le formalisme au niveau d'un noeud est l'abstraction du signal généré par l'activité de la population neuronale correspondante. Ceci nous conduit à représenter l'information cérébrale sous la forme d'un couple numérique/symbolique, permettant de tenir compte respectivement du niveau d'activation et de la configuration des neurones activés.

Ce travail se situe dans le prolongement d'un projet visant à développer une approche causale originale pour la modélisation du traitement de l'information dans des réseaux cérébraux à grande échelle et l'interprétation des données de neuroimagerie. L'aspect causal permet d'exprimer explicitement des hypothèses sur le fonctionnement cérébral. Notre contribution est double. Au niveau de l'intelligence artificielle, l'utilisation de variables aléatoires labellisées dans des réseaux bayésiens dynamiques nous permet de définir des mécanismes d'apprentissage non-supervisés originaux. Sur le plan des neurosciences computationnelles, nous proposons un nouveau formalisme causal, plus adapté à la représentation du fonctionnement cérébral au niveau des réseaux d'aires que les réseaux de neurones formels, et présentant plus de plausibilité biologique que les autres approches causales, en particulier les réseaux causaux qualitatifs.

MOTS CLES : neurosciences computationnelles – réseau bayésien dynamique – filtre de Kalman – neuroimagerie fonctionnelle – réseau cérébral à grande échelle – apprentissage adaptatif non-supervisé.