



**HAL**  
open science

# Aide au diagnostic et à l'apprentissage dans un domaine médical incertain, incomplet et évolutif: étude des méthodes existantes et proposition d'une méthodologie nouvelle

Vincent Rialle

## ► To cite this version:

Vincent Rialle. Aide au diagnostic et à l'apprentissage dans un domaine médical incertain, incomplet et évolutif: étude des méthodes existantes et proposition d'une méthodologie nouvelle. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1987. Français. NNT: . tel-00325679

**HAL Id: tel-00325679**

**<https://theses.hal.science/tel-00325679>**

Submitted on 30 Sep 2008

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

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

**UNIVERSITE SCIENTIFIQUE TECHNOLOGIQUE ET MEDICALE DE GRENOBLE**

**THESE**

Présentée par

**Vincent RIALLE**

Pour obtenir le titre  
de

Docteur en GENIE BIOLOGIQUE et MEDICAL

**AIDE AU DIAGNOSTIC ET A L'APPRENTISSAGE  
DANS UN DOMAINE MEDICAL INCERTAIN, INCOMPLET ET EVOLUTIF  
ETUDE DES METHODES EXISTANTES ET PROPOSITION  
D'UNE METHODOLOGIE NOUVELLE**

Soutenue le 20 Février 1987

**Membres du Jury**

Président :	Pr. J. DEMONGEOT
Examineurs :	Pr. J. P. LAURENT
	Pr. J. R. SCHERRER
	Pr. M. FIESCHI
	Pr. Y. CHIARAMELLA
	Dr. A. JOANNARD



**UNIVERSITE SCIENTIFIQUE TECHNOLOGIQUE ET MEDICALE DE GRENOBLE**

**THESE**

Présentée par

**Vincent RIALLE**

Pour obtenir le titre  
de

Docteur en GENIE BIOLOGIQUE et MEDICAL

**AIDE AU DIAGNOSTIC ET A L'APPRENTISSAGE  
DANS UN DOMAINE MEDICAL INCERTAIN, INCOMPLET ET EVOLUTIF  
ETUDE DES METHODES EXISTANTES ET PROPOSITION  
D'UNE METHODOLOGIE NOUVELLE**

Soutenue le 20 Février 1987

**Membres du Jury**

Président :	Pr. J. DEMONGEOT
Examineurs :	Pr. J. P. LAURENT
	Pr. J. R. SCHERRER
	Pr. M. FIESCHI
	Pr. Y. CHIARAMELLA
	Dr. A. JOANNARD



M E D E C I N E

PROFESSEURS CLASSE EXCEPTIONNELLE ET 1ère CLASSE.

MM.	AMBLARD	Pierre	Dermatologie
	AMBROISE-THOMAS	Pierre	Parasitologie
	BEAUDOING	André	Pédiatrie - Puériculture
	BEZES	Henri	Orthopédie - Traumatologie
	BONNET	Jean-Louis	Ophthalmologie
	BOUCHET	Yves	Anatomie Chirurgie Générale et Digestive
	BUTEL	Jean	Orthopédie - Traumatologie
	CHAMBAZ	Edmond	Biochimie
	CHAMPETIER	Jean	Anatomie Topographique et Appliquée
	CHARACHON	Robert	O. R. L.
	COLOMB	Maurice	Immunologie
	COUDERC	Pierre	Anatomie Pathologique
	DELORMAS	Pierre	Pneumophtisiologie
	DENIS	Bernard	Cardiologie
	GAVEND	Michel	Pharmacologie
	HOLLARD	Daniel	Hématologie
	LATREILLE	René	Chirurgie Thoracique et Cardiovasculaire
	LE NOC	Pierre	Bactériologie - Virologie
	MALINAS	Yves	Gynécologie et Obstétrique
	MALLION	Jean-Michel	Médecine du Travail
	MICOU	Max	Clinique Médicale et Maladies Infectieuses
	MOURIQUAND	Claude	Histologie
	PARAMELLE	Bernard	Pneumologie
	PERRET	Jean	Neurologie
	RACHAIL	Michel	Hépto-Gastro-Entérologie
	DE ROUGEMONT	Jacques	Neurochirurgie
	SARRAZIN	Roger	Clinique Chirurgicale
	STIEGLITZ	Paul	Anesthésiologie
	TANCHE	Maurice	Physiologie
	VERAIN	André	Biophysique
	VIGNAIS	Pierre	Biochimie

PROFESSEURS 2ème CLASSE.

MM.	BACHELOT	Yvan	Endocrinologie
	BARGE	Michel	Neurochirurgie
	BENABID	Alim, Louis	Biophysique
	BENSA	Jean-Claude	Immunologie
	BERNARD	Pierre	Gynécologie - Obstétrique
	BESSARD	Germain	Pharmacologie
	BOLLA	Michel	Radiothérapie
	BOST	Michel	Pédiatrie
	BOUCHARLAT	Jacques	Psychiatrie Adultes
	BRAMBILLA	Christian	Pneumologie
	CHIROSSEL	Jean-Paul	Anatomie - Neurochirurgie
	COMET	Michel	Biophysique
	CONTAMIN	Charles	Chirurgie Thoracique et Cardiovasculaire
	CORDONNIER	Daniel	Néphrologie
	COULOMB	Max	Radiologie
	CROUZET	Guy	Radiologie
	DEBRU	Jean-Luc	Médecine Interne et Toxicologie
	DEMONGEOT	Jacques	Biostatistiques et Informatique Médicale
	DUPRE	Alain	Chirurgie Générale
	DYON	J. François	Chirurgie Infantile
MME	ETERRADOSSI	Jacqueline	Physiologie
MM.	FAURE	Claude	Anatomie et Organogénèse
	FAURE	Gilbert	Urologie
	FOURNET	Jacques	Hépto-Gastro-Entérologie
	FRANCO	Alain	Médecine Interne
	GIRARDET	Pierre	Anesthésiologie
	GUIDICELLI	Henri	Chirurgie Générale et Vasculaire
	GUIGNIER	Michel	Thérapeutique et Réanimation Médicale
	HADJIAN	Arthur	Biochimie
	HALIMI	Serge	Endocrinologie et Maladies Métaboliques
	HOSTEIN	Jean	Hépto-Gastro-Entérologie
	HUGONOT	Robert	Médecine Interne
	JALBERT	Pierre	Histologie - Cytogénétique
	JUNIEN-LAVILLAUIROY	Claude	O. R. L.
	KOLODIE	Lucien	Hématologie Biologique
	LETOUBLON	Christian	Chirurgie Générale
	MACHECOURT	Jacques	Cardiologie et Maladies Vasculaires
	MAGNIN	Robert	Hygiène
	MASSOT	Christian	Médecine Interne
	MOUILLON	Michel	Ophtalmologie
	PELLAT	Jacques	Neurologie
	PHELIP	Xavier	Rhumatologie
	RACINET	Claude	Gynécologie-Obstétrique
	RAMBAUD	Pierre	Pédiatrie
	RAPHAEL	Bernard	Stomatologie
	SCHAERER	René	Cancérologie
	SEIGNEURIN	Jean-Marie	Bactériologie - Virologie
	SELE	Bernard	Cytogénétique
	SOTTO	J. Jacques	Hématologie
	STOEBNER	Pierre	Anatomie Pathologique
	VROUSOS	Constantin	Radiothérapie

## REMERCIEMENTS

Mes remerciements s'adressent tout d'abord au Professeur Jacques DEMONGEOT, de la Faculté de Médecine de Grenoble, pour son rôle d'inspiration d'un aspect important de cette recherche et pour sa collaboration et son soutien constants de mon travail. Il m'a transmis, au cours de cette collaboration, une part de son savoir-faire et de sa méthode de recherche. Je lui suis en cela extrêmement redevable.

Je le remercie également de l'honneur qu'il me fait en acceptant la présidence du jury.

Je remercie le Professeur Jean-Pierre LAURENT, du Laboratoire d'Informatique Appliquée -Université de Savoie-, d'avoir accepté d'être mon Directeur de Recherche pour un sujet dont l'Intelligence Artificielle, bien que partie intégrante, n'était pas l'unique objectif. Je le remercie pour les conseils avisés qu'il a su me donner durant ces trois années, et particulièrement dans les caps difficiles de cette recherche.



Je remercie particulièrement le Professeur Jean-Raoul SCHERRER, responsable de la Division d'Informatique de l'HSpital Cantonal Universitaire de Genève, haut lieu des applications médicales de l'informatique, d'avoir accepté de juger ce travail.

Je remercie le Professeur Marius FIESCHI d'avoir accepté d'être rapporteur de ce travail et de m'avoir fait l'honneur d'être membre du jury. Je le remercie également d'avoir été depuis longtemps pour moi, à travers sa propre thèse et son dynamisme en matière d'Intelligence Artificielle et Médecine, un initiateur et un modèle dans ce champ de recherche.

Le Professeur CHIARAMELLA, de l'Institut de Mathématiques Appliquées de Grenoble, a bien voulu être rapporteur de ce travail et me fait l'honneur d'être membre du jury; pour cela et pour les conseils qu'il m'a donnés concernant les développements à moyen terme de la Base de Données, je lui adresse mes plus vifs remerciements.

Je remercie très vivement le Docteur Alain JOANNARD, de la Clinique Médicale Infantile - CHU de Grenoble -, pour sa disponibilité envers ce travail ainsi que pour la confiance qu'il m'a témoignée durant cette longue marche de trois années pendant laquelle nous avons élaboré la première partie de la Base de Données. Je le remercie également d'avoir accepté d'être membre du jury.



Je suis très reconnaissant à Claudine ROBERT, assistante à la Faculté des Sciences de Grenoble, pour son attention ainsi que pour les conseils judicieux qu'elle a bien voulu me donner en matière d'Analyse des Données. Je la remercie également pour la compréhension profonde des problèmes décisionnels en matière de diagnostic médical qu'elle a su me faire partager.

Je remercie vivement les Docteurs Sylvie GOMES et Nathalie BAUDINO qui ont rassemblé avec patience les éléments (syndromes, symptômes, prévalences, etc...) qui nous ont permis de créer la Base de Données.

Je remercie Michel SIMONET du CNRS pour les échanges fructueux que nous avons eus à propos des domaines de connaissance vastes et évolutifs, centre d'intérêt que nous partageons.

Je tiens à remercier tout particulièrement André EBERHARD, ingénieur à l'Institut de Mathématiques Appliquées de Grenoble, pour l'aide qu'il m'a apportée dans la mise au point des programmes ainsi que pour les nombreux logiciels qu'il a mis à ma disposition avec une constante amabilité.

J'adresse mes remerciements à l'Institut de Mathématiques Appliquées de Grenoble pour avoir mis à ma disposition sa Bibliothèque d'Analyse Numérique (B.A.N.G.).



Je remercie Francois RECHENMANN, de l'INRIA, pour les nombreuses discussions que j'ai eues avec lui, et pour ses remarques et conseils de chercheur en Intelligence Artificielle, ainsi que pour sa constante et amicale disponibilité à mon égard.

J'ai la plus grande gratitude à l'égard d'André VIAL qui m'a efficacement aidé durant tout l'été de 1985, au cours de son stage de DESS Informatique-Double Compétence, pour la mise au point des programmes de gestion des fichiers de la base et pour la saisie des milliers de données relatives aux symptômes et aux syndromes. Le travail considérable qu'il a fourni m'a permis de franchir une étape décisive dans l'élaboration de ma propre recherche. Pour sa perspicacité et sa constante bonne humeur, je lui transmets mes remerciements les plus chaleureux.

Djamel BELLAÏD - I.B.M.-, a bien voulu répondre à mes questions et m'a apporté un soutien logiciel, qu'il en soit vivement remercié.

Mesdames Odette PAPINI et Rosette BONY, et Mademoiselle Fadila MERADJI ont contribué à la construction de la base de données et ont assuré la frappe de cette thèse avec une gentillesse et une patience remarquables, je tiens à les en remercier très sincèrement.

Je tiens à remercier toutes les personnes qui de près ou de loin, dans le milieu scientifique et médical et hors de ce milieu, m'ont permis d'avancer de manière décisive dans cette recherche.

## S O M M A I R E

PLAN SYNOPTIQUE DE LA THESE	p.17
INTRODUCTION GENERALE	p.18
I. MODELES MATHEMATIQUES D'ANALYSE DECISIONNELLE EN MEDECINE - ANALYSE COMPARATIVE -	p.27
I.1 Introduction	p.27
I.2 méthodes de décision à caractère probabiliste	p.29
I.2.1 Formule de Bayes	p.29
I.2.1.1 Rappel théorique	p.29
I.2.1.2 Une application probante : le système HELP	p.32
I.2.2 Modèle mathématique de Shortliffe	p.43
I.2.3 Indicateurs d'efficacité et valeurs prédictives	p.48
I.2.4 Analyse des proximités par corrélations	p.52
I.2.4.1 Introduction	p.53
I.2.4.2 Explicitation du modèle théorique	p.53
I.3 Théorie mathématique de l'évidence	p.60
I.3.1 Introduction	p.60
I.3.2 Fonctions de croyance et règle de combinaison de DEMPSTER - application : le système GENDIAG	p.64
I.4 Théorie des sous-ensembles flous et Théorie des possibilités	p.67
I.4.1 Eléments de base	p.67
I.4.2 Inférence en logique floue	p.71
I.4.2.1 Exemple de PROTIS	p.72
I.4.2.2 Exemple de SPHINX	p.76
I.5 Bases de Données, Bases de Connaissances	p.80
I.5.1 Introduction	p.80
I.5.2 La Base A.D.M.	p.83

## Sommaire

I.5.3	Eléments de prospective	p.86
I.6	Eléments de comparaisons et de discussion	p.87
I.6.1	Discussion des méthodes proposées	p.88
I.6.1.1	A propos de la Formule de Bayes	p.88
I.6.1.2	A propos de la Logique Floue	p.91
I.6.1.3	Remarques sur les coefficients numériques	p.94
I.6.1.4	La double philosophie de système SAM	p.96
I.6.2	Vers une synthèse des méthodes proposées	p.98
II.	UNE PRODUCTION DE CONNAISSANCES EXPERTES EN MEDECINE	p.101
II.1	Pourquoi une production de connaissances expertes ?	p.101
II.1.1	Point de vue médical général	p.101
II.1.2	Cas précis de la génétique médicale	p.107
II.2	Un exemple de production de connaissances : le système SEEK	p.110
II.3	Spécificité du domaine de la Génétique Médicale	p.114
II.4	Principe d'utilisation de l'Analyse des Données	p.116
II.4.1	Nature et origine des connaissances - pourquoi l'Analyse des Données ?	p.117
II.4.2	Exposé de principes fondamentaux - l'A.C.P. canonique	p.119
II.4.3	Conditions d'utilisation de l'Analyse des Données	p.124
II.4.4	Démarche formelle d'analyse	p.125
II.5	Modélisation du domaine	p.127
II.5.1	Eléments de base : les symptômes	p.127
II.5.1.1	Le concept de symptôme	p.127
II.5.1.2	Niveaux d'observation - partition ordonnée de l'ensemble des symptômes	p.128
II.5.1.3	Indice fréquentiel d'apparition	p.129
II.5.1.4	Arborescence des symptômes	p.130
II.5.2	Premier niveau de structuration :	

## Sommaire

les entités symptomatiques	p.130
II.5.2.1 Définition	p.130
II.5.2.2 Axe orienté associé à une entité symptomatique	p.132
II.5.2.3 Niveaux d'observation	p.133
II.5.2.4 Arborescence des entités symptomatiques	p.133
II.5.3 Second niveau de structuration : les super-entités symptomatiques de type SCORE	p.135
II.5.3.1 Définition	p.135
II.5.3.2 Finalité de cette structuration : pourquoi des SCORES ?	p.135
II.5.3.3 Choix des super-entités de type score	p.137
II.5.3.4 Problèmes inhérents au choix d'un score et solution retenue	p.140
II.5.3.5 La Roc-Analyse	p.142
II.5.4 Buts diagnostiques : les syndromes	p.147
II.5.4.1 Le concept de syndrome	p.147
II.5.4.2 Double représentation d'un syndrome	p.148
II.5.4.2.1 Représentation sémiologique	p.148
II.5.4.2.2 Représentation vectorielle	p.149
II.5.4.3 Résolution des informations incomplètes, Coefficient de Précision Descriptive d'un syndrome	p.149
II.5.4.4 Passage de la représentation sémiologique à la représentation vectorielle d'un syndrome	p.151
II.5.5 Espaces de travail	p.153
II.5.5.1 Définitions : espace complet, espaces conjecturaux	p.153
II.5.5.2 Construction des espaces de travail	p.153
II.6 Pré-étude statique du domaine	p.155
II.7 Production dynamique de Connaissances Nouvelles et Apprentissage	p.164
II.7.1 Postulat de base	p.164

## Sommaire

II.7.2	Cycle Théorique	p.165
II.7.3	Méthodologie Prospective	p.168
III.	SYSTEME HYBRIDE PROCEDURAL-DECLARATIF D'APPRENTISSAGE ET D'AIDE AU DIAGNOSTIC : PROJET GENESYS	p.179
III.1	Principe et schéma général du système	p.179
III.2	Base de Données Dynamique associée au système	p.181
III.2.1	Définition du modèle relationnel - formalisme Z0 -	p.181
III.2.2	Mise en oeuvre	p.185
III.2.3	Structuration de la base	p.186
III.2.3.1	Schéma fonctionnel - génération en cascade des fichiers F1 à F8	p.186
III.2.3.2	Fichiers relatifs aux symptômes	p.188
III.2.3.2.1	Informations de longueur fixe (F1)	p.189
III.2.3.2.2	Informations de longueur variable (F2)	p.189
III.2.3.2.3	Structure des listes de syndromes	p.190
III.2.3.2.4	Codage des cellules de syndromes	p.193
III.2.3.3	Fichiers relatifs aux syndromes	p.195
III.2.3.3.1	Informations de longueur fixe (F3)	p.195
III.2.3.3.2	Informations de longueur variable - listes de symptômes (F4) et de couples (es,v) (F5)	p.196
III.2.3.3.3	Structure des listes de symptômes	p.197
III.2.3.3.4	Codage des cellules de symptômes	p.198
III.2.3.3.5	Structure des listes de couples (es,v)	p.198
III.2.3.4	Fichier relatif aux entités symptomatiques (F6)	p.199
III.2.3.5	Fichier relatif aux scores (F7)	p.200
III.2.3.6	Fichier relatif aux axes de l'Analyse des Données (F8)	p.201
III.2.4	Gestion de la base - méthodes d'accès aux différents fichiers	p.202

III.2.4.1 Fichiers Dbase III	p.202
III.2.4.2 Fichiers pascal	p.203
III.3 Module d'Analyse Numérique	p.204
IV. CONCLUSION	p.207
BIBLIOGRAPHIE	p.214

## ANNEXES

1. Liste des axes.
2. Liste fragmentaire des entités symptomatiques.
3. Structures des fichiers DBASE III.
4. Programme CRES.PRG de création du fichier des entités symptomatiques.
5. Programme CREAM.PRG de génération et régénération de la Base de Données.
6. Programme GFSYND de gestion du fichier FSYNDROM.
7. Programme GENFSYMP de génération du fichier FSYPTOM.
8. Programme GENFES de génération du fichier FES.
9. Liste de valeurs propres et pourcentages simples et cumulés d'inertie (niveau 1).
10. Liste des premiers vecteurs propres de niveau 1.
11. Plans de projection de niveau 1.
12. Liste des syndromes.
13. Programme GFICHACP de création de la matrice (syndrome x axes).
14. Programme J de calcul de la matrice de variances-covariances et de recherche des valeurs et vecteurs propres.
15. Programme PROJEC de représentation graphique.
16. Programme LDISC de recherche des axes pertinents à partir des vecteurs propres.
17. Ensembles dichotomiques.
18. Liste des symptômes portés par les axes sélectionnés au niveau 1.
19. Valeurs et vecteurs propres de niveau 2 - exemple traité.
20. Super-entité symptomatique sélectionnée au niveau 2, liste des symptômes.

## PLAN SYNOPTIQUE DE LA THESE

Cette thèse poursuit un double objectif :

- Dans un premier temps: recenser, analyser et comparer les diverses théories et techniques de modélisation du Raisonnement Incertain et de l'Aide à la Décision, plus particulièrement dans le domaine médical; en ayant soin de préciser à chaque fois les structures mathématiques ainsi que l'orientation philosophique sous-jacentes.

- Dans un deuxième temps: s'attacher à un domaine particulier d'aide au diagnostic, celui de la Génétique Médicale, caractérisé par son étendue, son incertitude et son extrême évolutivité, et présenter une solution située au carrefour d'un certain nombre de grands domaines de l'informatique et des mathématiques. L'originalité du système tient principalement à l'utilisation de techniques d'Analyse des Données comme moteur d'apprentissage d'une connaissance destinée à un système expert.

Ce double objectif se reflète au niveau du plan général de la thèse:

- après une introduction générale,
- le chapitre I correspond à l'étude des modèles mathématiques d'Aide au Raisonnement et à la Décision Incertaine en Médecine,
- le chapitre II est consacré tout d'abord à l'exposé de la problématique de la production d'une connaissance experte dans le domaine considéré, ensuite à l'exposé du modèle théorique proposé, en relation avec les divers aspects du domaine,
- le chapitre III, plus technique, décrit la structure du logiciel correspondant à la mise en application du modèle théorique,
- le chapitre IV est consacré aux conclusions,
- une vingtaine d'annexes termine ce travail.

## INTRODUCTION GENERALE

"(...) La décision médicale est aujourd'hui en partie rationnelle. En partie seulement. De grands domaines d'ignorance partielle ou totale persistent ; certains désordres ne sont ni définis ni mesurés. C'est cette situation actuelle, cette alliance à l'origine de la décision médicale du rationnel et de l'irrationnel qui fait tout l'intérêt des questions que pose aujourd'hui cette décision médicale."

Jean BERNARD  
de l'Académie Française  
{BERNARD 77}

Le travail que nous présentons dans cette thèse repose sur les 3 axes fondamentaux suivants :

- 1 - un thème
- 2 - un champ
- 3 - une méthode

Le **thème** de notre recherche est l'innovation en matière d'aide au diagnostic à travers l'apport d'une méthodologie nouvelle et à la suite d'une étude approfondie et critique des diverses méthodes employées jusqu'à maintenant.

Le **champ** de cette recherche est en premier lieu la Génétique Médicale ; par extension, la méthodologie proposée pourra être adaptée à des champs possédant des caractéristiques similaires à ce

## Introduction

domaine particulier de la médecine.

La **méthodologie** présentée est inspirée à la fois par le courant de l'Analyse des Données et par le courant de l'Intelligence Artificielle et des Systèmes Experts.

Nous développons maintenant la problématique générale de notre travail, en relation avec les trois axes qui viennent d'être indiqués.

L'art de la décision médicale informatisée est au centre d'une intense réflexion, tant de la part de médecins que d'informaticiens ({POPLE 75 a}, {POPLE 75 b}, {PAUKER 76 a}, {PAUKER 76 b}, {SHAPIRO 76}, {ELSTEIN 78}, {KASSIRER 78}, etc.). L'innovation en matière d'aide informatique au diagnostic médical est une expérience à la fois stimulante, non dépourvue de risques, et toujours laborieuse. L'aide au diagnostic n'est pas le diagnostic lui-même, avec les connotations de gravité et de responsabilité qui s'y rattachent. Elle n'en constitue pas moins une étape parfois essentielle vers l'établissement de ce diagnostic. L'aide informatique au diagnostic est stimulante en ce qu'elle participe d'une part à l'étude du vivant, d'autre part aux progrès rapides de l'informatique, de l'Intelligence Artificielle (I.A.) et des théories sur l'aide à la décision.

L'innovation n'en est pas moins risquée. Ses risques se répartissent également sur les deux plans évoqués précédemment: l'étude du vivant et la technique informatique au sens large. En ce qui concerne le premier plan, le risque évident est lié à l'acte médical du diagnostic, du pronostic ou de la thérapeutique. Au plan de la technique informatique, le risque est multiple: heurter des

## Introduction

conceptions établies de l'aide à la décision, proposer des méthodes trop coûteuses eu égard aux capacités actuelles d'informatisation.

Le risque est également celui de l'échec possible d'une collaboration entre deux spécialistes d'orientation professionnelle très différente : le médecin et l'informaticien.

Enfin, l'innovation est chose laborieuse: le génie médical aussi bien qu'informatique exige une élaboration parfois en contradiction avec la rapidité des progrès réalisés globalement dans chacun de ces deux domaines.

Le travail que nous présentons dans cette thèse semble particulièrement représentatif des points évoqués ci-dessus.

Il se situe en effet au carrefour d'idées provenant de domaines de recherche aussi divers que ceux de la Médecine, de l'Intelligence Artificielle et de Systèmes Experts, de l'Apprentissage, de la Modélisation Mathématique du raisonnement incertain, de l'Analyse Multifactorielle des Données et enfin, de la Technique des Bases de Données.

Chaque branche de ce carrefour nécessite quelques phrases de présentation.

a - Dans le domaine tout d'abord médical, le travail que nous présentons tente d'apporter une réponse fondée sur une expérimentation réelle, à une demande d'aide au diagnostic dans le champ de la Génétique Médicale. A l'instar de la plupart des champs de la médecine, la Génétique Médicale a reçu nombre d'applications d'aide au diagnostic ({REALE 68}, {AYME 81}, etc.) dont certaines, nous le verrons au chapitre I, sont appliquées de façon régulière. On

## Introduction

notera, à ce propos, que la thèse que nous présentons est très liée au champ médical. Elle procède en effet d'une démarche pragmatique dont le point de départ est cette question: comment aider et améliorer le diagnostic des maladies congénitales ?

A partir de cette question a débuté un processus de recherche dont le seul but était de fournir les solutions les plus adaptées, puisées dans la diversité des directions actuelles de l'informatique, sans **a priori** majeur ou restrictif pour telle ou telle branche de la recherche. D'où une diversité de contributions, tant de la part de l'I.A. que de l'Analyse des Données, de la Technique des Bases de Données, etc. Cette diversité de contributions n'est pas le fruit d'une décision délibérée, mais bien celui du constat qu'aucun des domaines cités n'était en mesure, à lui seul, de répondre entièrement au problème du diagnostic en Génétique Médicale ni d'assumer entièrement les caractéristiques de ce domaine.

L'origine médicale, autant que la méthodologie employée, confèrent donc à ces travaux le caractère essentiel d'informatique appliquée à la médecine.

Le choix de la Génétique Médicale comme terrain de travail n'est pas neutre, il tient avant tout au caractère singulier de ce domaine au regard des méthodes d'aide au raisonnement. Néanmoins, si ce domaine, par ses caractéristiques particulières semble être un champ médical privilégié d'application des idées qui vont être exposées, il n'est pas unique dans l'ensemble des domaines d'investigation scientifique: la géologie, la botanique, par exemple, semblent remplir les conditions d'application des méthodes proposées.

b - En ce qui concerne l'Intelligence Artificielle , le travail présente un certain nombre de méthodes, impliquées dans l'organisation générale du système, étrangères à l'I.A. entendue au sens strict.

Ce travail défend l'idée que le mouvement des Systèmes Experts, extrêmement riche de concepts nouveaux et d'avancées théoriques et pratiques, connaît actuellement une phase critique, celle du dépassement de l'échelle des réalisations de laboratoire pour devenir un outil entièrement intégré à la fonction médicale. Une telle évolution ne peut être accomplie qu'en élargissant des idées (telles que la programmation entièrement déclarative, la base de connaissance par règles de production, etc...), qui ne résolvent pas toujours complètement le problème de l'aide au diagnostic. Certains systèmes experts constituent une contrainte pour le médecin, en n'ayant d'application que dans des domaines restreints ou en ne permettant que des déductions d'ordre trivial dans la connaissance médicale.

Une bonne partie du chapitre I témoigne cependant du rôle de l'I.A. dans la médecine {SZOLOVITS 82}, {LAURENT 83}, et des avancées théoriques importantes qu'elle permet.

De même, il sera montré, dans les chapitres II et III, la place effective, sinon prédominante, qu'occupe l'Intelligence Artificielle, dans l'orchestration générale des diverses parties du système proposé. L'I.A. permet en effet la création d'un "méta-programme" chargé d'assurer intelligemment le contrôle et l'enchaînement des divers processus constituant le système.

La notion de Système Expert intervient au niveau de la Base de

## Introduction

Connaissances obtenue par apprentissage, ainsi que du Moteur d'Exploitation de ces connaissances.

c - La notion d'apprentissage intervient au niveau de l'acquisition des connaissances nouvelles. Le système se comporte, en effet, de la façon suivante: en présence d'une observation de malade, soit il sait diagnostiquer le syndrome et il fait ce diagnostic, soit il apprend à le faire.

La base de connaissances du système est typiquement une **mémoire à long terme** en ce qu'elle conserve les prémisses ainsi que les conclusions des expériences passées. Lorsqu'une situation déjà vécue, c'est-à-dire lorsqu'une série d'observations pratiquées sur un malade se reproduit, la conclusion qui avait été formulée par le module d'aide à la production de connaissances nouvelles, lors de la première occurrence de cette situation, est utilisée de nouveau telle quelle.

L'étape d'apprentissage est unique dans la "vie" d'une connaissance. la connaissance qui en résulte reste dans la base de connaissances, à moins que cette connaissance ne soit modifiée ou supprimée par l'expert.

### **Schéma d'apprentissage:**

situation --> apprentissage --> conclusion

### **schéma de répétition de situations déjà vécues:**

situation --> conclusion

On trouve donc ici une seconde analogie avec le comportement humain: toute mise en application de connaissances correspond à la répétition d'une scène ou d'une situation vécue antérieurement, l'étape créative proprement dite étant celle de l'apprentissage.

## Introduction

d - En ce qui concerne la place des mathématiques dans l'aide au raisonnement et à la décision, on notera tout d'abord qu'un long chapitre, le premier, lui est presque entièrement consacré.

Bien avant l'émergence des Systèmes Experts, existait un axe de recherche sur l'aide à la décision dont l'objet principal était de "faire parler les chiffres" de la manière la plus rigoureuse possible (CROFT 74) (FOX 77). C'est ainsi que, dès les années cinquante, l'ordinateur était utilisé pour l'aide à la décision diagnostique à partir des grands tableaux de données d'observation. Le "mouvement bayésien", grâce à ce nouvel outil, a pu se développer et donner lieu à une pépinière d'applications ainsi qu'à une réflexion approfondie sur la place du calcul numérique en matière de décision médicale (WINKLER 67), (GORRY 68), (CROFT 72).

Le calcul des probabilités, dont la lourdeur apparaissait comme un élément rédhibitoire pour son utilisation courante, a cédé le pas en matière d'aide au diagnostic à des recherches sur la formalisation du raisonnement approximatif et incertain (FIESCHI 84 a et b).

Ces travaux, encore en plein développement, enrichissent les recherches en Intelligence Artificielle.

Le dialogue entre l'I.A. et la Modélisation Mathématique devrait développer l'idée qu'il n'existe pas de frontière très nette entre ces deux approches de l'aide au raisonnement. Si le calcul numérique peut être considéré comme "inintelligent" du point de vue du système expert, il n'en demeure pas moins un outil d'aide à l'exploitation de domaines connus à travers de très nombreuses

## Introduction

observations.

Les méthodes d'analyse multifactorielle, sur une grande masse de données, ont connu, depuis une vingtaine d'années et parallèlement à celui de l'I.A., un important développement (DURBEC 76). On dispose maintenant d'outils théoriques et logiciels capables véritablement d'extraire de la connaissance à partir d'une réalité observée de multiples fois au moyen de nombreux observables.

Il en va de même pour la ROC-Analyse dont le but est l'optimisation des scores, tant au point de vue de la qualité des variables que du point de vue des seuils de discrimination. L'usage de ces scores peut être traduit en terme de règles d'inférence puisqu'ils permettent de situer avec précision un individu dans une catégorie diagnostique, pronostique ou thérapeutique prédéfinie.

La méthodologie que nous exposerons au chapitre II accorde donc une place importante à l'utilisation des méthodes numériques pour l'exploration du domaine considéré.

e - un aspect "Base de Données" fait également partie de notre problématique générale. Cet aspect, plus particulièrement développé au cours du chapitre III, est étroitement lié à l'interface système expert et analyse des données.

f - le projet GENESYS, auquel le chapitre II est entièrement consacré, constitue un programme de travail à moyen et long terme, co-géré par la Faculté de Médecine et la Clinique Médicale Infantile de Grenoble. Nous précisons ici que c'est la méthodologie de réalisation du projet GENESYS qui constitue l'objectif principal de notre travail, et non pas la réalisation complète du programme.

## Introduction

Ce projet a pour but la réalisation d'un environnement de système expert capable à la fois d'une aide au diagnostic des syndromes génétiques, et d'une capacité de production de connaissances nouvelles dans ce domaine. Le processus d'apprentissage constitue donc une dimension majeure de ce système. Au cours de ce chapitre, seront présentés le plan d'ensemble ainsi que les premiers jalons du projet. Ces premiers jalons sont constitués par un certain nombre de programmes, qui ont valeur de maquette du programme final, ainsi qu'un échantillon de base de données à partir d'une centaine de syndromes et de près de 700 symptômes.

## CHAPITRE I

### LES MODELES MATHÉMATIQUES D'ANALYSE DÉCISIONNELLE EN MÉDECINE - ANALYSE COMPARATIVE -

#### I.1 INTRODUCTION

La quantification numérique et les méthodes objectives ont, de tout temps, attiré les médecins pour étayer leurs jugements, leurs intuitions et leurs décisions en matière de diagnostic, de pronostic ou de thérapeutique. Le développement très rapide de l'informatique, et plus particulièrement des micro-ordinateurs, permet aujourd'hui d'offrir aux médecins de véritables outils technologiques d'aide à la décision. Les performances techniques des micro-processeurs ne doivent cependant pas masquer les véritables sources de connaissances en matière d'aide à la décision médicale. Ces sources se situent au plan de la recherche fondamentale en modélisation, en analyse de données, ainsi qu'en intelligence artificielle. Nombre de travaux, issus de ces domaines scientifiques contribuent, peu ou prou, à la réalisation des Systèmes Experts.

Chacun sait qu'un système expert est un logiciel destiné à réaliser automatiquement des déductions à partir de faits dictés par une situation particulière, en utilisant des "connaissances" qui lui sont fournies par un ou plusieurs experts du domaine traité: "Le système est censé posséder toutes les connaissances d'un expert du domaine concerné, y compris tout ce qui relève plutôt de son expérience, de sa pratique." (LAURENT 84, p.140). Dans la constitution d'une base de connaissances, il n'y a pas lieu, comme nous le verrons dans ce chapitre, de séparer a priori

les méthodes numériques des méthodes spécifiques de l'intelligence artificielle, dont un des terrains de prédilection est constitué par les Systèmes Experts. Qu'elle soit factuelle, déductive ou numérique, une connaissance ne devrait en aucun cas être négligée dans la constitution d'un système expert.

Le travail que nous présentons ici s'attache plus particulièrement aux aspects numériques de la connaissance. Pour un examen plus complet des diverses méthodes utilisées en représentation des connaissances, on consultera (KAYSER 84). Nous verrons, dans les pages qui suivent, quelle est la véritable place des méthodes numériques dans les Systèmes Experts. Nous aborderons et essayerons de résoudre les questions suivantes:

Qu'entend-on par "intelligence" lorsqu'on parle d'I.A., et en quoi les méthodes numériques participent-elles de cette intelligence ?

Peut-on réellement séparer la partie "intelligente" d'un S.E. d'une autre partie, considérée comme non intelligente, dans laquelle seraient rangés les outils mathématiques ?

Nul n'ignore en effet, que nombre de systèmes experts dont les performances ne sont plus mises en doute, reposent largement sur des bases numériques. Citons le système MYCIN (SHORTLIFFE 76), d'inspiration probabiliste comme nous le verrons plus loin. Il y a aussi les systèmes basés sur la théorie des sous-ensembles flous tels SPHINX (FIESCHI 83) (FIESCHI 84 a) ou PROTIS (SOULA 83).

Nous verrons tout d'abord les méthodes numériques de décision probabiliste, puis les méthodes non probabilistes. Ces méthodes

## Chapitre I

seront exposées succinctement et assorties d'exemples précis d'applications avant d'être analysées et comparées chaque fois que cela sera possible.

On ne s'étonnera pas de voir certains systèmes experts privilégiés par rapport à d'autres. L'objet de ce travail n'est pas de créer une nouvelle classification des S.E. mais de réaliser une étude des méthodes mathématiques utilisées. Les S.E. cités le sont à titre d'illustration de l'exposé.

### **Nota Bene:**

1) dans tout l'exposé, nous ne ferons pas de distinction entre les mots "signe" et "symptôme", dont la nuance, dans le langage médical, n'a pas d'incidence sur les problèmes décisionnels.

2) l'ordre dans lequel sont introduites les diverses théories ne préjuge en aucune façon de leurs qualités respectives; cet ordre répond uniquement à un souci de clarté de l'exposé.

## **I.2 LES METHODES DE DECISION A CARACTERE PROBABILISTE**

### **I.2.1 FORMULE DE BAYES**

#### **I.2.1.1 RAPPEL THÉORIQUE**

Cette formule, rendue célèbre par un pasteur écossais, Thomas BAYES (BAYES 1763), appartient à la théorie des probabilités, elle permet de calculer des probabilités **a posteriori** d'événements, connaissant leur probabilité **a priori**, ainsi que les probabilités de réalisation de sous-ensembles de ces événements.

Donnons un exemple pour fixer les idées: il est possible avec

cette formule de calculer la probabilité qu'un malade soit atteint d'une maladie donnée,  $D_i$ , connaissant la probabilité  $p(D_i)$  de cette maladie dans la population dont est issu le malade (probabilité **a priori**), ainsi que les probabilités d'occurrence des divers symptômes provoqués par cette maladie:  $p(S_j/D_i)$  (probabilité du symptôme  $S_j$  chez un malade atteint de la maladie  $D_i$ ).

$$p(D_i/S_j) = \frac{p(D_i) \times p(S_j/D_i)}{\sum_{k=1}^r p(D_k) \times p(S_j/D_k)}$$

Les  $D_k$ , pour  $k$  variant de 1 à  $r$ , représentent les maladies ( $D_i$  comprise) dans lesquelles le symptôme  $S_j$  a une probabilité non nulle d'apparaître.

La formule peut se généraliser à une collection  $C$  de symptômes observés chez le malade:

$$p(D_i/C) = \frac{p(D_i) \times p(C/D_i)}{\sum_{k=1}^r p(D_k) \times p(C/D_k)}$$

$C$  représente l'événement  $(S_1 \cap S_2 \cap \dots \cap S_m)$  correspondant à l'observation de  $m$  symptômes.

Cette formule a la double qualité de se prêter parfaitement à un traitement algorithmique, et de reproduire parfois, mais parfois seulement, le raisonnement du médecin avec la rigueur qu'apportent les mathématiques. Ces quelques citations en témoignent:

- "(...) le théorème de Bayes permet de calculer les probabilités des maladies en présence de certains symptômes (...). C'est implicitement le raisonnement du médecin qui recherche un

## Chapitre I

diagnostic" (GREMY 69, p. 94).

- "Le bon sens n'est en fait que l'expression de données statistiques, c'est-à-dire les probabilités d'apparition d'une maladie dans une population donnée et la fréquence d'apparition dans une affection d'un tableau donné" (VALOIS 81, p. 44).

- "This sequential process of accumulating evidence to alter a disease probability accords well with the intuitive method of a doctor" (CARD 73, p. 424).

En ce qui concerne l'aspect algorithmique, plaçons-nous dans la situation d'un médecin en présence d'un malade en consultation. Ce malade présente une collection C de m symptômes. Le médecin peut obtenir les probabilités **a posteriori** de toutes les maladies dans lesquelles C peut être observé. En poussant plus avant l'interrogatoire du patient, ou encore à la suite d'un examen, il constate un nouveau symptôme :  $S_{m+1}$ , qui vient s'ajouter à la collection  $C_m$ :

$$C_{m+1} = C_m \cup (S_{m+1})$$

Toutes les probabilités **a posteriori** déjà calculées, deviennent maintenant les probabilités **a priori** des maladies dans une nouvelle étape du raisonnement:

$$p(D_i/C_{m+1}) = \frac{p(D_i) \times p(C_{m+1}/D_i)}{\sum_{k=1}^s p(D_k) \times p(C_{m+1}/D_k)}$$

s est inférieur (ou égal) à r puisque l'ajout d'un nouveau symptôme a forcément réduit le nombre des maladies possibles.

La formule de BAYES a reçu de très nombreuses applications

médicales, dont certaines, HEME (FLEHINGER 75), débouchent sur un véritable processus d'aide à l'affinement de la connaissance médicale diagnostique. Nous allons en détailler une, parmi les plus marquantes.

#### I.2.1.2 UNE APPLICATION PROBANTE : LE SYSTEME HELP

Le système HELP (WARNER 72 a) (WARNER 79) (PRYOR 82) est un important système hospitalier multifonction : gestion de dossiers médicaux, aide au diagnostic, au pronostic et à la thérapeutique, évaluation du système de santé, gestion administrative, aide à l'enseignement, etc. Il correspond certainement à l'un des plus performants Systèmes d'Information Hospitalière Intégrée (S.I.H.I) développé à ce jour (CINQUIN 84).

Si ce système tout à fait impressionnant est cité ici, ce n'est pas pour ses caractéristiques d'ampleur ou de performances, mais bien plutôt parce qu'il plonge ses racines les plus anciennes dans la décision probabiliste Bayésienne.

En effet, son principal instigateur, le Professeur H.R. WARNER, s'était intéressé, depuis les années cinquante, à la mise en oeuvre de la méthode bayésienne dans un centre hospitalo-universitaire (WARNER 61) (WARNER 72 b).

La base de données de malades, nécessaire dans toute application du théorème de BAYES, est allée en se développant, jusqu'à produire le système HELP actuel.

Pour donner un aperçu très bref de la structure de ce système, nous dirons qu'il est constitué d'un réseau de sous-systèmes, au

## Chapitre I

centre duquel se trouvent :

- une base de données de malades
- une base de connaissances médicales de type schémas (incluant la notion d'algorithme)
- un dictionnaire/interface de communication et de codification des données médicales
- un interprète de secteurs qui constitue le moteur d'exploitation de la base de connaissances.

L'inférence bayésienne, dans l'état actuel d'évolution du système, n'occupe plus une place centrale : elle est optionnelle parmi un certain nombre de méthodes inférentielles : inférences quasi-certaines dictées par la physio-pathologie, inférences probabilistes pour les domaines incertains.

La base de connaissances est constituée par un ensemble de secteurs, portions élémentaires de connaissance, au nombre d'environ 80 000.

Cette notion de secteur regroupe :

- l'algorithme de décision
- un texte en langage clair associé à la décision
- un système de contrôle d'informations demandées à la base, des dossiers médicaux et aux divers terminaux.

Les secteurs ressemblent à des fiches aide-mémoire très complètes, sur des points particuliers de la pathologie, et regroupent un ensemble de règles de décisions, de vérifications à pratiquer dans tel ou tel cas particulier, de conduites à tenir, etc. Le secteur est à rapprocher typiquement de la notion de **schéma** dans

son sens le plus large, à savoir celui qui inclut la notion d'algorithme local de décision.

L'ensemble des secteurs est évidemment partitionné en sous-ensembles correspondant aux diverses branches de la médecine, chacun de ces sous-ensembles étant sous la responsabilité d'un expert ou d'un comité d'experts médicaux. Ce groupe d'expert est chargé d'assurer la création, la validation, et la modification éventuelle des secteurs du domaine concerné.

La sélection du ou des secteurs appropriés, lors d'une requête décisionnelle, est facilitée par une immatriculation des secteurs selon trois numéros : numéro de classe de connaissance, numéro de rubrique dans la classe, numéro de secteur dans la rubrique.

Lors d'une demande d'aide au diagnostic, les informations concernant le malade sont codées de la même manière (classe de connaissance médicale, numéro de rubrique). La sélection des secteurs applicables au cas traité se fait par "matching" partiel sur les codes.

Par exemple, la source d'information : "critères pour block de branche gauche" fait référence, dans son codage, aux secteurs de l'électrocardiographie dont la classe est 03 et la rubrique est 64.

Chaque secteur peut faire appel à d'autres secteurs pour sa résolution. Les secteurs peuvent également être organisés selon un schéma décisionnel arborescent pour une partie de la pathologie.

L'expert qui modifie un secteur a la responsabilité d'en informer les experts responsables des secteurs référencés.

## Chapitre I

Un programme interactif de gestion (HCOM) permet d'effectuer sur la base de connaissances toutes les opérations nécessaires. C'est un langage de commandes d'édition de fichiers-secteurs.

Un secteur (MEDLAB) est composé de six chapitres:

1. Le texte à communiquer à la liste des destinations (3) si le secteur est satisfaisant.
2. Les divers codes de référence du secteur:  
code de propriété (classe, rubrique, numéro), code de sécurité et code de priorité d'exécution.
3. La liste des destinataires autorisés du texte de réponse (1)
4. Les conditions particulières d'application de l'algorithme décisionnel (4).
5. L'évaluation finale (le type de données produites par le secteur)
6. La logique du secteur : recherches et manipulations de données.

C'est ce sixième composant d'un secteur que nous détaillerons le plus pour l'intérêt qu'il présente dans notre étude sur les méthodes de décision médicale.

Cette partie de logique décisionnelle constitue un véritable **programme** à l'intérieur du secteur. Elle est constituée par une succession d'actions référencées chacune par une étiquette alphabétique.

L'association étiquette + action est appelée un item; chaque item fournit un résultat; l'étiquette de l'item sert à référencer l'item et à fournir le résultat (d'une manière similaire à une fonction sans paramètre en langage évolué tel que le pascal).

Il existe cinq actions possibles, c'est-à-dire cinq types d'items:

1. SEARCH: recherche d'une donnée particulière, soit dans les enregistrements propres à un malade, soit de manière interactive. Une action SEARCH associe à son étiquette 3 registres:

- . registre EXISTENCE,
- . registre DATE,
- . registre VALEUR,

dans lesquels sont rangés les résultats de la recherche.

Le registre EXISTENCE d'un item est généralement utilisé pour des actions conditionnelles:

exemple: l'expression IF EX : B (action), signifie:  
si le registre EXISTENCE de l'item B a pour valeur vrai,  
alors faire l'action indiquée.

2. ARITH : calcul arithmétique sur les données des registres de type VALEUR.
3. EXIST : test d'existence d'une donnée ; pratiqué sur les registres de type EXISTENCE.
4. CHRON : calculs chronologiques pratiqués sur des registres de type DATE.
5. PROB : analyse probabiliste bayésienne.

Voici un exemple de programme décisionnel :

```
A SEARCH : (A) [FC], [SCT] CHRONIC BRONCHITIS BY HISTORY =
B SEARCH : (A) [FC] CHEST AND LUNG EXAM, [N] PULMONARY
           AUSCULTATION, [ADJ] RHONCHI, B [DCM] PRESENT, (C) [DCM]
           ABSENT
IF EX     : A AND (B OR C), USE VAL : EX SUBITEM B + 10 *
           EX SUBITEM C
C PROB   : A, IF EX : B, USE VAL : B, MIN : (1,10)
```

## Chapitre I

```
TRUE : (0.7, 0.3), FALSE : (0.2, 0.8)
D SEARCH : (A) [FCJ] CHEST AND LUNG EXAM [NIJ] PULMONARY
AUSCULTATION, [ADJJ] WHEEZING, (B) [DCM] PRESENT,
(C) [DCM] ABSENT
IF EX : A AND (B OR C), USE VAL : EX SUBITEM B + 10
EX SUBITEM C
E PROB : C, IF EX : D, USE VAL : D, MIN : (1, 10)
TRUE : (0.6, 0.4), FALSE : (0.4, 0.6)
```

Cette suite d'instructions, plutôt hermétique pour un non-initié au langage HELP, correspond en clair à l'algorithme suivant:

**A** Rechercher la bronchite chronique dans le fichier histoire (résultat : la probabilité **a priori** de la maladie).

**B** Rechercher la présence ou l'absence de rhonchi par une auscultation des poumons.

**C** Calculer la probabilité **a posteriori** de la bronchite chronique avec les valeurs suivantes:

```
si rhonchi présents
alors p(s/m) = 0,7 et p(s/non m) = 0,2
sinon p(s/m) = 0,3 et p(s/non m) = 0,8
```

$$\text{Rappel : } p(m/s) = \frac{p(m) \times p(s/m)}{p(m) \times p(s/m) + p(\text{non } m) \times p(s/\text{non } m)}$$

Remarque :

a) Le calcul de probabilités est effectué avec les valeurs correspondant à l'occurrence ou la non-occurrence du symptôme.

b) La probabilité **a priori** est fournie par l'item A.

**D** Rechercher la présence ou l'absence de respiration asthmatiforme

**E** Calculer la probabilité **a posteriori** de la bronchite chronique

avec les valeurs suivantes :

si respiration asthmatiforme  
 alors  $p(s/m) = 0,6$  et  $p(s/\text{non } m) = 0,4$   
 sinon  $p(s/m) = 0,4$  et  $p(s/\text{non } m) = 0,6$

Remarque : la probabilité **a priori** est maintenant fournie par l'item C.

Le second exemple proposé montre l'importance de l'inférence bayésienne dans le contexte bien délimité d'un centre hospitalier, en l'occurrence celui du LDS Hospital, à Salt Lake City (UTAH, USA).

Nous le reproduisons **in extenso** et en langage clair. Le commentaire qui suit, du fait qu'il correspond à une traduction assez libre d'une partie de l'article (PRYOR 82), sera mis entre guillemets.

**Logique du secteur liquide pleural :**

- A X-ray, chest, specific exam
- B 10
- C If (A EQ 7) then B=524
- D If (A EQ 1) or (A EQ 2) or (A EQ 3) then B=150
- E If (A EQ 56) then B=79
- F Chest pain (reason)
- G ? Pleural effusion (reason)
- H Pre-op (general reason)
- I Followup known disease (general reason)
- J Chair (transportation)
- K Post-op (General reason)
- L Walk (transportation)
- M ? Pneumothorax (reason)
- N Coronary artery disease (reason)
- O ? Pulmonary embolus (reason)
- P Pleural fluid reported in last 2 days

	(ITEM)	(PRIOR PROB.)	P(S/D)	P(S/D)	
Q	Prob binary	F	B	.022	.070
R		G	Q	.193	.029
S		H	R	.009	.150
T		I	S	.210	.146
U		J	T	.546	.420
V		K	U	.395	.255
W		L	V	.004	.084

## Chapitre I

X		M	W	.059	.021
Y		N	X	.004	.030
Z		O	Y	.053	.018
AA		P	Z	.243	.036
BB	(A)	SMA-12, Albumin GMZ			
CC	(A)	CBC, Hemoglobin GMZ			
DD	(A)	SMA-12, Calcium MGZ			
EE	(A)	Age			

	ITEM	(P)	MIN	MAX	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FF Pr. dist	BB	AA	1.1	5.4	6/2	13/4	16/8	23/13	22/19	12/25	6/17	2/12
GG Pr. dist	CC	FF	4.7	18.0	5/2	14/7	18/9	23/11	18/19	13/23	6/16	2/14
HH Pr. dist	DD	GG	6.8	11.0	6/2	12/4	15/10	21/15	18/28	18/27	6/10	3/3
II Pr. dist	EE	HH	0	100	1/1	4/6	5/11	8/11	25/24	34/31	18/13	5/3

### COMMENTAIRE

"Ce secteur, mis en oeuvre pour chaque prescription de rayons X à un patient, a pour but le pronostic de la présence de liquide pleural par radiographie.

L'item A provoque une recherche, à travers la base de données du patient, d'une éventuelle demande d'examen du thorax qui n'aurait pas encore été effectuée par HELP. Si la réponse est positive, le registre valeur de l'item A est affecté d'un code numérique compris entre 1 et 8 selon le type d'examen qui a été demandé.

L'item B fournit la probabilité **a priori** à 1/100 près de la présence de liquide pleural pour un patient du LDS Hospital pour lequel un examen par rayons X a été pratiqué.

Les items C à E modifient la probabilité initiale fournie par l'item B selon la procédure spécifique de l'item A qui a été prescrite.

Lorsque la probabilité **a priori** de présence de liquide pleural a été établie en fonction du type d'examen à pratiquer, l'enregistrement du patient est recherché afin de tenir compte des

autres données qui vont modifier la probabilité de présence de liquide pleural.

Les items F à O recherchent les raisons spécifiques de prescription de l'examen, les conditions spéciales éventuelles de transport. Ces items sont nécessaires du fait qu'ils modifient de façon significative la probabilité de la maladie. Ils prennent pour valeur 1 (vrai) ou 0 (faux) selon, respectivement, l'existence ou la non-existence de la donnée recherchée dans la chaîne de données créée lors de l'ordonnance de l'examen. Ces valeurs sont utilisées dans les calculs de probabilité **a posteriori**.

L'item P prend pour valeur 1 si l'enregistrement du patient contient des examens radiologiques datant de moins de deux jours, ayant montré l'existence de liquide pleural.

Les items Q à AA calculent les probabilités bayésiennes basées sur les données des items F à P. Ces données sont des symptômes caractérisés par leur présence ou leur absence. Cette alternative présence-absence détermine les paramètres du calcul bayésien.

Dans l'item Q, les nombres 0,022 et 0,070 reflètent les probabilités que la douleur à la poitrine (item F= 1) soit donnée comme raison pour ordonner un examen radiologique pour un patient du LDS Hospital avec ou sans liquide pleural décelable radiologiquement. La valeur de l'item B contient la probabilité **a priori** qui doit être utilisée dans la formule de Bayes lorsque l'item Q est exécuté. Dans l'occurrence d'une douleur à la poitrine, la probabilité du liquide pleural est diminuée si cette douleur a constitué le motif de la demande d'examen. Si l'item à partir

## Chapitre I

duquel est appliqué un calcul bayésien (dans le cas de F, c'est la douleur à la poitrine) n'a pas la valeur vrai, alors, la probabilité **a posteriori** reste inchangée et garde donc la valeur de la probabilité **a priori**.

L'item R utilise ensuite le résultat de l'item Q comme probabilité **a priori** et calcule la probabilité conditionnelle suivante selon que le liquide pleural a été donné ou non comme motif d'examen.

Dans le cas d'une variable continue telle qu'un résultat de laboratoire, la formule de Bayes peut encore être appliquée, ceci par l'utilisation d'une distribution de probabilités. Dans cette éventualité, la distribution des valeurs des paramètres cliniques doit être connue pour des populations avec et sans la maladie. Ces probabilités peuvent alors être utilisées de la même manière que des variables binaires.

Dans le cas d'une distribution de probabilités, les paramètres de la distribution sont introduits au moment de la création de la partie logique du secteur. Cette distribution est obtenue en partitionnant les populations en classes puis en calculant, pour chaque classe, la probabilité pour que les patients du groupe des atteints et ceux du groupe des non-atteints par la maladie aient une valeur de paramètre qui les mettent dans cette classe.

La logique HELP utilise la valeur actuelle du paramètre clinique pour retrouver la probabilité appropriée pour le calcul.

Des calculs bayésiens sont effectués dans les items FF à II en utilisant les variables continues listées dans les items BB à EE. S'ils existent (registre EXISTENCE valant "vrai"), le programme

## Chapitre I

affecte aux registres VALEUR des items BB à EE, respectivement la plus récente valeur non nulle pour l'albumine du sérum (item BB), l'hémoglobine (item CC), le calcium (item DD) et l'âge (item EE).

L'item FF calcule la probabilité d'occurrence de liquide pleural à partir de la valeur de l'albumine pour le patient en utilisant l'item AA comme probabilité *a priori*. Le format du calcul bayésien distribué spécifie l'intervalle des valeurs possibles de l'albumine (entre 1,1 et 5,4 gr) et, pour huit classes, la probabilité pour qu'un patient présentant ou non du liquide pleural, ait les valeurs de son albumine dans cette classe. Ce sont les huit paires de nombres données en dernière position dans l'item FF.

L'item GG correspond également à un calcul de probabilités qui utilise la valeur de l'hémoglobine comme variable, et le résultat de l'item FF comme probabilité *a priori*.

Si les items BB, CC, DD ou EE ne figurent pas dans l'enregistrement du patient, leur valeur est celle de la probabilité *a priori* de cette ligne.

La probabilité finale du liquide pleural est la valeur de l'item II. Elle reflète l'influence des tests de laboratoire, des symptômes radiologiques détectés et des motifs de demande de l'examen. Seules les données qui modifient significativement la probabilité sont incluses dans la logique des secteurs HELP. Des efforts sont faits pour utiliser des paramètres indépendants.

Des secteurs utilisant ce type de logique ont été écrits pour une large variété d'applications."

**CONCLUSION** : Ce dernier exemple montre combien l'inférence

bayésienne est développée et utilisée régulièrement à grande échelle. Il s'agit, en effet, non seulement du LDS Hospital, mais également de plusieurs autres centres hospitaliers aux Etats-Unis et en Allemagne. Le système HELP est donc à considérer comme une application majeure de l'inférence bayésienne en milieu médical et dans un système expert assez complexe d'aide au diagnostic. Une telle application contredit l'idée que la méthode bayésienne est dépassée au regard des systèmes experts actuels, ou peu applicable à l'aide au diagnostic médical. Cette méthode reste un outil rigoureux d'aide à la décision, à la condition formelle de travailler dans un domaine bien connu de la pathologie (problème d'indépendance des symptômes) et de disposer de grandes populations. Il va sans dire que les probabilités **a priori** ou calculées sont vraies dans leur système de référence (en l'occurrence le centre hospitalier) et qu'il faut se méfier des extrapolations hâtives.

### I.2.2 MODELE MATHEMATIQUE DE SHORTLIFFE

La formule de Bayes et, plus généralement, le calcul des probabilités sont à l'origine l'une des premières réflexions importantes sur la modélisation du raisonnement approximatif en médecine: E.H. SHORTLIFFE (SHORTLIFFE 75) ( voir également (ADAM 76) (SHORTLIFFE 76), après avoir montré l'intérêt et les limitations de cette formule, a proposé un système de raisonnement approximatif basé sur le principe de quantification numérique de l'incertitude au moyen de deux mesures attribuées à une hypothèse H: la **mesure de confirmation** ("measure of belief") (MB) et la **mesure d'infirmité** ("measure of disbelief") (MD) de H. La

quantification numérique de l'incertitude se retrouve, malgré tous ses inconvénients (voir le paragraphe "Comparaison et discussion"), dans nombre de modèles de raisonnement incertain. On trouvera dans (FARRENY 85) un exposé détaillé du système expert MYCIN. Nous analyserons ici les bases mathématiques sous-jacentes à ce système.

Les mesures de confirmation et d'infirmité se traduisent, non seulement dans MYCIN, mais également dans d'autres systèmes (SOULA 83), par une valeur réelle comprise entre zéro et un. Par exemple : la confirmation  $C(H,E)$  d'une hypothèse (ou d'une conclusion)  $H$  en présence d'une observation  $E$  peut être évaluée à 0,7.

Cette mesure évoque la probabilité conditionnelle de  $H$  en présence de l'événement  $E$ . Cependant, l'évocation s'arrête là: un expert peut parfaitement attribuer la valeur 0,7 à  $C(H,E)$ , mais ne pas reconnaître comme juste la valeur 0,3 de  $C(\text{non } H,E)$  dictée par le calcul des probabilités. Dans l'esprit de l'expert, une observation  $E$  qui renforce une hypothèse  $H$  peut difficilement renforcer, même à un degré moindre, le complément de la même hypothèse. Ainsi, pour garder du calcul des probabilités ce qui va dans le sens de l'intuition, tout en supprimant le problème de la probabilité du complément, SHORTLIFFE a construit ses deux mesures de confirmation et d'infirmité de la façon suivante:

Soient

.la proposition: "si  $F$  alors  $H$  avec la probabilité  $p(H/F)$ "

.la probabilité **a priori** de  $H$ , notée  $p(H)$ ,

la mesure de confirmation s'exprime par:

$$MB(H,F) = \begin{cases} 1 & , \quad \text{si } p(H) = 1 \\ \frac{p(H/F) - p(H)}{1 - p(H)} & , \quad \text{si } p(H/F) > p(H) \\ 0 & , \quad \text{si } p(H/F) \leq p(H) \end{cases}$$

et la mesure d'infirmité s'exprime par:

$$MD(H,F) = \begin{cases} 1 & , \quad \text{si } p(H) = 0 \\ \frac{p(H) - p(H/F)}{p(H)} & , \quad \text{si } p(H/F) < p(H) \\ 0 & , \quad \text{si } p(H/F) \geq p(H) \end{cases}$$

On remarquera aisément que ces deux mesures sont "symétriques"; pour tout événement, il n'existe que trois possibilités:

- F confirme l'hypothèse H :  $MB(H,F) > 0$  &  $MD(H,F) = 0$
- F infirme l'hypothèse H :  $MB(H,F) = 0$  &  $MD(H,F) > 0$
- F est indépendant de H :  $MB(H,F) = MD(H,F) = 0$

Une troisième mesure: le **facteur de crédibilité** CF ("certainty factor"), donne une valeur résultant des deux mesures précédentes:

$$CF(H,F) = MB(H,F) - MD(H,F)$$

Ses valeurs sont donc comprises entre - 1 et 1. L'interprétation du facteur de crédibilité se déduit aisément de l'interprétation, exposée précédemment, des mesures de confirmation et d'infirmité.

Le facteur de crédibilité évoque parfois fortement la probabilité conditionnelle: supposons le cas où  $p(H)$  est faible et  $p(H/F)$  proche de 1. En médecine, nous dirons que la prévalence de la maladie H est faible et que le symptôme F est toujours présent pour cette maladie. Il est pathognomonique si, de plus,  $p(F/M_i)$  est peu différent de zéro lorsque  $M_i$  est différent de M. Nous pouvons

## Chapitre I

écrire:

$$CF(H,F) = \frac{p(H/F) - p(H)}{1 - p(H)}$$

La valeur de cette expression est peu différente de  $p(H/F)$ .

Cependant, une fois de plus, le raisonnement probabiliste est différent de la théorie des facteurs de crédibilité. Il suffit, pour le montrer, de calculer le facteur de crédibilité du complément de H en présence de F, ( $p(H/F)$  restant proche de 1 et  $p(H)$  restant faible):

$$CF(\text{non } H, F) = MB(\text{non } H, F) - MD(\text{non } H, F)$$

$$CF(\text{non } H, F) = 0 - \frac{p(\text{non } H) - p(\text{non } H/F)}{p(\text{non } H)}$$

$$CF(\text{non } H, F) = \frac{(1 - p(H/F)) - (1 - p(H))}{p(\text{non } H)}$$

$$CF(\text{non } H, F) = \frac{p(H) - p(H/F)}{1 - p(H)}, \quad \text{peu différent de } -p(H/F),$$

alors que  $p(\text{non } H/F) = 1 - p(H/F)$ .

Les problèmes de diagnostic reposent en général sur une série d'observations simultanées ou séquentielles. Nous avons vu que la formule de Bayes se prête bien au traitement d'une série d'observations pour le calcul d'une probabilité **a posteriori**. Quel est alors l'équivalent de ce traitement selon la théorie des facteurs de crédibilité ? E.H. SHORTLIFFE propose les formules suivantes:

Chapitre I

$$(a) \text{ MB}(H, F1 \ \& \ F2) = \begin{cases} 0 & \text{si } MD(H, F1 \ \& \ F2) = 1 \\ \text{MB}(H, F1) + \text{MB}(H, F2) \cdot (1 - \text{MB}(H, F1)) & \text{sinon} \end{cases}$$

$$(b) \text{ MD}(H, F1 \ \& \ F2) = \begin{cases} 0 & \text{si } MB(H, F1 \ \& \ F2) = 1 \\ \text{MD}(H, F1) + \text{MD}(H, F2) \cdot (1 - \text{MD}(H, F1)) & \text{sinon} \end{cases}$$

la formule (a) s'interprète ainsi : la mesure de confirmation résultant d'une seconde observation F2 s'ajoute à la mesure résultant d'une première observation F1, à un facteur près:  $(1 - \text{MB}(H, F1))$ . Ce facteur est le résidu de confirmation issu de l'observation précédente. Cette méthode est logique et conserve les mesures de confirmation dans l'intervalle  $[0, 1]$ . La formule (b) s'interprète de manière symétrique par rapport à la formule (a).

Remarquons que la définition des mesures de confirmation et d'infirmité n'est pas exempte de contradiction; la formule définissant  $\text{MB}(H, F1 \ \& \ F2)$  implique:

$$\text{MB}(H, F2) = \frac{\text{MB}(H, F1 \ \& \ F2) - \text{MB}(H, F1)}{1 - \text{MB}(H, F1)}$$

Or, par définition de MB, si  $P(H/F1)$ ,  $P(H/F2)$  et  $P(H/F1 \ \& \ F2)$  sont supérieurs à  $P(H)$ :

$$\text{MB}(H, F2) = \frac{P(H/F2) - P(H)}{1 - P(H)} \quad \text{d'une part,}$$

$$\text{MB}(H, F2) = \frac{P(H/F1 \ \& \ F2) - P(H/F1)}{1 - P(H/F1)} \quad \text{d'autre part, si l'on}$$

applique, pour le calcul de  $\text{MB}(H, F1 \ \& \ F2)$ , la formule initiale de définition de  $\text{MB}(H, F)$ .

Il est clair que ces deux quantités peuvent être tout à fait

distinctes, car les probabilités  $P$  et  $P (./F1)$ , qui interviennent de la même manière dans les 2 formules, sont a priori très différentes. Cela réduit, selon nous, l'intérêt des mesures de confirmation et d'infirmité.

Les problèmes de conjonction et de disjonction d'hypothèses sont pris en compte dans les formules suivantes, dans lesquelles  $E$  est un ensemble non vide d'observations:

- (c)  $MB(H1 \& H2, E) = \min (MB(H1, E), MB(H2, E) )$
- (d)  $MD(H1 \& H2, E) = \max (MD(H1, E), MD(H2, E) )$
- (e)  $MB(H1 \text{ ou } H2, E) = \max (MB(H1, E), MB(H2, E) )$
- (f)  $MD(H1 \text{ ou } H2, E) = \min (MD(H1, E), MD(H2, E) )$

Ce modèle n'est pas sans rappeler la théorie des sous-ensembles flous, abordée plus loin, qui utilise en partie les mêmes outils.

### 1.2.3 INDICATEURS D'EFFICACITE ET VALEURS PREDICTIVES

Si la formule de Bayes a pour fonction principale de calculer des probabilités de maladie et donc de s'attacher presque exclusivement au diagnostic, nous avons vu qu'elle péchait particulièrement en ce qui concerne les définitions même de ces maladies par des symptômes.

Les indicateurs d'efficacité (GALEN 75), (GRENIER 83) s'attachent au contraire à démontrer "l'efficacité" d'un symptôme dans l'évocation d'une maladie. Ils sont construits statistiquement au moyen d'un groupe de patients testés dans un service médical. Dans ce groupe déterminé de patients, on s'attache à analyser un symptôme particulier ou une conjonction particulière de symptômes. Ce symptôme peut ressortir à l'une quelconque des catégories:

## Chapitre I

manifestation clinique, résultat biologique, radiographie, etc. Un symptôme peut être toujours présent dans une maladie et ne se rencontrer dans aucune autre, on parle alors de symptôme **pathognomonique**. D'autre part, si un symptôme est toujours au rendez-vous, lorsque la maladie qui le provoque est déclarée, il est dit **"totalement sensible"** ou **constant** pour cette maladie. Il entre alors en général dans la définition nosologique de la maladie (ou du syndrome) en question.

Dans la pratique, les symptômes constants et, plus encore, les symptômes pathognomoniques, sont rares. On dispose alors de trois outils statistiques importants, constitués : d'une part de deux indicateurs d'efficacité : **la sensibilité** et **la spécificité** d'un symptôme; d'autre part, du **rapport de vraisemblance** d'une maladie en fonction d'un symptôme. La valeur de leur estimateur se calcule d'une manière très simple à partir d'un tableau dit **tableau à quatre cases** ou encore **Matrice de Décision**:

		Maladie m		
		Atteints	Non-atteints	
Symptôme (ou groupe de symptômes) s	Présent	a	b	N
	Absent	c	d	
	Total	a + c	b + d	

a : effectif des patients atteints de la maladie m présentant le symptôme s

b : effectif des patients non atteints de la maladie m, mais présentant néanmoins le de symptôme s

a + c : effectif des malades atteints de la maladie m

b + d : effectif des malades non-atteints de m

N : effectif général.

N.B. On trouve une importante utilisation de la Matrice de Décision dans la ROC-Analyse (§ II.5.3.5).

#### INDICATEURS D'EFFICACITE:

##### 1) Sensibilité:

La sensibilité  $Se$  du symptôme  $s$  pour la maladie  $m$  est définie par la fréquence de  $s$  dans  $m$ :

$$Se = \frac{a}{a + c} \quad (1)$$

##### 2) Spécificité:

La spécificité  $Sp$  du symptôme  $s$  pour la maladie  $m$  est définie par la fréquence de l'absence de  $s$  dans les maladies autres que  $m$ :

$$Sp = \frac{d}{b + d} \quad (2)$$

La fréquence du symptôme  $s$  en l'absence de la maladie  $m$  est égale à  $b / (b+d)$  soit  $1-Sp$ ;  $Se$  et  $Sp$  varient entre zéro et un.

On peut établir un **intervalle de confiance** pour chacun des rapports (1) et (2). En effet,  $Se$  et  $Sp$  sont des fréquences empiriques qui suivent des lois gaussiennes lorsque les effectifs  $a+c$  et  $b+c$  sont supérieurs ou égaux à 30.

On pourra ainsi affirmer que le symptôme  $s$  est totalement sensible dans la maladie  $m$  avec 1% de chance de se tromper si le test gaussien à 1% montre que  $1-Se$  est non significativement  $\neq 0$ .

De même, s sera totalement spécifique de m (ou pathognomonique) à 1% d'erreur si le test gaussien à 1% montre que  $1 - S_p$  est non significativement  $\neq 0$ .

Le coefficient de Youden (LACAINE 79):  $Y = S_e + S_p - 1$ , permet de regrouper en une seule formule les deux indicateurs d'efficacité.

**3) Rapport de vraisemblance :**

Le rapport de vraisemblance L permet également de mesurer, au moyen d'une seule formule, la vraisemblance du diagnostic de la maladie m par rapport au symptôme s.

$$L = \frac{S_e}{1 - S_p}$$

**4) Taux des faux positifs :**

$$TFP = \frac{b}{b + d}$$

**5) Valeur prédictive positive**

$$VP+ = \frac{a}{a + b}$$

**6) Valeur prédictive négative :**

$$VP- = \frac{d}{c + d}$$

Lorsqu'on cherche à infirmer un diagnostic, on choisit de

préférence un signe à sensibilité élevée ; un signe à spécificité élevée sera par contre recherché pour confirmer un diagnostic.

La valeur prédictive positive apporte un élément de réponse à la question : "si le signe est présent, quelle est la probabilité d'avoir la maladie ?". La valeur prédictive négative répond à la question opposée : "si le signe est absent, quelle est la probabilité de ne pas avoir la maladie ?".

Le Tableau à Quatre Cases représente un modèle théorique de référence. Il reste néanmoins d'un usage limité pour deux raisons principales:

- d'une part, il est très ponctuel et ne concerne qu'un seul signe et qu'une seule maladie à la fois. Il faudrait, logiquement, autant de tableaux que de couples (signe, maladie) différents, ce qui deviendrait démesuré et utilisable uniquement par une machine.
- d'autre part, il est tributaire d'un lieu, d'un laboratoire, d'une région, etc. La spécificité et la sensibilité des signes dépendent toutes deux de nombreux facteurs environnementaux: climat, conditions sanitaires, niveau de vie, etc.

Voici l'expression du théorème de Bayes en fonction de la sensibilité, de la spécificité d'un signe S et de la prévalence P d'une maladie M (LACAINE 78):

$$P(M/S) = \frac{Se \cdot P}{Se \cdot P + (1 - Sp) \cdot (1 - P)}$$

#### I.2.4 ANALYSE DE PROXIMITES PAR CORRELATIONS

#### I.2.4.1 INTRODUCTION

Une manière élégante d'envisager le problème de la recherche des meilleurs diagnostics pouvant expliquer un cas précis d'observations est de considérer la maladie, encore inconnue, du patient comme une "forme" à reconnaître parmi un ensemble prédéfini de formes fixées : les maladies.

Un outil privilégié de comparaison entre le profil symptomatique du malade et les profils de syndromes et de maladies enregistrés dans la machine est ici le **coefficient de corrélation** non centré, entre les deux profils.

L'idée originelle de cette méthode a été proposée par Claudine ROBERT (laboratoire de Mathématiques Pures, Université Scientifique, Technologique et Médicale de Grenoble) et appliquée à un problème de recherche du meilleur remède homéopathique à proposer à un patient.

#### I.2.4.2 EXPLICITATION DU MODELE THEORIQUE

##### A - REPRESENTATION DU DOMAINE

##### a) LES DIAGNOSTICS :

Nous emploierons ici le terme de diagnostic dans un sens large pouvant signifier soit un remède (dans le cas de l'homéopathie), soit une maladie ou un syndrome (dans le cas de l'allopathie).

L'ensemble des diagnostics en fonction des symptômes est représenté sous la forme d'une matrice  $(m \times n)$ :  $M$ . Les  $m$  lignes correspondent aux différents symptômes, les  $n$  colonnes correspon-

## Chapitre I

dent aux diagnostics. Une valeur  $a_{ij}$  située à l'intersection de la ligne  $i$  et de la colonne  $j$  représente l'occurrence du symptôme  $i$  dans le diagnostic  $j$ .

Les valeurs d'occurrences sont :

- 0 : symptôme n'apparaissant pas,
- 1 : symptôme apparaissant quelquefois,
- 2 : symptôme apparaissant souvent,
- 3 : symptôme apparaissant toujours dans le diagnostic.

Il est évident que cette échelle pseudo-quantitative de fréquences d'apparition de symptômes dans des maladies (ou pour des remèdes homéopathiques) laisse une très large place à l'incertitude. Les valeurs  $a_{ij}$  de la matrice  $M$  représentent les "poids" des symptômes dans les diagnostics. Ces poids peuvent résulter soit directement de l'estimation du spécialiste, soit d'une analyse statistique ou épidémiologique.

$$\begin{array}{c} S_1 \\ \vdots \\ S_i \\ \vdots \\ S_n \end{array} \begin{bmatrix} D_1 & & D_j & & D_n \\ & & \vdots & & \\ & & \vdots & & \\ \dots & \dots & a_{ij} & \dots & \\ & & \vdots & & \\ & & \vdots & & \end{bmatrix}$$

où les  $S_i$  représentent les symptômes et les  $D_j$  les diagnostics.

### b) LE PATIENT

Le patient est représenté par un vecteur colonne à  $m$  composantes. Ce vecteur représente l'observation du patient par le médecin. Il est donc sujet à évolution au cours des diverses étapes

## Chapitre I

de la recherche du diagnostic (examen clinique, examens biologiques, radiologiques, etc.)

Appelons P le vecteur issu de l'observation clinique initiale:

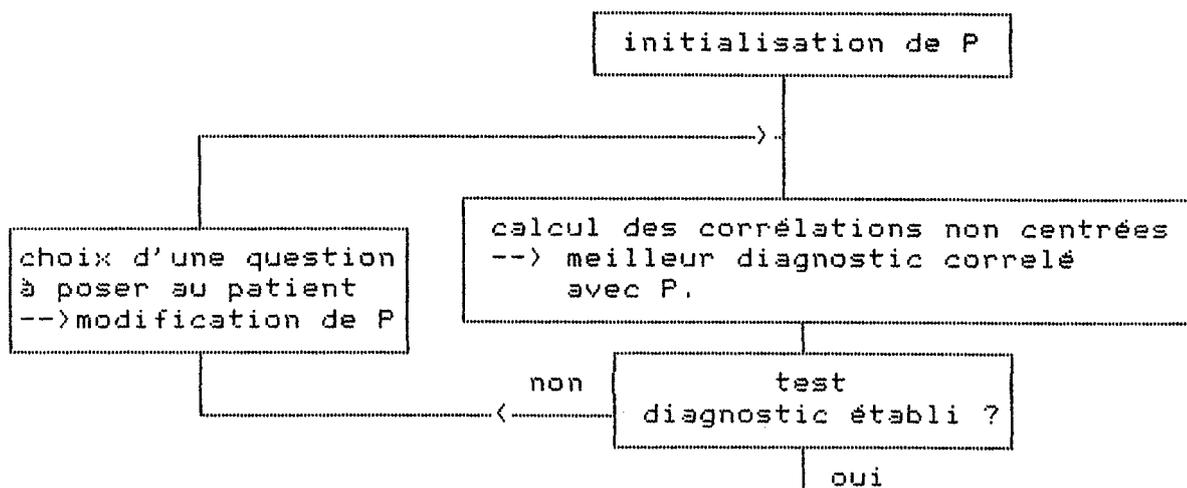
$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_m \end{bmatrix}$$

avec  $P_i = \begin{cases} 1 & \text{si le symptôme } S_i \text{ est observé} \\ 0 & \text{dans le cas contraire} \end{cases}$

### B - PROCESSUS DE RAISONNEMENT

#### a) ORGANIGRAMME GENERAL

Le processus de raisonnement automatique va constituer un cycle illustré par l'organigramme suivant:



#### b) NOTION D'ECLAIRAGE

Etant donné un diagnostic  $D_j$ , l'éclairage du patient par  $D_j$

consiste à remplacer, dans le vecteur du patient, les valeurs 1 des symptômes observés, par les valeurs de ces symptômes dans  $D_j$ .

Cette opération confère au vecteur P un caractère de "caméléon". Elle n'affecte en rien la qualité de l'observation par le médecin.

Le but poursuivi par ce moyen est l'obtention d'un coefficient de corrélation égal à 1 dans le cas limite où les deux vecteurs, patient et diagnostic sont identiques. Cette identité ne saurait avoir lieu si les valeurs du vecteur patient étaient limitées à l'alternative 0 ou 1, alors que celles du vecteur diagnostic pourraient varier entre 0 et 3.

La règle de transformation est la suivante:  
appelons PE le vecteur représentant l'éclairage momentané de P par un diagnostic  $D_j$ .

Si  $P_i = 1$  et  $a_{ij} \neq 0$   
Alors  $PE_i$  prend pour valeur  $a_{ij}$   
Sinon  $PE_i$  prend pour valeur  $P_i$

Cette opération n'est autre que le produit  $\otimes$  de deux vecteurs, termes à termes, ces deux vecteurs étant P et  $\text{Sup} ( D_j, \mathbf{1} )$ :

$$PE = P \otimes \text{Sup} ( D_j, \mathbf{1} )$$

$\mathbf{1}$  désigne le vecteur dont toutes les composantes sont égales à 1.

### c) CALCUL DU COEFFICIENT DE CORRELATION NON CENTRE

Appelons  $D_j$  le jème diagnostic de la base des diagnostics et  $r\langle j \rangle$  le coefficient de corrélation non centré entre le vecteur patient éclairé PE et le diagnostic  $D_j$ .

## Chapitre I

La corrélation est calculée sur des variables non centrées afin de favoriser les occurrences effectives de symptômes. Lorsque la variable est non-centrée, la non-occurrence d'un symptôme se traduit par la valeur nulle. Dans le cas d'un centrage sur la moyenne, le changement d'origine sur l'axe des valeurs confère à cette absence de symptôme une valeur non-nulle qui influence le coefficient de corrélation centré. Dans ce dernier cas, deux maladies totalement différentes, mais ayant en commun un grand nombre de symptômes non-occurents, seraient étroitement corrélées.

L'expression arithmétique du calcul est la suivante:

$$r\langle j \rangle = \frac{m * \sum_i PE_i * a_{ij}}{\sqrt{(m \sum_i PE_i^2) * (m \sum_i a_{ij}^2)}}$$

Rappelons que  $m$  est le nombre total de symptômes admis par le système. Les sommes sont effectuées sur l'indice  $i$  avec  $1 \leq i \leq m$ .

Rappelons également que le coefficient de corrélation  $r$  prend ses valeurs dans l'intervalle  $[-1, +1]$  des nombres réels, et qu'il s'interprète de la manière suivante:

Lorsque  $r$  est proche de  $+1$  ou de  $-1$ , les variables sont corrélées respectivement positivement (même sens de variation) ou négativement (sens inverse de variation).

Lorsque  $r$  est proche de zéro, les variables sont non-corrélées.

### d) FILTRAGE DES DIAGNOSTICS

Appelons,  $F$  le vecteur  $(r\langle 1 \rangle, \dots, r\langle n \rangle)$  des corrélations.

$F$  va servir de "filtre" de la base des diagnostics : nous retiendrons en effet comme diagnostics candidats uniquement ceux

dont le coefficient de corrélation sera "suffisamment grand".

Il conviendra pour cela de fixer un seuil  $a$  d'acceptation : un diagnostic  $D_j$  sera retenu comme candidat si  $r(j) > a$ .

Cette valeur  $a$  (supérieure à 0,5) sera choisie vraisemblablement expérimentalement ou à la suite d'une simulation.

#### e) CHOIX DU DIAGNOSTIC FINAL

Les diagnostics candidats étant ordonnés par ordre de corrélation décroissante, deux cas peuvent se produire :

- soit le premier des "candidats" est unique ou "loin" devant les autres, auquel cas il sera proposé comme diagnostic final,
- soit le premier des "candidats" est très proche (c'est-à-dire a un coefficient de corrélation peu différent) du ou des suivants. Dans ce cas, il faudra proposer une question pertinente à poser au patient pour départager les diagnostics candidats.

#### f) CHOIX D'UNE QUESTION PERTINENTE

La détermination optimale d'une question à poser à un stade donné de l'interrogatoire du malade est un problème important en médecine.

Le principe intuitivement le plus logique, semble être celui qui consiste à choisir un fait qui, s'il est confirmé, réduira de manière maximale le nombre des diagnostics candidats.

**N.B. :** Une formule un peu plus complexe, proposée par (SCHILD 78), est la suivante :  
soit  $S = \{S_1, \dots, S_k\}$  l'ensemble des  $k$  symptômes non observés et donc candidats à l'observation.

## Chapitre I

L'indice  $J$  du meilleur symptôme à proposer quelle que soit la réponse (présence ou absence) est celui qui correspond au minimum suivant:

$$\min_{i \in \{1, \dots, k\}} [K \max (D_{i+}, D_{i-}) + \min (D_{i+}, D_{i-})]$$

où  $D_{i+}$ , est le nombre de maladies restantes si le fait  $S_i$  est présent et  $D_{i-}$  est le nombre de maladies restantes si le fait  $S_i$  est absent.

La constante  $K$  est ajustée expérimentalement. La valeur 3 semble d'après les auteurs, donner les meilleurs résultats.

Les critères de choix d'une question pertinente sont assez nombreux. L'un des plus naturels consiste à proposer à l'observation le symptôme qui aura la meilleure "spécificité", c'est-à-dire celui qui désignera plus particulièrement un diagnostic parmi les candidats.

Le coefficient de spécificité sera exprimé par la formule:

$$Sp\langle j \rangle i = \frac{a_{ij}}{\sum_k a_{ik}}$$

avec:

\*  $Sp\langle j \rangle i$  : spécificité du symptôme  $i$  dans le diagnostic  $j$

\* la somme du dénominateur, variant sur  $k$ , est effectuée sur l'ensemble des diagnostics candidats.

Ce coefficient de spécificité est compris entre 0 et 1 et sera d'autant plus proche de 1 que le symptôme  $i$  sera absent des autres diagnostics.

La spécificité sera calculée pour les symptômes  $i$  non encore observés chez le patient ( $P_i = 0$ ), et apparaissant au moins une

## Chapitre I

fois dans les diagnostics candidats.

Après avoir posé la question, un nouveau vecteur P est proposé au calcul de corrélation (retour à c)

### g) FIN DU PROCESSUS

Le processus prend fin lorsque le diagnostic final est unique ou qu'il devance largement ses concurrents, ou encore lorsqu'il n'y a plus de questions à poser.

On peut également estimer que le processus est susceptible d'être arrêté à tout moment par le médecin si celui-ci a obtenu une liste de diagnostics candidats qui lui apporte suffisamment d'informations.

## I.3 THEORIE MATHEMATIQUE DE L'EVIDENCE

### I.3.1 INTRODUCTION

La théorie mathématique de l'évidence, proposée par Glenn SHAFER (SHAFER 76), a pour objectif de proposer un modèle à la fois rigoureux et naturel de raisonnement incertain qui tente d'échapper au cadre purement probabiliste.

En effet, ce qui est essentiellement reproché au calcul des probabilités, c'est de ne pas "épouser" les nuances d'un jugement. Par exemple, pour un événement e donné, la formule  $P(e) + P(\text{non } e) = 1$  a pour conséquence que, si l'on pense que l'événement e a de fortes chances de se produire (par exemple  $P(e) = 0,85$ ), cela implique que la probabilité qu'il ne se produise pas devient, du même coup, très faible (en l'occurrence = 0,15). Or, le jugement humain ne procède souvent pas selon ce principe : si je crois à

## Chapitre I

la réalisation d'un événement  $e$  et que j'attribue à cette croyance une valeur (comprise entre 0 et 1) proche de 1, la plausibilité de l'événement contraire peut être sensiblement plus importante dans mon esprit que le complément à 1 de ma croyance en  $e$ .

Une autre faiblesse du calcul probabiliste, montré par (PRADE 84), surgit lorsque l'on ne sait rien *a priori* de la véracité d'une proposition  $e$ . On est par conséquent conduit à prendre  $P(e) = P(\text{non } e) = 0,5$ .

Pour contourner les difficultés présentées par le calcul des probabilités, on assiste, depuis le début des années 70, à diverses propositions de modélisation du raisonnement approximatif. Le modèle mathématique de SHORTLIFFE et BUCHANAN (SHORTLIFFE 75) en constitue, à travers le système MYCIN, un exemple des plus marquants.

Les propriétés communes et minimales de chacun de ces modèles sont les suivantes:

soit  $m$  une mesure d'incertitude d'un ensemble de propositions  $P$  dans  $[0,1]$ .  $m$  représente la part de croyance accordée à une proposition.

Propriétés:

- 1)  $m(\text{faux}) = 0$
- 2)  $m(\text{vrai}) = 1$
- 3)  $( P \Rightarrow Q ) \Rightarrow ( m(P) \geq m(Q) )$

Ces trois propriétés sont dictées par le simple bon sens, et ne constituent pas, à elles seules, une stratégie de choix d'objet ou

d'action pour un système expert. Il faut alors introduire des outils de modélisation qui permettent de rendre compte des étapes du raisonnement incertain proprement dit, et en particulier de la non-complémentarité à 1 des mesures d'événements incompatibles.

Ceci implique qu'il n'existe pas une mesure unique de l'événement (sa probabilité), mais plusieurs mesures "parallèles et simultanées" dont la combinaison produira, ou concourra à produire des déductions.

C'est ainsi que G. SHAFER a proposé un modèle complet de raisonnement basé sur deux principes : la **crédibilité** et la **plausibilité** d'un événement. On emprunte à (PRADE 83) et (PRADE 84) les notions qui suivent:

- La mesure de base  $m$  proposée par SHAFER satisfait les trois propriétés énoncées haut, plus une quatrième:

$$\sum_{p \in P} m(p) = 1$$

- La mesure de crédibilité, notée CR, basée sur  $m$  s'exprime alors par:

$$\text{pour tout } q \in P, CR(q) = \sum_{p/p \Rightarrow q} m(p)$$

La crédibilité d'une proposition  $q$  s'obtient par la sommation des parts de croyance de toutes les propositions  $P$  qui entraînent  $q$ .

On définit, par dualité, la plausibilité PL de la façon suivante:

$$PL(q) = 1 - CR(\text{non } q) \quad (2)$$

$$\text{équivalent à } : CR(q) = 1 - PL(\text{non } q) \quad (3)$$

$$\text{c'est-à-dire : } PL(q) = \sum_{p/\text{non}(p) \Rightarrow \text{non } q} m(p) \quad (4)$$

## Chapitre I

La plausibilité d'une proposition  $q$  s'obtient par la sommation des parts de croyance dans les propositions qui **ne nient pas**  $q$ .

On vérifie alors que :  $CR(p) + CR(\text{non } p) \leq 1$  (5)

$PL(p) + PL(\text{non } p) \geq 1$  (6)

Dans le cas de l'ignorance totale, on aura :

$$CR(p) = CR(\text{non } p) = 0$$

et  $PL(p) = PL(\text{non } p) = 1$

On déduit de (2) et (5) que  $PL(p) \geq CR(p)$

La plausibilité d'une proposition est au moins égale à sa crédibilité.

La relation (3) satisfait l'intuition qui veut qu'une proposition soit d'autant plus crédible que sa proposition contraire est peu plausible.

Enfin, remarquons que si chaque proposition  $p$  telle que  $m(p) > 0$  est incompatible avec toute proposition qu'elle n'entraîne pas, les mesures de crédibilité et de plausibilité définies par (1) et (4) sont alors égales à une même mesure de probabilité  $P$ , et (2) devient  $P(p) + P(\text{non } p) = 1$ .

Lorsque les diverses propositions sont "consonantes", c'est-à-dire qu'elles sont liées par des relations d'implication sans boucles :

$$p_n \rightarrow p_{n-1} \rightarrow \dots \rightarrow p_1,$$

on peut alors montrer que :

a)  $CR(p \text{ et } q) = \min (CR(p), CR(q))$

b)  $PL(p \text{ ou } q) = \max (PL(p), PL(q))$

### I.3.2 FONCTIONS DE CROYANCE ET REGLE DE COMBINAISON DE DEMPSTER - LE SYSTEME GENDIAG

La théorie des Fonctions de Croissance énoncée par SHAFER connaît une application en Génétique Médicale, à l'Hôpital d'Enfants de la Timone (MARSEILLE).

Mis au point par J.GOUVERNET en collaboration médicale avec S. AYME (AYME 81), le système GENDIAG (GOUVERNET 85) est un système d'aide au diagnostic des syndromes génétiques à partir d'un thésaurus arborescent de signes et de symptômes et d'une base de syndromes (descriptions diagnostiques). Les auteurs ont mené une réflexion analogue à la nôtre (§ II.1), à savoir que:

- a) le généticien ne peut conserver en mémoire la totalité des symptômes et des syndromes rencontrés en génétique médicale. Leur nombre (plus de 3000 actuellement) s'accroît régulièrement chaque année,
- b) il est hors de question d'adopter un modèle inférentiel basé sur un procédé purement bayésien (imprécision des connaissances) pas plus qu'un modèle de type logique booléenne: le domaine est trop vaste, complexe et incertain pour que l'on puisse espérer construire des règles à partir de la pratique diagnostique des généticiens.

Ce constat étant fait, la solution a consisté à adopter, comme mécanisme d'inférence, les fonctions de croyances combinées par la règle de DEMPSTER.

La présentation de la base théorique (GOUVERNET 79) est la

## Chapitre I

suivante:

considérons  $D$  l'ensemble des diagnostics tels que définis en I.2.4 (un diagnostic est défini par une liste de symptômes).

$$D = \{D_1, \dots, D_n\}$$

Appelons  $H$  un sous-ensemble de  $D$  constituant une **hypothèse**.

$H$  représente un ensemble de diagnostics candidats à l'explication d'une série d'observations de symptômes appelée  $S$ .  $H$  est inclus dans  $D$ , et  $S = \{S_1 \& S_2 \& \dots \& S_k\}$ .

Une **fonction de croyance** ("belief function") est une application  $B(H/S)$  de l'ensemble des parties de  $D$  dans  $[0,1]$  qui, à toute hypothèse  $H$  incluse dans  $D$ , associe un nombre correspondant au degré de croyance accordée à l'hypothèse.

Une fonction de croyance liée à une observation  $S$  obéit aux axiomes suivants:

1° -  $B(\emptyset/S) = 0$

2° -  $B(D/S) = 1$

3° - pour toute hypothèse  $H = \{D_1, \dots, D_r\}$ ,  $r > 0$ , incluse dans  $D$ , alors:

$$B(D_1 \cup D_2 \cup \dots \cup D_r/S) \geq \sum_{I \subset \{1, \dots, r\}} (-1)^{|I|+1} B(\bigcap_{i \in I} D_i/S)$$

Pour définir pratiquement la fonction de croyance, on définit d'abord une loi de probabilité sur l'ensemble des parties de  $D$  soit  $m_S$ , liée à l'observation  $S$ .

On rappelle les axiomes d'une loi de probabilité:

$m_S(\emptyset) = 0$

$\sum m_S(d) = 1$

DED

La fonction de croyance  $B(. / S)$  est définie de la manière suivante:

$$\text{pour tout } D \in \mathcal{D}, B(D/S) = \sum_{C \subset D} m_S(C)$$

récioproquement, si  $B$  est donnée,  $m(S)$  est déduite par:

$$m_S(D) = \sum_{C \subset D} (-1)^{|D-C|} \cdot B(C/S)$$

La combinaison des fonctions de croyance est effectuée au moyen de la règle de DEMPSTER. L'application de cette règle n'est pas soumise à l'hypothèse d'indépendance des symptômes.

$$m_{S1S2}(D) = \frac{\sum_{A_i \cap B_j = D} m_{S1}(A_i) \cdot m_{S2}(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_{S1}(A_i) \cdot m_{S2}(B_j)}$$

où les  $A_i$  et  $B_j$  sont des sous-ensembles de  $D$  tels que

$$m_{S1}(A_i) > 0 \text{ et } m_{S2}(B_j) > 0, \text{ quelques soient } i \text{ et } j.$$

$m_{S1}$  et  $m_{S2}$  sont des probabilités élémentaires appliquées à

$B(. / S1 S2)$  si celle-ci existe, c'est-à-dire si

$$\sum_{A_i \cap B_j = \emptyset} m_{S1}(A_i) \cdot m_{S2}(B_j) < 1$$

La règle de combinaison de DEMPSTER permet donc de déduire  $B(. / S1 S2)$  à partir de  $B(. / S1)$  et  $B(. / S2)$  sans autre connaissance que celle de  $m_{S1}$  et  $m_{S2}$

Mise en oeuvre du calcul:

Dans la liste des symptômes définissant un syndrome  $D$ , chaque symptôme est affecté d'un coefficient fréquentiel prenant ses valeurs dans l'ensemble [ 0, 0,1 , 0,5 , 0,8 , 1 ]

## Chapitre I

Pour l'observation d'un symptôme S, la part de croyance accordée au diagnostic D est calculée de la manière suivante:

si valeur de S                      alors:

dans D égale :

0	$mS(D) = 0,$	$mS(D - D) = 1$
0,1	$mS(D) = 0,6,$	$mS(D - D) = 0,4$
0,5	$mS(D) = 0,7,$	$mS(D - D) = 0,3$
0,8	$mS(D) = 0,8,$	$mS(D - D) = 0,2$
1	$mS(D) = 1 - a,$	$mS(D - D) = a$

Le paramètre a est laissé au choix de l'utilisateur et permet de "régler" le caractère confirmatif de la fréquence égale à 1 d'un symptôme.

Enfin, le Système GENDIAG fonctionne également avec une phase différentielle: l'utilisateur se voit proposer des signes importants apparaissant dans les diagnostics les plus crédibles par rapport à la sémiologie observée. Cette phase de proposition d'observation, après une sélection ordonnée de diagnostics possibles, confère au système une dimension plus "intelligente", au sens de l'I.A., que le simple fait de proposer des maladies comme le ferait un système de Base de Données classique. Néanmoins, le système fonctionne de manière entièrement procédurale, et n'a pas réellement la dimension et la souplesse d'un système expert proprement dit.

Ajoutons enfin que ce système est utilisé "en routine" depuis plus de 6 ans, et qu'il constitue un acte d'informatique médicale assimilable à un acte médical courant de laboratoire.

### I.4 THEORIE DES SOUS-ENSEMBLES FLOUS ET THEORIE DES POSSIBILITES

#### I.4.1 ELEMENTS DE BASE

Nous abordons maintenant une théorie qui semble rompre, dans son esprit sinon dans sa lettre, avec les méthodes numériques plus classiques telles que le calcul des probabilités, la statistique ou l'analyse de données. Comme nous pourrons le constater, cette rupture n'en est pas vraiment une, puisque cette théorie utilise des concepts voisins de ceux utilisés dans ces disciplines. Elle se présente essentiellement comme une formalisation de l'imprécision: imprécision des connaissances, imprécision des faits, imprécision des conclusions.

La théorie des possibilités réalise une synthèse de méthodes et de concepts principalement issus de la théorie des sous-ensembles flous et se définit essentiellement comme un outil de formalisation du raisonnement incertain. Elle fait l'objet actuellement de nombreux travaux (ZADEH 78) (FARRENY 84).

Notre but ici étant de présenter des modèles de décision en relation directe avec le domaine médical, il est bien sûr hors de propos de présenter la théorie dans sa totalité : bien que récente, elle offre en effet un volume d'ouvrages de base et de publications assez important. Je m'appuierai donc, pour une large part, sur des travaux de recherche appliquée en relation avec les systèmes experts. La théorie des sous-ensembles flous occupe une place importante en matière d'aide au raisonnement diagnostique (WECHSLER 76) (ADLASSNIG 80). Parmi les travaux les plus marquants de ce domaine, nous trouvons ceux du C.H.U. de Marseille (SANCHEZ 79) (FIESCHI 83) (SOULA 81) (SOULA 82) (SOULA 83).

Le point de départ de la théorie des sous-ensembles flous (T.S.E.F.) est une nouvelle définition de l'appartenance d'un

élément donnée à un sous-ensemble particulier A d'un ensemble général E, appelé univers du discours. A est dit "sous-ensemble flou" de E s'il existe une fonction  $\mu$  définie sur E, à valeurs dans le sous-ensemble [0,1] de R. Cette fonction, dénommée "fonction d'appartenance", quantifie le degré d'appartenance à A d'un élément de E.

$$\begin{array}{l} \mu(A): E \longrightarrow [0,1] \\ e \longrightarrow \mu(A)(e) \end{array}$$

Cette fonction est analogue à la fonction caractéristique d'un sous-ensemble T de U qui, à tout élément e de U, fait correspondre la valeur un si e appartient à T et la valeur zéro dans le cas contraire.

On peut dire qu'un sous-ensemble flou A a une frontière large, qui n'est pas franchie instantanément, de sorte que tout élément qui se trouve dans cette bande-frontière n'est ni vraiment dans A, ni vraiment hors de A.

L'intérêt d'un tel concept est évident en médecine puisqu'il n'existe souvent pas de frontière précise entre l'état de maladie et l'état de bonne santé. Si E représente l'ensemble des états physio-pathologiques d'un individu DUPOND, on peut concevoir l'expression :  $\mu(\text{grippe})(\text{DUPOND}) = 0,8$ .

De même, la T.S.E.F. permet de définir des **propositions floues** au moyen de leurs **distributions de possibilités**.

Une proposition telle que "le poids est normal" peut s'interpréter comme une variable : le poids, à valeurs dans le référentiel R.

$$R = (\text{faible}, \text{normal}, \text{élevé})$$

## Chapitre I

La notion de normalité s'exprime d'une manière floue par une distribution de possibilités  $\pi \langle XY \rangle$ , où X désigne le poids et Y l'événement "poids normal".  $\pi \langle XY \rangle (e)$  appartient à  $[0,1]$ , où e est à valeurs dans un ensemble E de poids ( $E = [0,200]$ , par exemple).

$\pi \langle XY \rangle (90)$  (figure I.1) peut s'interpréter comme la possibilité que X prenne la valeur 90 sachant Y:

Possibilité  $(X = 90/Y) = 0,3 = \pi \langle Y \rangle (90)$  c'est-à-dire, en d'autres termes: la possibilité pour qu'un poids, considéré comme normal, soit de 90 kgs, est évaluée à 0,3.

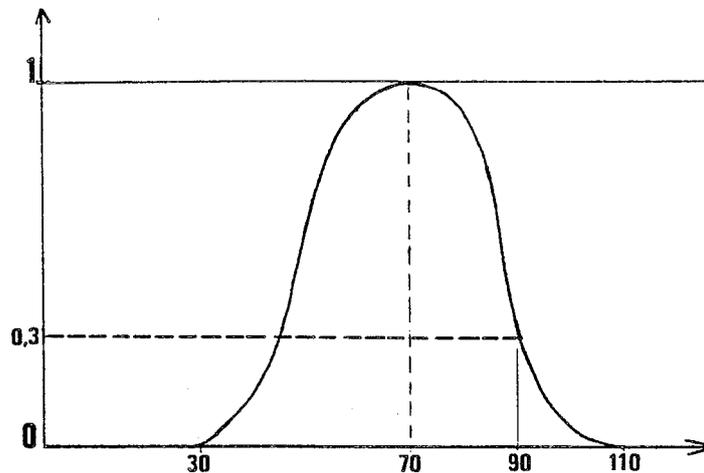


figure I.1

Dès lors, il va être possible de définir des **Propositions Floues de Références** (P.F.R.). Ces P.F.R. sont considérées, comme des connaissances acquises, que l'on ne remet généralement plus en cause. On les utilisera comme propositions de conditionnement dans les distributions de possibilités définies ci-dessus.

On va alors définir une **mesure de possibilité**  $\pi(Z/Y)$  pour qu'une proposition telle que "X est Z" soit vraie étant donné une P.F.R. : "X est Y". Cette mesure traduira le degré de similitude

## Chapitre I

entre la proposition de référence et la proposition à analyser. Elle s'exprime par la formule:

$$\pi(Z/Y) = \sup_{e \in E} (\mu(Z)(e) \cap \mu(Y)(e)) ,$$

dans laquelle la fonction Sup demande une explication.

Prenons l'exemple simple de la température; la P.F.R. : "la température est normale" ("X est Y") se traduit par la Distribution de Possibilités représentée en trait continu sur la figure I.2. Soit une proposition (représentée en pointillés) : la température est de 38° ("X est Z") à tester par rapport à la P.F.R. précédente.

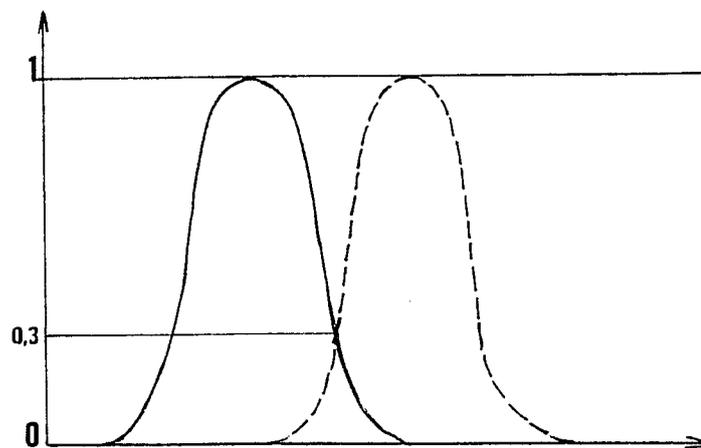


figure I.2

$$\pi(Z/Y) = 0,3$$

En clair, la possibilité floue pour qu'une température de 38° soit considérée comme "normale" est égale à 0,3.

Il est à noter que  $\pi(Z/Y) = \pi(Y/Z)$ .

### I.4.2 INFERENCE EN LOGIQUE FLOUE

Voyons maintenant comment la "logique floue" peut contribuer à la représentation et à l'exploitation de connaissances expertes.

Considérons connu le principe général de fonctionnement d'un système expert classique (CORDIER 84), (LAURENT 84).

Dans un tel système, une condition nécessaire (mais non suffisante) de déclenchement d'une règle en chaînage avant, est que chaque prémisses soit vérifiée par des objets de la base de fait ("pattern-matching"). Toutes les nuances des prémisses doivent être concrétisées dans des règles parfaitement définies, ce qui peut donner lieu, suivant les applications, à une pléthore de "petites" règles. En outre, la formalisation de l'incertitude de l'expert quant à une règle ou à une prémisses reste un problème mal résolu.

Il existe plusieurs méthodes de raisonnement en logique floue, (LESMO 82) nous en étudierons deux: celle des règles d'inférence floue et celle du Modus Ponens Etendu et du Modus Tollens Etendu.

#### I.4.2.1 EXEMPLE DE PROTIS

Nous retiendrons pour cet exposé une méthode qui a été testée et validée dans le système expert PROTIS (SOULA 83).

Dans ce système, une règle floue prend la forme suivante:

SI	condition	ALORS	évocation (C, e)
		SINON	rejet (C, r)
		LIEN	(S)

La condition est une combinaison de plusieurs distributions de possibilités représentant des propositions floues.

C représente la conclusion

e est une valeur de pondération de la conclusion, comprise

## Chapitre I

entre 0 et 1.

$r$  est une valeur de pondération de la non-conclusion (rejet), ou disconfirmation, comprise également entre 0 et 1.

Les valeurs  $e$  et  $r$  sont données par l'expert. Si la condition est satisfaite, la conclusion sera d'autant plus affirmée que la valeur  $e$  sera proche de 1. À l'inverse, si la condition est insatisfaite, la conclusion sera rejetée à son tour avec une force d'autant plus grande que  $r$  sera proche de 1. La conclusion  $c$  est totalement rejetée ( $r=1$ ) lorsque la condition est totalement insatisfaite.

Quel est alors le rôle des liens ? Un lien représente un moyen d'ajuster la connaissance contenue dans la règle (condition, évocation, rejet) au contexte dans lequel elle intervient.

Un lien a lui-même une structure de règle floue :

SI condition ALORS action.

La condition est représentée par une série de distributions de possibilités.

L'action pose deux questions :

a) quel élément (distribution de possibilités, pouvoir d'évocation ou pouvoir de rejet de la conclusion) doit-elle affecter ?

b) quelle action (accroissement ou abaissement, décalage à gauche, décalage à droite, amplitude, etc) doit-elle exécuter ?

Prenons l'exemple d'une règle floue portant sur la glycémie. Chez une personne âgée, la fourchette de la glycémie normale est

décalée vers la droite. Le lien d'une telle règle aura donc pour condition: "la personne est âgée", et pour conclusion, en clair: "décaler de 0,8 g/l vers la droite les valeurs de la glycémie".

Le moteur d'exploitation de telles règles fonctionne suivant un principe d'associations floues (Fuzzy Pattern Matching) entre faits et conclusions.

Considérons un fait : "X et F" défini par une distribution de probabilités, et une condition : "X et C" définie également par une distribution de probabilités.

L'association du fait et de la condition s'établit à partir du calcul de deux valeurs : la possibilité (POS) et la nécessité (NEC) de cette association, définies de la manière suivante:

$$POS(F/C) = \pi(F/C)$$

$$NEC(F/C) = 1 - \pi(F / \text{non } C)$$

$$\text{ou encore : } POS(F/C) = \text{Sup}_{e \in E} ( \mu(F)(e) \cap \mu(C)(e) )$$

$$NEC(F/C) = 1 - \text{Sup}_{e \in E} ( \mu(F)(e) \cap \mu(\text{non } C)(e) )$$

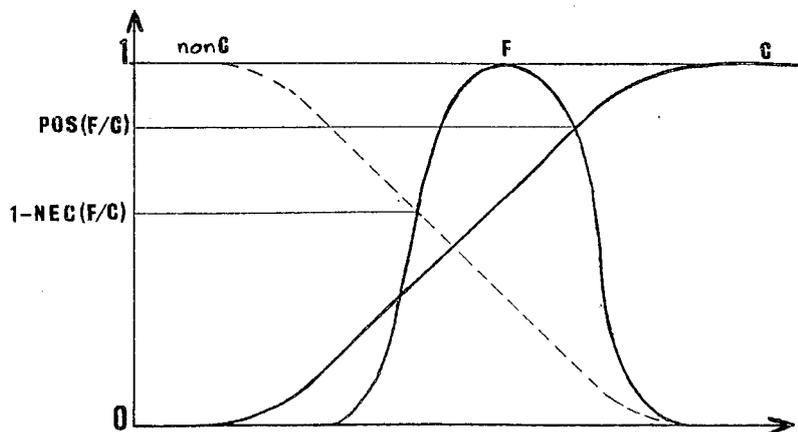


figure I.3

## Chapitre I

On vérifie graphiquement les deux implications suivantes:

$$\text{NEC}(F/C) = 1 \quad ==> \quad \text{POS}(F/C) = 1$$

$$\text{POS}(F/C) = 0 \quad ==> \quad \text{NEC}(F/C) = 0$$

Par contre, si  $\text{POS}(F/C)$  est égal à 1,  $\text{NEC}(F/C)$  peut parfaitement prendre une valeur non nulle.

L'association entre le fait  $F$  et la conclusion  $C$  sera d'autant plus forte que  $\text{NEC}$  et  $\text{POS}$  seront proches de un, et d'autant plus faible que  $\text{NEC}$  et  $\text{POS}$  seront proches de zéro. Entre les deux, l'association est "floue".

Les valeurs  $\text{POS}$  et  $\text{NEC}$  ne déterminent pas à eux seuls l'acceptation ou le rejet de la conclusion, mais n'en sont que des facteurs. Il faut tenir compte en effet des pondérations d'acceptation ( $e$ ) et de rejet ( $r$ ) fournies par l'expert.

La pondération finale d'acceptation sera:  $Ac = \text{MAX}(\text{NEC} + e - 1, 1)$   
La pondération finale de rejet sera:  $Re = \text{MIN}(\text{POS} + 1 - r, 1)$

$Ac$  et  $Re$  appartiennent tous deux à l'intervalle  $[0,1]$  de  $\mathbb{R}$ .

On remarque que:

$$\begin{aligned} Ac &= 1 \text{ lorsque } \text{NEC} = e = 1, \\ Re &= 0 \text{ lorsque } \text{POS} = 0 \text{ et } r = 1. \end{aligned}$$

Un chaînage avant de règles floues donne lieu à une cascade de couples  $(Ac, Re)$ :

$$\begin{array}{lcl} \text{FR1} & \text{-----}> & (Ac1, Re1) \\ \text{FR2} & \text{-----}> & (Ac2, Re2) \\ & \vdots & \\ & \vdots & \\ \text{FRn} & \text{-----}> & (Acn, Ren) \end{array}$$

Une décision  $d$ , issue de cette cascade de  $n$  règles, est

assortie d'un intervalle (Ac<sub>d</sub>, Re<sub>d</sub>), obtenu à partir des couples précédents:

$$Ac_d = \text{MAX}_{i \in [1,n]} Ac_i$$

$$Re_d = \text{MIN}_{i \in [1,n]} Re_i$$

Cet intervalle fournit une évaluation de  $d$ , et s'interprète de la façon suivante:

Si  $Ac_d$  et  $Re_d$  sont tous deux proches de un, la décision est fortement proposée. Si, à l'inverse,  $Ac_d$  et  $Re_d$  sont proches de zéro,  $d$  est faiblement proposée. Les résultats seront d'autant plus précis que  $Ac_d$  et  $Re_d$  seront proches.

L'ensemble complet des règles activées constitue une arborescence dont les faits initiaux constituent la racine, et les décisions sont les feuilles. Le système fournit la liste, classée par ordre décroissant des valeurs  $(Ac+Re)/2$  des décisions, depuis la plus fortement suggérée, jusqu'à la plus faiblement suggérée.

#### 1.4.2.2 EXEMPLE DE SPHINX

Le système expert SPHYNX développé par M.FIESCHI (FIESCHI 83) (FIESCHI 84 a) est un système très complet qui assemble judicieusement et tire le meilleur parti d'un certain nombre de concepts de l'Intelligence Artificielle tels que schémas de Minsky, règles de production, logique floue, logique du premier ordre,...

La logique floue et la théorie des possibilités sont utilisées pour le mécanisme d'évocation d'un contexte de maladie. Ce mécanisme est déclenché une seule fois, au début d'une session de

travail, immédiatement après l'entrée du tableau symptomatologique évident du malade. Le contexte de la maladie est alors déduit au moyen de **règles d'évocation** mettant en oeuvre un module de "**pattern matching**" flou.

Ce module, utilisé une seule fois, a pour but d'initialiser un **agenda** d'hypothèses.

Un **contexte**, dans SPHINX, est défini par un ensemble de diagnostics connus du système. Ces diagnostics sont exprimés par des **prototypes** d'entités, c'est-à-dire des instanciations prédéfinies de schémas.

Les concepts utilisés par le discours médical sont en effet structurés selon des schémas: chaque concept est défini par un certain nombre d'attributs, lesquels attributs sont soit décomposés en une nouvelle série d'attributs, soit affectés d'une valeur qualitative prise dans la **catégorie sémantique** propre à chaque attribut.

Par exemple : l'entité APPETIT est qualifiée par l'attribut MODIFICATION selon le schéma suivant:

```
APPETIT
  MODIFICATION
    AUGMENTE
      TRES AUGMENTE
      PEU AUGMENTE
    NORMAL
    DIMINUE
      TRES DIMINUE
      PEU DIMINUE
```

La relation liant un niveau arborescent au niveau inférieur (situé à droite) est appelée FILS DE. Ainsi, nous avons:

PEU AUGMENTE = FILS DE (AUGMENTE).

Nous nous trouvons dans un univers entièrement qualitatif (non numérique). Le problème consiste alors à calculer un **degré de conformité sémantique** entre une observation, c'est-à-dire une variable affectée d'une valeur, et un attribut situé dans un prototype appartenant au contexte sélectionné par le "pattern matching" flou.

La connaissance mise en jeu dans la phase d'évocation fait appel à un raisonnement approximatif d'évocation.

M. FIESCHI s'est inspiré des principes d'inférence proposés par ZADEH (ZADEH 75a) et SOULA (SOULA 81), en les adaptant à des variables prenant leurs valeurs dans des champs purement linguistiques.

D'une manière plus formelle, étant donné un prototype E ( $V_1, V_2, \dots, V_k$ ), les  $V_i$  désignant les valeurs respectives des attributs de E, et la donnée du patient P ( $O_1, O_2, \dots, O_k$ ) désignant l'ensemble des valeurs observées, le degré  $t$  de conformité sémantique entre E et P est calculé selon l'algorithme suivant:

$t = 1$  si  $V_i = O_i$  c'est-à-dire si les valeurs sont égales ou telles que  $O_i = \text{FILS DE } V_i$

$t = 1 - \frac{1}{(n+1)^2}$  si  $V_i = \text{FILS DE } O_i$

$t = 1 - \frac{1}{n+1}$  dans les autres cas.

$n$  désigne le niveau, dans l'arborescence, du prédécesseur commun à

## Chapitre I

$V_i$  et  $O_i$ .

Pratiquement, dans l'exemple de l'appétit,  $n$  prend ses valeurs dans l'ensemble  $\{0,1,2\}$  et  $t$ , après calcul, dans l'ensemble  $\{0, 0.5, 0.75, 1\}$

exemples: \* si  $O_i =$  "peu augmenté" et  $V_i =$  "très augmenté"  
alors  $n = 1 \rightarrow "t = 0.5"$   
\* si  $O_i =$  "diminué" et  $V_i =$  très "diminué"  
alors  $n = 1 \rightarrow "t = 0.75"$

Chaque prototype est affecté, par l'expert, d'un **pouvoir évocateur**  $e$  du contexte dans lequel il se trouve ( $0 \leq e \leq 1$ ), ainsi que d'un **pouvoir de rejet**  $r$  de ce même contexte ( $0 \leq r \leq 1$ ):

Les formules arithmétiques suivantes donnent, en fonction de  $t$  et de  $e$  ou  $r$ :

- la **force d'évocation**  $\alpha$  du contexte à partir d'une observation:

$$\alpha = \text{Max} (t + e - 1, 0) \quad 0 \leq \alpha \leq 1$$

- la **force de rejet**  $\beta$  du contexte à partir d'une observation:

$$\beta = \text{Min} (t + 1 - r, 1) \quad 0 \leq \beta \leq 1$$

Le choix du contexte est effectué en tenant compte de l'étendue et de la valeur des bornes de tous les intervalles  $[\alpha_{ij}, 1 - \beta_{ij}]$  obtenus à partir des entités observées  $i$  et des contextes  $j$ .

Le calcul de l'intervalle propre  $[a_j, b_j]$  du contexte  $j$  résulte du calcul suivant:

$$a_j = \text{Max}_i \alpha_{ij} \quad \text{et} \quad b_j = \text{Min}_i \beta_{ij}.$$

Le meilleur contexte sera celui qui maximise la valeur médiane:

$$\frac{a_j + b_j}{2}$$

2

S'il existe plusieurs valeurs médianes égales, le choix se porte sur le contexte qui a la borne inférieure (pouvoir d'évocation) la plus haute.

## I.5 BASES DE DONNEES - BASES DE CONNAISSANCES

### I.5.1 INTRODUCTION

Délaissées, voire ignorées par le récent essor des systèmes experts, les méthodes d'aide au diagnostic basées sur les principes de la logique documentaire demeurent les seules à pouvoir répondre à un certain nombre de critères inhérents à la pratique du diagnostic médical. Ces critères, voisins de ceux énoncés par les Pr. LENOIR (LENOIR 80) ou H.R WARNER (WARNER 70), nous semblent être les suivants:

a) **l'étendue**: une véritable intégration de l'outil informatique dans le domaine du diagnostic médical est subordonnée à l'étendue des connaissances couvertes par cet outil. L'idée ici est celle de la non-spécificité de l'outil à un champ particulier et restreint de la médecine, ce qui n'est pas le cas actuellement pour la grande majorité des systèmes existants.

b) **l'ouverture** : l'utilisateur de l'outil d'aide au diagnostic ne doit en aucun cas se sentir limité, retenu ou contraint par la machine dans son mode de raisonnement.

c) **la convivialité** : l'outil doit être le plus possible à la portée du praticien tant sur le plan technique (un système d'interrogation sur le lieu de travail) que sur le plan procédural (un système utilisable sans formation particulière si ce n'est une heure au maximum d'initiation à la pratique de l'outil).

d) **la prégnance** : l'outil doit assister le médecin dans un large éventail de services centrés sur la connaissance médicale: le diagnostic, le pronostic, la thérapeutique, les interactions médicamenteuses, etc. Il est par exemple nécessaire qu'un système d'aide au diagnostic permette une information du médecin sur les récentes évolutions de la connaissance médicale à l'endroit de telle ou telle maladie ou technique thérapeutique. On peut opposer à ce critère l'objection selon laquelle le problème du diagnostic est un problème spécifique, différent de celui de la formation ou de l'information du médecin. Ce serait là opérer une dichotomie que le médecin lui-même n'effectue pas dans son esprit: la nécessité d'apprendre et de comprendre est indissociable de celle de soigner.

Il existe un certain nombre de tentatives réussies de thésaurisations sectorielles de la connaissance médicale sous forme de bases de données, ainsi qu'un grand nombre de systèmes experts développés sur un champ restreint de la médecine; il existe d'autre part plusieurs tentatives de nomenclatures universelles (SNOP 65), (SNOMED 75), (ICD) (FRUTIGER 77) des signes, symptômes, maladies ; il existe enfin de vastes bases de données documentaires et bibliographiques (MEDLINE, BIAM, MEDITEX, etc.). Il n'existe pas vraiment, sinon sous forme embryonnaire, de système complet rassemblant les possibilités ci-dessus.

La conséquence de la mise en application de ces critères serait

la création d'un puissant outil de médecine informatique capable aussi bien d'engranger une grande partie de la connaissance médicale, que de la mettre en oeuvre dans l'aide à la décision, et la rendre disponible, telle une encyclopédie, au plus grand nombre de médecins. Ce qui permettrait enfin de l'utiliser comme aide à l'enseignement (E.A.O) et d'en faire un outil de recherche médicale.

On aura compris, par l'exposé qui précède, que le problème de l'aide au diagnostic n'est pas le seul apanage des systèmes experts. Plus précisément, il serait faux de considérer que le seul moyen pour réaliser une bonne aide au diagnostic serait d'abord de réaliser un bon "moteur" au sens large (qu'il soit non algorithmique, probayésien ou autre) puis d'appliquer ce principe inférentiel à une base de connaissances. La bonne vieille méthode de recherche documentaire par simple coïncidence entre une observation et une description, peut encore constituer la meilleure démarche de base, si l'on veut être sûr de ne déclarer que des assertions vraies, quitte éventuellement à les trouver trop imprécises ou surabondantes.

La recherche de diagnostic par juxtaposition, comparaison et recherche de coïncidences constitue plus une méthode prospective qu'inférentielle proprement dite: à partir d'une observation, l'ordinateur fournit tous les scénari possibles, du plus vraisemblable au moins vraisemblable, selon la norme triviale qui consiste à "compter les points" c'est-à-dire, en l'occurrence, le nombre de coïncidences. Etant donné qu'un symptôme peut, dans la plupart des cas, être rangé dans l'une des catégories {introuvable, rare, régulier, fréquent, constant} ou dans un sous-ensemble de ces

## Chapitre I

catégories, vis-à-vis d'une maladie donnée, la coïncidence entre l'observation et la maladie sera plus ou moins exacte. La maladie la plus vraisemblable sera par conséquent celle qui aura le plus de symptômes spécifiés "toujours présent" ou "fréquent", parmi les symptômes observés.

Cette étape de recherche constitue le degré zéro de l'inférence, qui consiste simplement à produire l'éventail des maladies possibles à partir du cas traité.

L'accessibilité des informations offertes par le système de banque de données permet une exploitation de la connaissance non seulement en tant qu'aide à la décision, mais également en tant qu'aide à l'enseignement, recherche bibliographique, consultation de type encyclopédique, accessibilité multiutilisateurs, etc.

Nous allons voir plus concrètement, dans le système qui va être présenté maintenant, les possibilités et les perspectives offertes par une telle approche.

### I.5.2 LA BASE A.D.M.

La base A.D.M. (Aide au Diagnostic Médical) {LENOIR 80 a} {LENOIR 80 b} {LENOIR 80 c} {LENOIR 81 b}, constitue une tentative de grande envergure de mise en mémoire de la connaissance médicale et de mise au point de services centrés sur cette connaissance. Développée depuis 1972 dans le service du Professeur P. LENOIR (CHU de RENNES), elle fournit un remarquable exemple de base de données multifonctions et multi-usages telle que nous venons d'en tracer le profil général.

## Chapitre I

La cible fondamentale de ce travail est le médecin praticien auquel elle offre, à son stade actuel de développement, cinq grandes fonctions:

- 1 - l'aide à l'évocation de maladies devant une séméiologie donnée, observée chez un malade,
- 2 - l'aide documentaire par la description des agents pathogènes,
- 3 - l'aide au diagnostic différentiel,
- 4 - l'aide à la surveillance thérapeutique,
- 5 - la reprise d'interrogatoires mis en mémoire.

La base fournit:

- un thésaurus de 60 000 entités médicales élémentaires (descriptions): maladies, syndromes, signes, symptômes, etc.
- un ensemble de descriptions de maladies et de syndromes au nombre de 6 000 actuellement (4 000 maladies, 1 300 syndromes, 700 formes cliniques), en accroissement d'un millier environ par an ;
- un interface de communication en langage clair doté d'un thésaurus de 100 000 mots, expressions composées, flexions grammaticales etc., ainsi qu'un dispositif de reconnaissance d'erreurs de frappe.

Les liens entre les divers descripteurs de la base sont:

- l'identité (synonymie, 1 500 identités)
- l'inclusion : thésaurus sous forme d'un réseau de poly-hiérarchies (2,5 millions d'inclusions)
- la proximité (homonymies, facettes sémantiques)

## Chapitre I

- le commentaire.

Les champs couverts par la base, avec leur pourcentage approximatif de recouvrement, sont les suivants:

- CARDIOLOGIE/ANGEIOLOGIE	90 %	HEMATO/IMMUNO	60 %
- APPAREIL LOCOMOTEUR	90 %	NEPHROLOGIE	60 %
- DIGESTIF PLEIN	80 %	ENDOCRINO/NUTRITION	60 %
- DIGESTIF CREUX	80 %	TOXICO/MAL. PROF.	60 %
- GENETIQUE	80 %	ORL	50 %
- MALADIES IATROGENES	80 %	DERMATOLOGIE	30 %
- MALADIES INFECTIEUSES	75 %	PEDIATRIE	30 %
- PNEUMOLOGIE	75 %	OPHTALMOLOGIE	15 %
- GYNECOLOGIE/ANDROLOGIE	70 %	PSYCHIATRIE	15 %
- NEUROLOGIE PERIPHERIQUE	70 %	NEUROLOGIE CENTRALE	5 %
- UROLOGIE	70 %		

(d'après les informations recueillies auprès du Pr. P.LENOIR)

Le mode principal d'acquisition des données médicales est constitué par des thèses de doctorat de médecine, au nombre actuel d'environ 150.

Implantée sur un gros ordinateur du CHU de Rennes, la base ADM est accessible à partir du lieu de travail du médecin au moyen de terminaux classiques et de minitels.

La base, actuellement dans sa 12ème version, augmente au cours de son évolution le nombre de ses potentialités.

Elle est structurée en 4 modules de base:

- une banque médicale textuelle
- un interface de communication
- un module d'aide au diagnostic
- un module d'E.A.O.

Elle offre une somme importante de renseignements utiles à la recherche médicale (LENOIR 81 a), et développe un service d'aide à l'enseignement : l'AEDM (PEDRONO R. 79) (CHALES 81). L'ouverture

sur les techniques de l'intelligence artificielle et les systèmes experts est également présente (PEDRONO M.J. 81).

### I.5.3 ELEMENTS DE PROSPECTIVE

Les systèmes experts tels qu'ils sont connus et développés depuis quelques années constituent très probablement une propédeutique à de futurs gros systèmes intégrés centrés sur la connaissance.

Un mouvement est en train de se dessiner actuellement ,qui tente d'associer deux grands courants d'investigation informatique: les Systèmes de Gestion de Base de Données (S.G.B.D.) et l'Intelligence Artificielle (I.A.) (BENSAÏD 86). Citons en exemple le système PATREC (MITTAL 84) qui se présente comme une Base de Données dirigée par la connaissance et orientée vers un Système Expert d'Aide au Diagnostic.

Il est fort probable que nous assisterons, dans les années qui viennent, à une certaine fusion de ces deux domaines à partir de laquelle devraient émerger des Systèmes de Gestion de Bases de Connaissances (S.G.B.C.).

Ces S.G.B.C. intégreront, à l'ensemble des outils d'exploitation des bases de données, des outils et des concepts issus de travaux sur l'apprentissage, la modélisation des processus cognitifs, le raisonnement incertain, les logiques non-monotones, etc.

Cette relation entre S.G.B.D. et I.A. est non seulement rendue nécessaire par le besoin d'extraire de nouvelles connaissances à partir de bases de données factuelles, mais elle reflète également

## Chapitre I

la relation qui existe dans le cerveau humain entre la fonction de mémorisation pure et celle de déduction.

D'une façon assez pragmatique, nous pouvons reprendre le tableau proposé par GARDARIN (GARDARIN 86) montrant les coïncidences et différences que l'on peut apporter entre les actuels systèmes experts, et les futurs S.G.B.C.

CONCEPTS	SYSTEMES EXPERTS	S.G.B.C.
	3	6
nombre de faits	10	10
	3	6
nombre de règles	10	10
contraintes d'intégrité	-	+
méthodes d'accès	-	+
indépendance règles/faits	+	+
explications	+	?
langages algorithmiques	-	+
langages de requêtes	-	+
langages naturels	+	?
partage multi-utilisateurs	-	+
protections	-	+
sécurité	-	+
inférences	+	+
calcul de possibilités	+	?
heuristiques	+	?
interactivité	+	+

### I.6 ELEMENTS DE COMPARAISON ET DE DISCUSSION

### I.6.1 DISCUSSION DES METHODES PROPOSEES

#### I.6.1.1 A PROPOS DE LA FORMULE DE BAYES

La formule de Bayes a un aspect séduisant, en ce qu'elle constitue un véritable petit "moteur" d'entraînement d'un raisonnement; ce raisonnement propose des choix (les probabilités **a posteriori**) à partir de faits (les observations de symptômes, résultats d'examens, etc). E.H. SHORTLIFFE (SHORTLIFFE 76, p. 163 ) fait, à juste titre, remarquer que l'expression de la probabilité conditionnelle  $P(D_i/S_k) = x$  peut être considérée comme une règle de production:

Si : le patient présente le symptôme  $S_k$   
Alors : conclure que le patient a la maladie  $D_i$  avec la probabilité  $x$ .

Cette qualité n'est cependant pas déterminante, car la formule de Bayes ne peut s'appliquer que dans un cadre d'hypothèses rigoureuses, généralement très difficile à atteindre en pratique.

En effet, il existe deux niveaux de restrictions pour son application: un niveau théorique et un niveau pratique. En ce qui concerne le niveau théorique, les événements "maladies" doivent être indépendants, c'est-à-dire qu'un patient ne peut pas être atteint de deux maladies simultanément, ce qui de toute évidence est contredit par la réalité ; d'autre part, ce qui est plus gênant encore, les symptômes doivent être **indépendants** pour permettre un calcul aisé des probabilités conditionnelles conjointes du type de celles rencontrées plus haut:  $p(C/D_i)$ , où  $C=S_1 \& S_2 \& \dots \& S_m$ . En effet, la connaissance de la pathologie n'est jamais telle que des probabilités aussi fines que ces  $p(C/D_i)$  puissent être évaluées

directement. Le calcul le plus simple consiste en fait à supposer que les  $S_i$  sont indépendants (§ I.2.5) conditionnellement à la maladie  $D_i$ , c'est-à-dire que :

$$P(C/D_i) = \prod_{j=1}^m p(S_j/D_i)$$

Ces problèmes pourraient être résolus théoriquement en "créant" de nouvelles maladies qui seraient des combinaisons de plusieurs maladies, ainsi que de nouveaux symptômes, composés de plusieurs symptômes unitaires. Mais ces efforts théoriques ne sauraient supprimer l'obstacle de la mise en pratique : second niveau de restriction. Il existe néanmoins diverses manières de contourner ce problème de la dépendance des symptômes (NORUSIS 75) (CUMBERBATCH 76) (FRYBACK 78).

La mise en pratique de la Formule de Bayes dans des conditions expérimentales acceptables est extrêmement difficile à réaliser en raison du nombre prohibitif de données cliniques et épidémiologiques nécessaires pour établir les prévalences des maladies (probabilités expérimentales) ainsi que les probabilités d'occurrence des symptômes dans chacune des maladies.

Le système HELP (§ I.2.1.2) constitue l'une des rares (sinon la seule) applications de grande envergure, dans de bonnes conditions, de l'algorithme décisionnel bayésien. Ce système marque une étape dans une direction importante, encore peu développée malgré les moyens actuels : celle de la mise à disposition des praticiens d'un centre hospitalo-universitaire d'un outil décisionnel dépendant de ce centre, source inépuisable de connaissances inférentielles à proposer à toute la communauté

## Chapitre I

médicale.

Enfin, il ne faut pas oublier que le système bayésien est totalement dépourvu de moyens de décision quant au choix des meilleurs symptômes à proposer à l'observation entre deux "cycles" de calcul de probabilité. Le nombre de possibilités de symptômes à rechercher, de tests cliniques à effectuer, d'examens biologiques à demander pour déceler le symptôme le plus pertinent pour la recherche d'un diagnostic est en général très élevé (SZOLOVITS 78).

Pour corser la difficulté, il faut savoir que les prévalences des maladies ne sont pas stables d'une région à une autre; elles dépendent du lieu géographique et du type de population concernée. Si bien qu'une étude statistique, aussi fine soit-elle, ne peut nullement être extrapolée à d'autres populations ou à d'autres climats.

Si cette formule est si délicate à appliquer, on peut se demander alors pourquoi elle est autant utilisée. Deux raisons peuvent être invoquées à ce sujet. Dans un secteur de la connaissance médicale très localisé ou dans un lieu géographique (une population) bien déterminé, le théorème de Bayes peut conduire à des résultats relativement fiables (SALAMON 76). A un niveau de recherche théorique sur les méthodes de raisonnement artificiel, le théorème de Bayes est source d'inspiration (DUDA 76) (SZOLOVITS 78) (CHEESEMAN 83). Bien que d'emploi pratique limité, il constitue néanmoins un modèle rigoureux de raisonnement mathématique.

On trouvera une intéressante étude comparative entre l'inférence bayésienne et l'inférence par règle de production dans (FOX 80).

### I.6.1.2 A PROPOS DE LA LOGIQUE FLOUE

Nous avons survolé (§ I.4) la théorie des sous-ensembles flous, et nous avons vu quelques-unes de ses contributions au vaste problème du raisonnement incertain, lot par excellence du domaine médical. Nous renvoyons le lecteur, pour une étude et des informations plus approfondies, aux ouvrages et publications fondamentaux en la matière, tels que: {ZADEH 75 a}, {KAUFMAN 75}.

Ce bref exposé appelle néanmoins une série de remarques. On peut tout d'abord se pencher sur la question du "flou" et se demander en quoi consiste réellement la modélisation de ce flou, incontestable dans les raisonnements incomplets ou incertains. Le flou des concepts médicaux tels "la température est élevée", "hypoglycémie" ou "personne âgée" semble bien modélisé par des distributions de possibilités de forme ou gaussienne. De même, l'association floue entre faits et conditions satisfait le bon sens naturel. Mais on peut se demander sur quelles réalités ou sur quelles expérimentations sont fondées les D.P. N'y a-t-il pas un "flou" qui échappe totalement à la théorie en ce qui concerne par exemple le degré d'aplatissement de la courbe en cloche d'un fait ? S'il ne s'agit là que d'une remarque, ce qui paraît en revanche beaucoup plus important, c'est l'usage à maintes reprises de coefficients numériques soit calculés, telles les pondérations d'acceptation (A) et de rejet (R) d'une conclusion, soit demandés à l'expert, telles les valeurs de pondération  $e$  et  $r$  associées à chaque règle.

La logique floue semble préférer le traitement des connaissances numériques au traitement des connaissances non-

numériques. Comme le fait remarquer M. FIESCHI (FIESCHI 83, p. 85), les divers exemples d'application de la logique floue au raisonnement approximatif portent sur des variables sémantiques dont l'univers du discours est numérique, condition *sine qua non* pour définir rigoureusement une fonction d'appartenance.

Nous allons, pour démontrer cette idée, reprendre les figures de G. SOULA (SOULA 83) à propos de la proposition "la glycémie est augmentée". La représentation floue de cette proposition est :

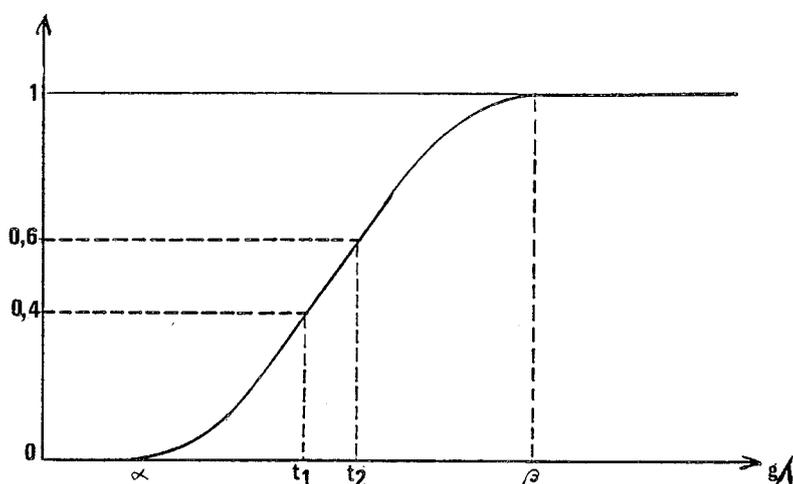


figure I.4

L'interprétation en est la suivante: étant donné la proposition "la glycémie est augmentée", les possibilités que cette glycémie prenne les valeurs  $t_1$  ou  $t_2$  sont respectivement 0,4 et 0,6. Il existe naturellement deux autres cas de figures correspondant aux concepts de "diminution" et "d'intervalle". Les trois cas de figures peuvent s'interpréter comme étant une analogie, en logique floue, des opérateurs de comparaison de l'algèbre classique:  $>$  flou,  $<$  flou et intervalle flou (une extension de la notion d'égalité floue).

Nous obtenons donc les trois schémas suivants (figure I.5) :

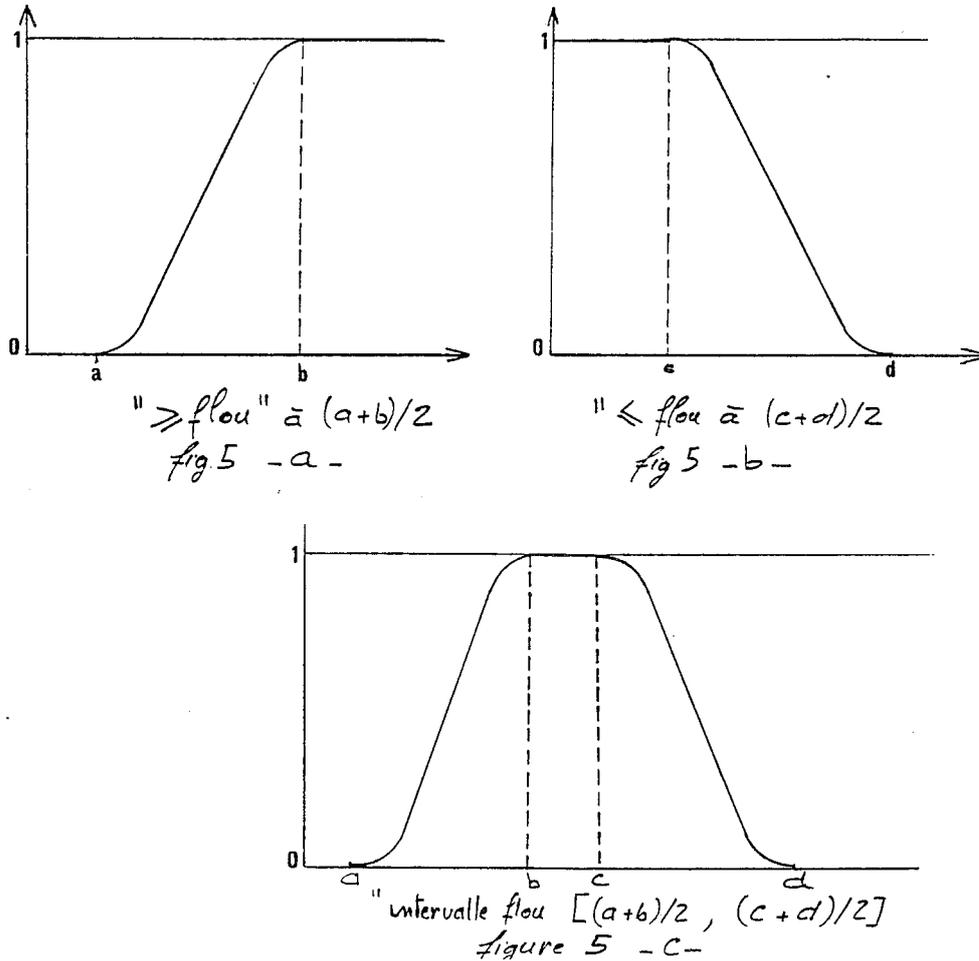


figure I.5

Pratiquement, l'expert se voit demander pour chacune de ces propositions floues, les valeurs (au nombre de 2 ou de 4) des frontières floues.

Une telle démarche, parfaitement cohérente avec des variables numériques, semble plus difficile à appliquer à des variables sémantiques qualitatives, présentant un caractère également flou, telle la variable "teinte de la peau", dont les valeurs proviennent de diverses causes génétiques.

On peut donc, sans équivoque, affirmer que certains domaines médicaux (le traitement du diabète par exemple) se prêtent beaucoup mieux à une modélisation en logique floue que d'autres.

Il faut noter cependant que ZADEH (ZADEH 75 b) a élaboré un certain nombre d'idées quant à l'application de la T.S.E.F. au traitement des variables sémantiques qualitatives.

### **I.6.1.3 REMARQUES SUR LES COEFFICIENTS NUMERIQUES**

Ici se pose à nouveau la question cruciale du sens réel des coefficients numériques fournis par l'expert. Cette question, déjà posée par MYCIN ou LITHO (BONNET 82), resurgit également dans la théorie des sous-ensembles flous ou la théorie des possibilités. Elle semble parfaitement étanche à tout apport nouveau de la part de ces théories, et marque une fois de plus la limite des modèles proposés. L'incertitude reçoit une tentative de modélisation au niveau des faits ou des concepts de base, ainsi qu'au niveau du moteur d'exploitation des objets de la connaissance (faits et règles), mais l'incertitude au sujet des règles de production, principal point d'achoppement de la représentation des connaissances incertaines, reste totalement entre les mains de l'expert. Il ne s'agit pas, insistons sur ce point, d'enlever à l'expert toute possibilité d'associer une notion d'incertitude à chacune de ses règles, mais au contraire de lui proposer un autre modèle que la simple pondération entre deux bornes (0 et 1 par exemple) dictées par le système. Car de tels systèmes présentent deux failles majeures:

a) La première faille consiste en ce qu'ils proposent une solution contraire au code de pensée naturel. Lorsqu'on émet un syllogisme quelconque ("Tous les hommes sont mortels, or Socrate est un homme, donc Socrate est mortel"), il n'est pas naturel de lui adjoindre

automatiquement un coefficient (0.3) de pondération. Une telle méthode conduit forcément l'expert à être en désaccord avec lui-même : quel moyen a-t-il de faire la différence entre 0,66 et 0,75 pour pondérer une connaissance incertaine ? On joue ici sur deux tableaux antinomiques : l'incertitude en relation avec l'intuition, et la rigueur des chiffres. Citons M. FIESCHI à ce sujet (FIESCHI 83, p 86):

"Cette façon d'exprimer d'une manière numériquement pondérée ce que l'on ressent (douleurs,...) ou ce que l'on juge (appétit diminué) n'est pas naturelle. Elle est, nous semble-t-il, de nature à introduire une "fausse" quantification du fait et, des études l'ont montré, on n'utilise pas un segment continu pour exprimer de tels jugements, mais plutôt une échelle de 3 à 5 valeurs; d'autre part, la notion d'"augmenté" ou "diminué" étant par nature linguistiquement floue, il est très difficile d'apprécier la signification du coefficient de pondération donné par l'utilisateur.

b) La seconde faille de ce type de système consiste en ce qu'ils transforment la base de connaissances en un immense tableau de bord présentant autant de boutons (ou couples de boutons) que de règles. On sait que le système commence à être opérationnel lorsqu'on a réglé tous ces boutons entre eux de manière à les rendre cohérents, sinon les enchaînements de mauvais réglages conduisent à des conclusions inattendues et erronées.

Il n'est pas juste, en outre, d'affirmer que, dans un tel système, des connaissances peuvent être ajoutées "en vrac" à la base de connaissances (thème important des S.E.). L'ajout d'une règle est au contraire délicat en raison de la perturbation que celle-ci

introduit dans l'arbre des états possibles, et du retentissement qu'elle a sur des règles voisines.

On peut regretter, par ailleurs, que la méthode empirique des coefficients de pondération de règles, fournis de la main même de l'expert, ait toujours aussi bonne audience. Cette méthode, vieille d'une quinzaine d'année, a maintes fois démontré ses limites. S'il en avait été autrement, le moteur EMYCIN aurait certainement connu un plus grand nombre d'applications. Elle est représentative d'une première génération de S.E., mais devrait céder le pas maintenant à des représentations à la fois plus "naturelles" et plus efficaces de l'incertitude.

#### **I.6.1.4 LA DOUBLE PHILOSOPHIE DU SYSTEME SAM**

Pour terminer ce débat sur le problème des coefficients de vraisemblance des faits et des règles, il convient de citer l'expérience tentée par Gascuel (GASCUEL 81) de marier deux sources de coefficients totalement différentes.

L'originalité du modèle mathématique de raisonnement incertain qu'il propose tient au fait que ce modèle met en concurrence, à chaque étape du raisonnement un modèle probabiliste (statistique) et un modèle logique de calcul de coefficients de vraisemblance.

La philosophie du modèle est la suivante:

- si un fait, en prémisses d'une règle, est totalement incertain, on utilise le modèle probabiliste qui en l'occurrence, est le mieux à même de décider,
- si une certitude, positive ou négative, existe à propos d'un

fait, on utilise alors le modèle logique.

Le mécanisme, très brièvement décrit, est le suivant:

SAM est un environnement de système-expert qui utilise:

- des règles de productions de type:

EST EN FAVEUR  
INDIQUE

- des tautologies:

CONSTITUENT  
S'EXCLUENT

- des faits dont la particularité est de posséder **deux** coefficients de vraisemblance (appelés arguments):

a) un argument statistique variant dans l'intervalle [0,1], nommé ARGS,

b) un argument logique à trois valeurs possibles: 0, 0,5 ou 1, correspondant respectivement aux notions de faux, d'incertain et de vrai.

La présence de ces deux coefficients distincts atteste et prend en compte la dualité du principe de raisonnement: tantôt expérimental et probabiliste, tantôt théorique et logique.

L'alternance entre ces deux principes est contrôlée par l'argument logique, comme l'indique la règle ci-après:

Pour tout fait A:

si ARGL(A) = 0,5  
alors ARG(A) <----- ARGS(A)  
sinon ARG(A) <----- ARGL(A)

L'argument statistique obéit à la loi de complémentarité à 1:

$$\text{ARGS}(\text{non } A) = 1 - \text{ARGS}(A).$$

Lorsque le fait A résulte de plusieurs règles, l'argument statistique doit être indépendant de l'ordre d'application des règles. La solution a consisté à définir ARGS de A comme une moyenne entre deux valeurs: CVO ET CVN. CVO est l'un indice du nombre de confirmation de A; à l'inverse, CVN est l'indice du nombre d'infirmité de A. D'où  $\text{ARGS}(A) = (\text{CVO}(A) - \text{CVN}(A) + 1)/2$ .

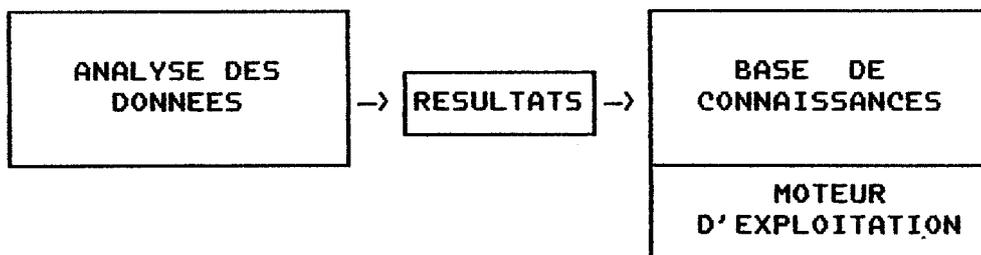
### I.6.2 VERS UNE SYNTHÈSE DES MÉTHODES PROPOSÉES

La multiplicité des méthodes d'aides à la décision médicale conduit à s'interroger sur leur dénominateur commun. Nous avons tenté jusqu'ici d'analyser la spécificité de ces méthodes ainsi que leur capacité à aborder et résoudre les divers problèmes posés par le raisonnement et la décision incertains. Existe-t-il alors, au-delà des différences ou des oppositions, une manière de les combiner en vue d'une intégration dans un système de type expert, de plus haut niveau ? Il semble en effet que l'on s'oriente de plus en plus, en matière de systèmes experts, vers des systèmes intégrés complexes (tel est le cas par exemple du projet EDORA (ROUSSEAU, 86)) pouvant inclure des techniques de traitement du signal.

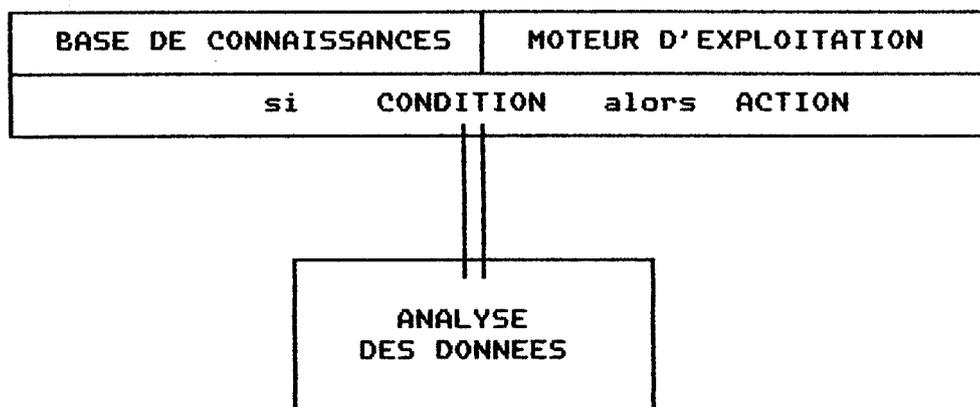
Compte tenu de ce qui a été dit jusqu'ici, les méthodes de quantificatives d'une part, les méthodes symboliques ou logiques d'autre part, apparaissent comme concurrentielles. Etant issues de philosophies différentes, leur mariage dans une même application apparaît au prime abord impossible. Leur point commun essentiel étant de construire une connaissance exploitable automatiquement, on peut alors transposer la question au niveau des connaissances

exprimées dans les divers formalismes: comment exploiter le plus rationnellement possible des connaissances issues de formalismes et de philosophies différentes ?

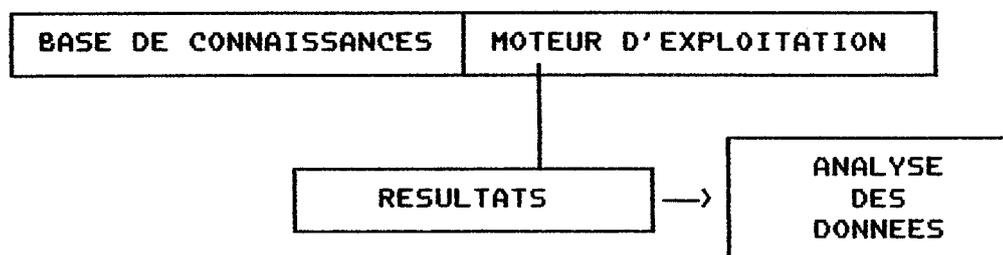
Si on décide d'associer les traitements symboliques aux traitements numériques des connaissances, alors, on obtient un système hybride du type A, B ou C.



A - MOTEUR AVAL



B - MOTEUR PARALLELE



C - MOTEUR AMONT

L'organisation de type B, consistant en l'évaluation de

## Chapitre I

prémisses (ou d'attribut de schémas) au moyen d'une analyse numérique, semble correspondre au mariage le plus étroit entre traitement symbolique et traitement numérique.

## CHAPITRE II

### UNE PRODUCTION DE CONNAISSANCES EXPERTES EN MEDECINE

#### II.1 POURQUOI UNE PRODUCTION DE CONNAISSANCES EXPERTES ?

##### II.1.1 POINT DE VUE MEDICAL GENERAL

Le concept de système expert est apparu et s'est rapidement répandu à la fin des années 70. Beaucoup ont vu en lui une libération du carcan que représentait la programmation procédurale. Désormais, la connaissance n'est plus figée dans les instructions d'un programme, mais devient accessible et modifiable à souhait par l'expert, ultime détenteur de cette connaissance. Seule revient à l'informaticien la tâche de mettre en application, au moyen d'un moteur d'exploitation, les connaissances engrangées dans la base de connaissances. Néanmoins, comme le fait remarquer R.S. MICHALSKI, le problème de la construction d'une base de connaissances, constitue l'un des obstacles majeurs à la réalisation d'un système expert: "(...) knowledge acquisition is presently the most limiting 'bottleneck' in the development of modern knowledge-intensive artificial intelligence systems" (MICHALSKI 83, p.112).

En ce qui concerne la médecine, deux raisons peuvent être avancées pour expliquer cette difficulté: soit le formalisme de représentation des connaissances gêne, pour diverses raisons, l'expert médical: inadaptation au domaine, nécessité de considérer des calculs d'incertitude, langage peu naturel, etc; soit c'est la connaissance médicale elle-même qui n'existe pas

## Chapitre II

encore de manière opérationnelle pour constituer une base de connaissances. Tel est le cas des domaines en cours d'exploration ou à forte croissance de connaissances.

Ce problème a reçu, dans le cadre même des systèmes experts, un certain nombre de solutions et constitue plus que jamais un terrain de recherches ouvert. Les classiques outils de recherche d'incohérence ou d'inconsistance de la base de connaissances constituent déjà des aides à l'expression de cette connaissance. Ceux-ci ne sont cependant pas suffisants, dans la mesure où ils n'interviennent que dans la phase de représentation et d'organisation de la connaissance, et non pas dans la phase de production proprement dite des connaissances.

La question que l'on peut alors se poser est celle de la pertinence et de la légitimité de l'emploi de l'ordinateur dans la production des connaissances médicales.

Sans entrer trop à fond dans ce problème, nous évoquerons simplement le fait qu'une connaissance s'acquiert, le plus souvent par la pratique. Au cours de cette pratique, intervient un certain nombre d'outils parmi lesquels on compte désormais l'outil informatique. Ce dernier ne devrait-il pas, en tout état de cause, avoir pour vertu essentielle de faciliter et d'accélérer la mise en forme de la connaissance ?

Un exemple typique, pour illustrer ce propos, est celui de la base de connaissances réalisée au service d'endocrinologie du Sheba Medical Center (Israël) (SCHILD 78).

Le principe de construction en a été le suivant:

## Chapitre II

a) sélection d'une large population de sujets présentant une absence de fertilité d'origine endocrinienne, avec, pour chaque sujet, un diagnostic clairement établi, ainsi que la description précise des symptômes présentés,

b) partition de la population en sous-groupes, chaque sous-groupe représentant un échantillon de malades atteints d'une maladie hypothalamique, hypophysaire ou ovarienne,

c) expression, dans chaque groupe, du champ de variation de la maladie, au moyen de modèles (ou exemplaires), avec relations entre ces exemplaires. Par exemple, l'apparition de règles spontanées ou provoquées après administration de progestérone a cinq conséquences importantes:

- 1 - la présence d'un utérus ayant un endomètre répondant normalement aux stéroïdes ovariens.
- 2 - la présence d'une certaine activité oestrogénique ovarienne qui indique:
- 3 - la présence d'une activité ovarienne minimale et,
- 4 - une stimulation gonadotrophique suffisante pour provoquer une maturation folliculaire ainsi que,
- 5 - une activité hypothalamique suffisante pour donner lieu à une stimulation basale hypophysaire.

Cet exemple suffit à illustrer le fait que la connaissance médicale est d'élaboration complexe. La perspective de réalisation d'une base de connaissances peut constituer l'aiguillon principal de cette élaboration. On peut également voir dans ce travail d'élaboration une importante retombée de l'utilisation de l'informatique pour le diagnostic médical.

Néanmoins, le problème de l'aide à l'extraction et à la formulation des connaissances est encore loin d'être résolu d'une

manière systématique, au moyen d'outils appropriés.

On peut distinguer deux orientations d'aides possibles (nous plaçons en-dehors à part de ces deux orientations un système tel que EURISKO (LENAT 83) dont la production de concepts nouveaux constitue un but à part entière et non pas une simple aide):

La première consiste en une aide à la mise en mémoire des connaissances selon un formalisme précis et plus ou moins complexe de représentation. C'est dans cette orientation que se situent tous les outils "satellites" d'un logiciel de gestion de base de connaissances: vérification de cohérence et de consistance, outils d'explication du mode de fonctionnement de la base, outils de trace du raisonnement, etc. Certains systèmes, employant des pondérations numériques de faits ou de règles de production (§ II), proposent une analyse de l'incidence de ces coefficients sur la succession des déductions (GASCHING 79). Le système TEIRESIAS (DAVIS 79) guide l'acquisition de nouvelles règles de production et réalise ainsi un transfert interactif des connaissances.

La seconde orientation consiste à intégrer à l'environnement "système expert" des outils et méthodes d'investigation situés généralement hors du champ de l'Intelligence Artificielle. Ces outils et méthodes sont: la modélisation biomathématique, la simulation fonctionnelle, les techniques d'analyse des grands tableaux (ACP, analyse factorielle, analyse discriminante, etc.), le traitement du signal, l'imagerie médicale, etc... Cette seconde orientation constitue une extension de l'orientation purement "système expert" en ce qu'elle conserve la méthodologie du traitement non algorithmique d'informations symboliques, tout en

## Chapitre II

intégrant des outils d'analyse numérique adaptés à la production de connaissances nouvelles dans certains domaines.

Elle introduit également un important changement: la remise en cause de l'idée d'**imitation** du raisonnement humain. Le paradigme du système expert est sous-tendu en effet par l'idée d'une similitude plus ou moins proche entre le mode de fonctionnement du logiciel et celui de l'activité cérébrale de l'homme. En fait, ce paradigme se manifeste beaucoup plus, à l'heure actuelle, dans la simple recherche d'une **similitude de résultats** produits dans les deux cas.

Dans cette optique, le système expert est destiné à reproduire la démarche déductive pas-à-pas de l'expert lui-même, ceci en raisonnant à chaque étape sur un nombre limité de faits et d'actions. L'objectif est alors de produire un double de l'expert, appréciable non pas pour son infailibilité (l'expert lui-même ne l'est pas), mais pour l'**exhaustivité** de ses cheminements déductifs ainsi que pour la **permanence** de sa rigueur logique (deux qualités parfois fluctuantes chez l'homme).

Cette idée de similitude d'activité entre l'expert et le système est remise en cause pour la simple raison que les deux protagonistes ont chacun des potentialités difficilement comparables. La plupart des comparaisons aboutissent en effet à des positions triviales telle que celle-ci: l'ordinateur peut traiter sans se tromper une très grande quantité d'informations numériques ou symboliques alors que l'homme en est en général incapable. A l'inverse, l'homme **crée** selon des processus qui nous échappent pour la plupart, alors que l'ordinateur est, et restera probablement encore longtemps, un **exécutant**, aussi sophistiqué soit-il.

Il y a confusion, semble-t-il, entre le fait de raisonner sur un nombre limité de données simultanées (démarche diagnostique

classique du médecin) et le fait de raisonner "pas à pas", c'est-à-dire une étape après l'autre, sans préjuger de la quantité de données qui intervient à chaque étape.

Il est évident qu'un raisonnement suit une démarche **pas à pas**, puisqu'il y a cheminement dans une arborescence de possibilités. S'il n'y avait qu'une étape, ce ne serait plus un raisonnement, mais une simple déduction. L'utilisation, par exemple, de règles heuristiques correspond à des étapes de raisonnement. Un raisonnement qui ne procède pas par étapes successives ne serait plus un raisonnement, mais une intuition magique au regard d'un observateur extérieur. Un raisonnement correspond nécessairement à une démarche qui peut être expliquée. De plus, une explication n'est autre qu'une **succession de relations causales**. Si, dans une explication, une seule de ces relations causales vient à manquer, c'est toute l'explication qui est remise en cause. Il ne saurait donc exister de raisonnements autres que pas à pas.

Par contre, on peut discuter sur ce qu'est véritablement un **pas** de raisonnement. Ce **pas** de raisonnement humain est le passage d'un certain nombre de prémisses à une conclusion locale, laquelle conclusion pourra éventuellement devenir prémisses d'un nouveau **pas** de raisonnement.

Un **pas** de raisonnement humain n'a aucune raison **a priori** de ressembler au **pas** de raisonnement d'une machine. L'homme produit des machines pour compléter ses capacités manuelles et intellectuelles. La vocation de la machine est donc de surpasser, pour une action précise, les capacités humaines.

En résumé, l'expert est très limité du point de vue de

l'ordinateur et vice-versa. Il n'est donc pas toujours judicieux de chercher à appliquer à celui-ci des méthodes propres à celui-là.

### II.1.2 CAS PRECIS DE LA GENETIQUE MEDICALE

La Génétique Médicale, du point de vue de l'Aide au Diagnostic, constitue un domaine relativement singulier. En effet, cette branche de la médecine traite un nombre de symptômes, de syndromes et de maladies extrêmement important. Par exemple, le nombre de syndromes actuellement répertoriés dépasse le chiffre de 1200, et augmente rapidement, d'année en année, grâce au développement de la recherche dans ce domaine. Par voie de conséquence, le nombre de symptômes connaît lui aussi une croissance rapide. Il ne faut pas voir dans ce phénomène une augmentation réelle du nombre des maladies génétiques, mais la simple évolution, grâce aux techniques nouvelles, du tableau des maladies reconnues qui conduit plus à un éclatement qu'à une réduction de la classification nosologique. Cette évolution rapide de la Génétique Médicale est sans aucun doute une conséquence naturelle d'importants progrès scientifiques récemment réalisés dans ce domaine, progrès qui donnent lieu à un nombre proportionnellement croissant d'ouvrages et de publications dans plusieurs pays.

On peut donc caractériser ce domaine par les données suivantes:

- plusieurs milliers de symptômes,
- plus d'un milliers de syndromes et maladies,
- un important volume de connaissances,
- un nombre important de chercheurs répartis à travers le monde, d'où:
- une évolution rapide et ininterrompue des connaissances,
- des syndromes caractérisés par leur rareté (faible prévalence),

## Chapitre II

- des symptômes quelque-fois difficiles à identifier formellement sur un malade,
- un diagnostic pouvant, comme dans toute branche médicale, avoir un caractère urgent.

Dès lors, le problème auquel se heurte le praticien est celui de la prise en compte, dans une pratique quotidienne, de cette masse considérable d'informations médicales. Quelques tentatives de compilation des travaux et publications existent sous la forme de volumineux recueils contenant plusieurs centaines de syndromes (§ II.3), recueils régulièrement mis à jour tous les trois ou cinq ans.

Lors de l'examen d'un patient, la difficulté de considérer la totalité des connaissances diagnostiques est accrue par le caractère rarissime de certains syndromes et l'imprécision de la description littérale.

A ces difficultés majeures, s'ajoutent diverses difficultés secondaires, mais non moins présentes, comme celle de la **reconnaissance** des symptômes malformatifs: il existe en effet des cas litigieux dans lesquels il est malaisé de reconnaître s'il y a vraiment malformation ou non.

Ainsi, le travail de recherche de syndromes à partir de symptômes, dans la masse d'informations livresques que nous avons citée, et en présence de cas rares et mal connus, est extrêmement fastidieux. Le praticien ne s'y résoud qu'avec réticence, et seulement pour un nombre réduit de cas.

Il demeure que cette quasi-impossibilité matérielle de conduire de manière approfondie une identification de syndrome porte à conséquence. D'une part, un certain nombre de maladies rares pas-

## Chapitre II

sent inaperçues dans le cabinet du médecin, ce qui diminue leur connaissance statistique et épidémiologique, et cause éventuellement des erreurs de traitement; d'autre part, un fossé de plus en plus large risque de se creuser entre le développement de la connaissance génétique théorique et celui de la pratique médicale *in situ*. Enfin, le conseil génétique, dans le cas de maladies génétiques héréditaires, s'en trouve appauvri.

Ces considérations, parmi d'autres, motivent le développement d'un outil de recherche et d'aide au raisonnement en génétique médicale capable d'augmenter efficacement les capacités intellectuelles et associatives du médecin.

C'est ici que se pose le véritable problème de l'aide au raisonnement et à la déduction dans un domaine caractérisé, nous l'avons vu, par :

- 1 - son étendue,
- 2 - son incertitude,
- 3 - son incomplétude,
- 4 - son évolutivité.

Il est évident qu'un tel domaine échappe totalement à l'emprise d'un seul expert ou d'un groupe restreint d'experts.

L'expertise proprement dite, c'est-à-dire la somme totale des connaissances sur le domaine, n'existe pas sous une forme synthétique et communicable par un expert. Cette connaissance, nous l'avons vu, existe néanmoins sous forme encyclopédique (études, articles, ouvrages de synthèse, etc...).

Le schéma classique : expert -> expertise -> système expert ne saurait donc être appliqué.

C'est la raison pour laquelle un outil d'aide à la production et à l'apprentissage de connaissances, intermédiaire entre la connaissance globale existant à l'état diffus et une véritable base de connaissances exploitable, semble indispensable.

En outre, la littérature apparaît, pour de nombreux praticiens, comme la principale source d'informations. Cependant, dans la masse considérable de syndromes répertoriés, nombre d'entre eux ont des liens de ressemblance. L'idée que certains d'entre eux ne seraient que des expressions différentes d'une même cause n'est pas à écarter. Cette incertitude sur les syndromes résulte du fait qu'il est, à l'heure actuelle, très difficile de relier une maladie malformative à son origine génique, ou d'identifier à quel moment de l'embryogénèse ou du développement du fœtus est survenue le facteur déclenchant la malformation.

L'étude du rapprochement entre le syndrome et son origine est subordonnée au nombre et à la quantité des diagnostics effectués dans le plus grand nombre de centres médicaux. Or nous avons vu que la quantité des syndromes ainsi que leur faible prévalence rendent cette tâche extrêmement difficile. La nécessité, voire l'urgence, du nombre de diagnostics établis renforce donc l'intérêt d'un outil d'aide à la production de connaissances dans ce domaine.

### II.2 UN EXEMPLE DE PRODUCTION DE CONNAISSANCES : LE SYSTEME SEEK

Réalisé par P. POLITAKIS et S.M. WEISS, le système SEEK (POLITAKIS 84) est un système expert caractérisé par ses capacités, d'une part à renseigner l'expert sur les performances de sa base de connaissances, d'autre part à formuler des conseils pour

## Chapitre II

l'affinement des règles de production. Ces conseils concernent la **généralisation** ou la **spécialisation** de certaines règles. Ce système a été appliqué en rhumatologie.

Alors que, dans l'exemple de SCHILD (§ II.1), l'élaboration et l'affinement des connaissances se situaient hors du système expert, le système SEEK, également fondé sur une étude de cas expérimentaux, intègre l'aide à la production de connaissances dans l'environnement même du système expert.

Une **table de critères** est associée à chaque diagnostic candidat à l'explication d'une sémiologie de malade.

Chaque table possède deux parties :

- a) l'ensemble des faits, majeurs ou mineurs, en faveur du diagnostic,
- b) l'ensemble des règles conduisant au diagnostic. Par exemple, pour le diagnostic d'une maladie du tissu conjonctif, la liste des faits (partie a) pourra être :

critères majeurs	critères mineurs
1. mains moites	1. inflammation muscul. moyenne
2. sclérodactylie	2. anémie
3. phénomène de Raynaud (hypomotricité oesophagienne)	3. péricardite
4. inflammation musc. sévère	4. arthrite, 4 à 6 semaines
5. coeff. de diffus. du CO < 70	5. pleurite
	6. alopecie

La partie b de la table est un tableau à deux dimensions:

- 3 colonnes constituées par les niveaux de certitude:  
**possible, probable, certain.**

## Chapitre II

- 3 lignes correspondant respectivement:

- au nombre d'observations majeures et mineures,
- au critère minimum,
- aux exclusions.

Exemple: la règle permettant de conclure à une maladie du tissu conjonctif peut être représentée sous la forme suivante:

	<b>certain</b>	<b>probable</b>	<b>possible</b>
<b>observations</b>	4 majeurs	2 majeures 2 mineures	3 majeures
<b>critère minimum</b>	anticorps RNP positif	anticorps RNP positif	néant
<b>exclusions</b>	anticorps SM positif	néant	néant

Ce tableau s'interprète de la façon suivante:

si le malade a au minimum 4 observations majeures (prises parmi les éléments de la partie a),

et si l'anticorps RNP est positif,

et si l'anticorps SM est non positif,

alors conclure avec certitude à la maladie du tissu conjonctif.

Dans la plupart des applications, de multiples règles sont établies pour chaque niveau de certitude.

L'exemple précédent de la table de critères est appelé un **sous-modèle du modèle** général constitué par la rhumatologie.

Ainsi, un modèle est une collection de sous-modèles, c'est-à-dire de tables de critères, correspondant à chaque diagnostic.

Lors d'une consultation, le système recherche d'abord l'ensem-

ble des sous-modèles applicables, puis, pour chacun d'eux, calcule sa performance. Cette performance est un nombre représentatif à la fois du degré de coïncidence entre la sémiologie observée et celle du modèle, et d'une mesure statistique de performance des règles de la table. Cette dernière mesure est établie par application du modèle sur un ensemble de dossiers complets de malades (sémiologie et diagnostic) pris comme référence.

L'aide à l'affinement des règles se fait à travers deux notions, déjà définies par (MICHALSKI 83), celle de **généralisation** et celle de **spécialisation**.

La généralisation consiste en des modifications destinées à affaiblir une règle R : une nouvelle règle Rg telle que  $R \Rightarrow Rg$  en résulte.

La généralisation peut être obtenue par exemple en diminuant le nombre d'observations majeures ou mineures ou encore en supprimant une observation du critère minimum ou des exclusions.

A l'inverse, la spécialisation consiste en des modifications destinées à renforcer une règle R: une nouvelle règle Rs telle que  $Rs \Rightarrow R$  en résulte.

Il suffit pour cela d'ajouter une observation soit au critère minimum, soit aux observations majeures et mineures, soit aux exclusions.

En conclusion, nous pouvons retenir du système SEEK le fait qu'il se démarque du système expert classique en proposant à l'expert des moyens de vérification et de renforcement de sa connaissance. Ce système travaille sur des règles pour en proposer

des améliorations. Cependant, il ne propose pas à proprement parler de nouvelles règles, inconnues de l'expert. C'est sur cette dernière éventualité que nous allons tenter de travailler maintenant.

### II.3 SPECIFICITES DU DOMAINE DE LA GENETIQUE MEDICALE

Nous venons de voir que la Génétique Médicale se caractérise par un nombre impressionnant et sans cesse croissant de syndromes et de symptômes. C'est la raison pour laquelle nous avons réduit la base de données au cas particulier des **syndromes malformatifs** incluant une malformation **labiale ou labio-palatine** (BONAITI 82).

L'élaboration complète d'une base de données de syndromes et symptômes demandera un délai certainement long. En effet, la réalisation de la base dans son état actuel de 102 syndromes et 675 symptômes, nous a demandé un investissement temporel de 2 ans.

Nous n'avons pas négligé le fait (BAUDINO 85) qu'un certain nombre de réalisations de thésaurus et bases de données informatisés existent ou sont en cours de développement: système GENDIAG à Marseille (§ I.5.2), système ADM à Rennes (§ I.5.2), etc.

Nous nous en sommes souvent inspirés. Cependant, notre approche particulière de la symptomatologie au travers:

- d'entités symptomatiques munies de valeurs,
- de scores,
- de combinaisons linéaires d'entités symptomatiques,

a nécessité une redéfinition complète, d'une part, d'un thésaurus

## Chapitre II

de symptômes (§ II.5.1 et III.2.2.3), d'autre part, d'une hiérarchie d'entités symptomatiques (§ II.5.2 et III.2.2.5).

Ce travail étant aujourd'hui bien avancé, il ne serait pas inutile d'envisager des collaborations scientifiques avec d'autres centres de recherche. De telles collaborations permettraient d'augmenter considérablement le nombre de symptômes et de syndromes de notre base, et elle permettrait certainement en retour d'enrichir des bases voisines. Le système d'aide à l'apprentissage et à la production de connaissances que nous présentons ne fournira en effet sa pleine capacité qu'en présence d'une base extrêmement riche et complète.

Pour l'heure, l'expérimentation effectuée sur une centaine de syndromes a essentiellement une valeur illustrative des idées et des méthodes utilisées.

Voici quelques éléments caractéristiques du domaine:

L'ensemble des syndromes malformatifs se partitionne selon les 4 classes suivantes, avec, pour chacune d'elles, le nombre de syndromes actuels:

- **classe 1: syndromes d'origine génétique:**  
maladies ou malformations entraînées par un gène délétère.  
52 syndromes.
- **classe 2: syndromes d'origine chromosomique:**  
maladies ou malformations entraînées par une anomalie chromosomique (mongolisme, trisomie 9, etc.). 26 syndromes.
- **classe 3: syndromes d'origine environnementale:**  
principalement des accidents survenus en cours de grossesse  
3 syndromes.
- **classe 4: syndromes d'origine inconnue.** 21 syndromes.

Le nombre de symptômes descriptifs d'un syndrome s'étend de 3

## Chapitre II

à 80, avec une moyenne de 23.

Les descriptions des syndromes d'origine chromosomique sont les plus riches.

Les plus pauvres descriptions reviennent aux syndromes d'origine inconnue.

L'étude de la hiérarchie des entités symptomatiques a nécessité un travail préalable de réalisation, sur le papier, d'un thésaurus hiérarchisé des symptômes. Ce thésaurus (BAUDINO 85) est structuré en quinze chapitres:

- I - GROSSESSE-NAISSANCE
- II - CROISSANCE-LONGEVITE-PUBERTE
- III - REGION CRANIO-FACIALE
- IV - PEAU-ANNEXE-DERMATOGLYPHE
- V - TRONC
- VI - MEMBRE
- VII - APPAREIL GENITAL EXTERNE
- VIII - APPAREIL CARDIO-VASCULAIRE
- IX - APPAREIL RESPIRATOIRE
- X - APPAREIL DIGESTIF
- XI - APPAREIL GENITO-URINAIRE
- XII - SYSTEME HEMATOPOIETIQUE ET IMMUNOLOGIQUE
- XIII - SQUELETTE
- XIV - SYSTEME NERVEUX CENTRAL, PERIPHERIQUE, APPAREIL SENSORIEL
- XV - GLANDE ENDOCRINE-ONCOLOGIE-METABOLISME

Cette structuration a été établie par le docteur A. JOANNARD, selon la logique de la démarche diagnostique et des regroupements anatomiques. Elle tient compte des classifications actuelles: classification de la Société Américaine de chirurgie de la main, classification internationale des maladies osseuses constitutionnelles, etc. Elle correspond, du point de vue hiérarchique, au premier niveau de l'arborescence des entités symptomatiques (§ II.5.2.4).

### II.4 PRINCIPE D'UTILISATION DE L'ANALYSE DES DONNEES

#### II.4.1 NATURE ET ORIGINE DES CONNAISSANCES - POURQUOI L'ANALYSE DES DONNEES

L'idée directrice de ce système, est qu'il n'existe pas de différence qualitative **a priori** entre la connaissance élaborée et mûrie par un expert humain, et celle qui est issue de l'analyse mathématique des données expérimentales, ceci pour deux raisons fondamentales :

- d'abord, le principe même d'un Système Expert suppose que la base de connaissances soit enrichie par un grand nombre de connaissances déclaratives sans préjuger de la provenance de ces dernières: intuition, expérimentation, théorie, etc., sous réserve que ces connaissances restent cohérentes entre elles. Le mode de production d'une connaissance n'intervient pas directement sur son mode d'utilisation de même que le fabricant d'un marteau a peu de raisons de définir **a priori** le champ d'utilisation de ce marteau. C'est à l'utilisateur, d'en décider après l'avoir expérimenté. Il en est de même pour les outils d'Analyse des Données.

- Par ailleurs, l'Analyse des Données fournit un **point de vue** sur une question au même titre qu'un expert humain. Ce point de vue peut être considéré comme fiable puisque fondé sur des expériences ou des observations et bâti dans un cadre rigoureux: l'algèbre linéaire. Bien entendu, l'Analyse des Données ne saurait remplacer la présence de l'expert humain qui aura choisi et défini, selon ses propres codes de valeur, les variables observées sur chaque objet d'observation. L'incertitude et l'erreur ne sauraient épargner cette démarche. La seule affirmation que l'on peut poser est que les opérations algébriques et matricielles, effectuées sur les données, sont exactes.

## Chapitre II

Les raisons principales de l'utilisation de l'Analyse des Données (BENZECRI 81), (DE LAGARDE 83) sont les suivantes:

- elle est adaptée au traitement des grands ensembles de données,
- elle permet d'extraire des relations empiriques entre un certain nombre de variables,
- elle constitue en elle-même un outil à part entière d'aide à la décision,
- elle fournit un cadre mathématique rigoureux pour la modélisation des données,
- elle permet une visualisation de phénomènes complexes (A.C.P.),
- elle fournit d'importants éléments de décision, propres à être intégrés dans un ensemble de connaissances.

On peut donc considérer l'Analyse des Données aussi bien comme une entité "expert" à qui l'on fournit des données et qui en restitue un point de vue, que comme un outil efficace et rigoureux, utilisable par un expert, pour augmenter, renforcer ou vérifier ses propres connaissances.

En outre, l'un des fondements de l'Analyse des Données est l'aide à la décision. Nul n'ignore d'autre part que l'aide à la décision est l'une des principales raisons d'être des S.E. (décisions diagnostique, thérapeutique, pronostique). Il n'y a donc aucune rupture épistémologique dans le rapprochement entre l'Analyse Discriminante et les travaux sur le raisonnement de l'expert. En effet, si l'expert était doté de la capacité de

construire mentalement la matrice d'inertie d'un système, d'en découvrir les axes principaux ainsi que les positions relatives des objets par rapport à ces axes, il est indéniable qu'il utiliserait cette investigation.

#### II.4.2 EXPOSE DE PRINCIPES FONDAMENTAUX - L'A.C.P. CANONIQUE

Voici quelques brefs rappels sur l'Analyse en Composantes Principales Canonique:

appelons individus les syndromes de la base de données, et caractères les variables construites à partir des entités symptomatiques.

On appelle ACP Canonique, l'ACP pratiquée sur le nuage centré des individus: pour chaque variable, un individu a pour valeur sa valeur réelle moins la moyenne des individus sur cette variable.

On part d'un tableau de n individus et de p caractères noté X, et l'on considère tout d'abord que tous les individus ont un même poids égal à 1/n.

Construisons le tableau centré X :

$x_{ij}$  : valeur de la variable j sur l'individu i

$$X_j(i) = x_{ij} - \frac{1}{n} \sum_{i=1}^n x_{ij}$$

	- X <sub>i</sub> .. X <sub>J</sub> .. X <sub>p</sub>
1	
⋮	
i	X <sub>J</sub> (i)
⋮	
n	

X  
tableau (n,p)

On appelle variable n°j la colonne j de X :  $X_j \in \mathbb{R}^n$

On appelle individu n°i la ligne i de X :  $i \in \mathbb{R}^p$

$$i = \begin{bmatrix} X_1(i) \\ \vdots \\ X_p(i) \end{bmatrix} \quad X_j = \begin{bmatrix} X_j(1) \\ \vdots \\ X_j(n) \end{bmatrix}$$

Pour effectuer l'ACP canonique, on diagonalise la matrice carrée d'ordre p, notée M, obtenue de la manière suivante :

$$M = \text{transposée}(X) \cdot X$$

Les valeurs propres ordonnées par valeurs décroissantes sont notées  $\Gamma_1, \Gamma_2, \dots, \Gamma_p$  avec  $\Gamma_1 \geq \Gamma_2 \dots \geq \Gamma_p \geq 0$

Les vecteurs propres unitaires associés sont appelés **vecteurs principaux** et notés par exemple  $v_1, \dots, v_{p'}$ , où p' désigne le nombre de vecteurs propres non nuls. Dans la suite, nous supposons que  $p' = p$ .

$$v_1 = \begin{bmatrix} v_{11} \\ \vdots \\ v_{1p} \end{bmatrix} \in \mathbb{R}^p \quad v_{ij} : j \text{ ème composante de } v_i$$

## Chapitre II

L'axe  $\delta v_i$  déterminé par  $v_i$  est l'axe principal n°i.

$$C_1 = \sum_{i=1}^p v_{1i} X_i \quad \text{1ère composante principale}$$

$$C_j = \sum_{i=1}^p v_{ji} X_i \quad \text{jème composante principale}$$

$$\text{On a : } \sigma^2(C_j) = \Gamma_j \quad (\text{variance de la composante } C_j)$$

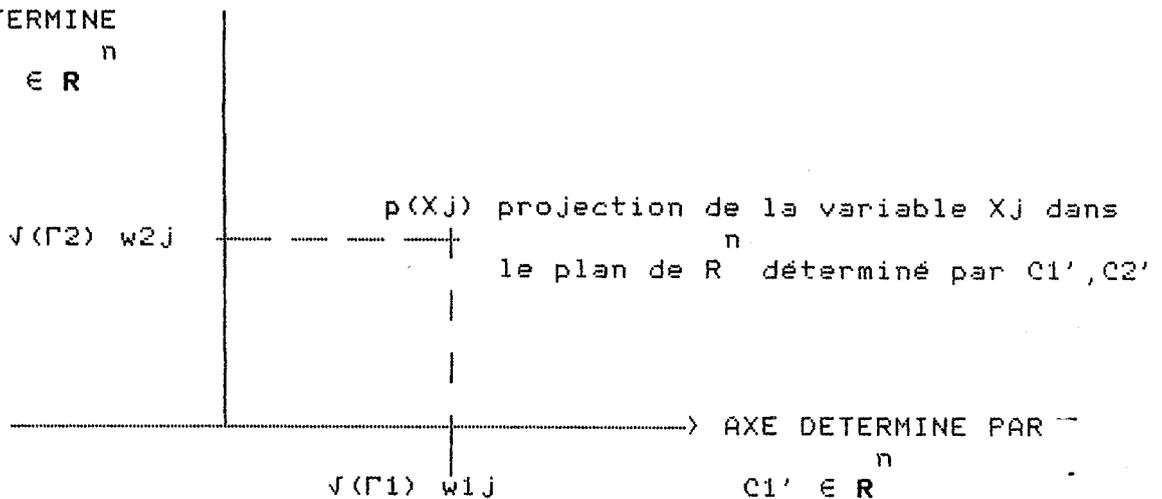
On appelle  $C_j' = \frac{C_j}{\sqrt{\Gamma_j}}$  la jème composante principale réduite.

Inversement, nous noterons :  $X_i = \sum_{j=1}^n w_{ij} \cdot C_j$ , la matrice  $W$

égale à l'inverse de la matrice  $V$  d'élément général  $v_{ij}$

### Représentations euclidiennes des variables dans $R^n$

AXE DETERMINE  
PAR  $C_2' \in R^n$



Les  $F$  coordonnées des  $p$  variables sont donc données par :

$$\sqrt{\Gamma_1} w_1 = \begin{bmatrix} \sqrt{\Gamma_1} w_{11} \\ \vdots \\ \sqrt{\Gamma_1} w_{1p} \end{bmatrix}$$

Chapitre II

Les 2e coordonnées des p variables sont données par les coordonnées du vecteur  $\sqrt{(\Gamma_2)} w_2$ .

La qualité globale d'une telle représentation est donnée par :

$$q_{1,2} = \frac{\Gamma_1 + \Gamma_2}{\sum \Gamma_i}$$

$q_{1,2}$  est égal à la proportion de variance expliquée par les 2 premières composantes principales.

La qualité de la représentation d'une variable dans cette représentation est donnée par :

$$q_{1,2} (X_j) = \frac{\Gamma_1 \cdot w_{1j}^2 + \Gamma_2 \cdot w_{2j}^2}{\sigma^2 (X_j)}$$

où  $\sigma^2(X_j)$  = jème terme diagonal de M.

Pour que la variable  $X_j$  soit bien représentée, il faut que  $q_{1,2} (X_j)$  soit proche de 1. A l'inverse, si  $q_{1,2} (X_j)$  est proche de 0,  $X_j$  est mal représentée.

Il faut donc, pour pouvoir interpréter la représentation des variables, le tableau à p lignes, 4 colonnes suivant :

Représentation et qualité sur  $\mathcal{SC}_1'$

Représentation et qualité sur  $\mathcal{SC}_2'$

COMPOSANTES	QUALITE	COMPOSANTES	QUALITE
$\sqrt{(\Gamma_1)} \cdot w_{11}$	$\Gamma_1 \cdot w_{11}^2 / \sigma^2 (X_1)$	$\sqrt{(\Gamma_2)} \cdot w_{21}$	$\Gamma_2 \cdot w_{21}^2 / \sigma^2 (X_1)$
⋮	⋮	⋮	⋮
$\sqrt{(\Gamma_1)} \cdot w_{1p}$	$\Gamma_1 \cdot w_{1p}^2 / \sigma^2 (X_p)$	$\sqrt{(\Gamma_2)} \cdot w_{2p}$	$\Gamma_2 \cdot w_{2p}^2 / \sigma^2 (X_p)$

## Chapitre II

### Représentation des individus:

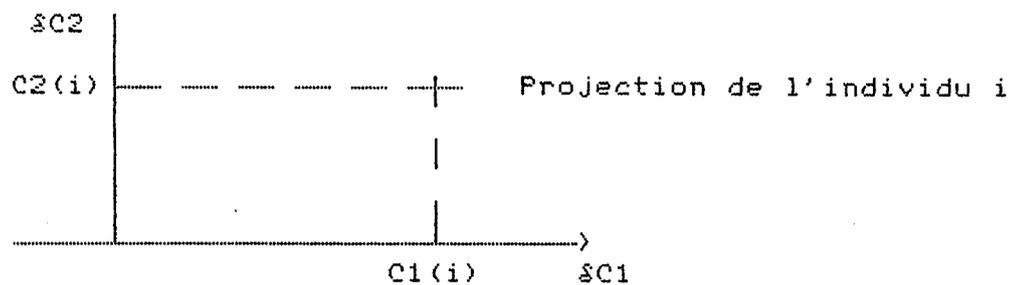
Nous avons :

$$C1(i) = \sum_{j=1}^P v1j Xj(i)$$

$$C2(i) = \sum_{j=1}^P v2j Xj(i)$$

$$\vdots$$
$$\vdots$$
$$Cp(i) = \sum_{j=1}^P vpj Xj(i)$$

d'où la représentation des individus dans le plan des variables  $C1$  et  $C2$  (ou premier plan factoriel) :



Indice de qualité de représentation de l'individu  $i$  sur  $\$C1$  (contribution de l'axe 1 à l'individu  $i$ ):

$$q1(i) = \frac{C1(i)^2}{\sum_{j=1}^P [Cj(i)]^2}$$

De même,

$$q2(i) = \frac{C2(i)^2}{\sum_{j=1}^P [Cj(i)]^2}$$

La qualité de la représentation de  $i$  sur le premier plan factoriel ( $\$C1, \$C2$ ) est:

$$\cos^2\theta = q_1(i) + q_2(i) = \frac{C_1(i)^2 + C_2(i)^2}{\sum_{j=1}^p C_j(i)^2},$$

où  $\theta$  est l'angle formé par le segment liant l'origine à l'individu  $i$  avec le plan de projection.

Il est également utile d'avoir la contribution du point  $i$  à l'axe 1 :

$$CTA_1(i) = \frac{p_i \cdot C_1(i)^2}{\Gamma_1},$$

où  $p_i$  est le coefficient de Précision Descriptive (CPD) du syndrome (§ II.5.4.3). Donc, pour pouvoir interpréter une représentation des individus sur le plan ( $\$C_1$ ,  $\$C_2$ ), il faut un tableau de  $n$  lignes et 6 colonnes :

$C_1(1)$	$q_1(1)$	$CTA(1)$	$C_2(1)$	$q_2(1)$	$CTA_2(1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_1(n)$	$q_1(n)$	$CTA(n)$	$C_2(n)$	$q_2(n)$	$CTA_2(n)$

#### II.4.3 CONDITIONS D'UTILISATION DE L'ANALYSE DES DONNEES

Les conditions premières d'utilisation de l'Analyse des Données sont les suivantes :

- a) variables quantifiables (continues ou discrètes),
- b) addition et multiplication par un nombre réel doivent être possibles,

## Chapitre II

- c) nombre assez élevé de niveaux,
- d) le nombre n de points du nuage doit être sensiblement plus élevé que le nombre m d'entités symptomatiques.

Pour satisfaire c et d, on utilisera des scores, par sommation des valeurs d'entités symptomatiques présentant une ou plusieurs parentés suivantes :

- même fonction physiologique
- même localisation anatomique
- même stade d'examen
- mesure commune (poids, dimension, ...)

Le score correspond à une parenté sémantique entre les entités symptomatiques (§ II.5.3).

### II.4.4 DEMARCHE FORMELLE D'ANALYSE

Nous exposons ici la démarche formelle d'apprentissage en mode hybride (branche procédurale de l'organigramme complet (§ II.7.2)). Une pré-étude statique du domaine complet est exposée en II.6.

- 1 - Point de départ: le tableau clinique du malade, correspondant par exemple à l'observation des symptômes : S1, S2, S3, S4.
- 2 - Expression des hypothèses diagnostiques (syndromes définis par leurs listes de symptômes) soit par exemple:

M1:	[ S1, S2, S3, S4 ]	S5, S6, S7, S8
M2:	[ S1, S2, S3, S4 ]	S5, S6, S9
M3:	[ S1, S2, S3, S4 ]	S5, S10, S11
M4:	[ S1, S2, S3, S4 ]	S6, S12, S13
M5:	[ S1, S2, S3, S4 ]	S6, S14, S16
M6:	[ S1, S2, S3, S4 ]	S7, S15, S17
M7:	[ S1, S2, S3, S4 ]	S8, S18, S19, S20

Les symptômes du tableau clinique ont été rassemblés au début.

3 - Chaque symptôme correspond à un couple (entité symptomatique, valeur): on en déduit les entités symptomatiques candidates:

E1, E2, E3, E4, E5, E6, E7, E8, E9, E10, E11, E12, E13

**N.B.:** Il y a moins d'entités symptomatiques que de symptômes candidats étant donné qu'une e.s. est porteuse d'un ou plusieurs symptômes.

4 - Ces entités symptomatiques candidates constituent un espace vectoriel dans lequel se positionnent les syndromes candidats. Le module de calcul (§ III.3) déduit, par l'Analyse des Données, l'entité symptomatique discriminant le mieux les syndromes candidats, au niveau d'observation qui suit celui qui nous a fourni les premiers symptômes, ici S1, S2, S3, S4.

soit, par exemple: observer l'entité symptomatique: E5

Ce module a pour objectif de proposer, à chaque étape de l'observation d'un malade, une orientation vers telle ou telle entité symptomatique (localisation physiologique, examen complémentaire, etc.) de manière à rendre optimale la découverte du diagnostic.

Les transitions d'états de la base de faits s'effectuent en deux étapes:

a) choix d'une entité symptomatique, résultant d'une analyse numérique ou de l'apprentissage antérieur.

b) demande d'observation par le médecin.

L'apprentissage antérieur est contenu, dans une base

d'apprentissage, sous forme de couples (BF, conséquence) dans lesquels la conséquence correspond soit à des propositions d'observation d'entités symptomatiques (révélées par l'analyse des données), soit à un ou plusieurs syndromes-diagnostic. La base d'apprentissage constitue une partie de la Base des Connaissances, et correspond exclusivement aux connaissances obtenues par analyse numérique, les autres connaissances étant fournies par l'expert.

### II.5 MODELISATION DU DOMAINE

#### II.5.1 ELEMENTS DE BASE : LES SYMPTOMES

##### II.5.1.1 LE CONCEPT DE SYMPTOME

Le symptôme constitue l'élément de base du système à partir duquel les autres éléments vont être élaborés.

Le domaine médical utilise largement, pour décrire maladies et syndromes, le concept de symptôme. Un symptôme est un trouble fonctionnel ou une lésion, c'est-à-dire la manifestation d'une maladie. Il représente un **fait** dont le praticien doit tenir compte dans l'établissement d'un diagnostic. Le concept de symptôme est donc identique à celui de **fait** généralement utilisé dans les systèmes experts. Il peut être exprimé aussi bien à l'aide d'une valeur numérique (température de 38 degrés) que par une valeur purement sémantique littérale (lèvre fendue).

Bien que les mots "faits" et "symptômes" représentent un seul et même concept, nous avons préféré le second au premier en raison de l'enracinement médical de ce travail. Cependant, ce choix de vocabulaire n'entache en rien la portée générale de la

## Chapitre II

méthode présentée. Cette remarque s'appliquera également au mot "syndrome" qui n'est autre qu'un **but** du point de vue des systèmes experts.

Signalons que nous emploierons le terme de symptôme comme synonyme de signe . En effet, bien qu'un "signe", d'un point de vue médical, ne désigne pas exactement la même chose qu'un "symptôme", la différence entre les deux concepts n'a aucune incidence sur le problème du diagnostic qui est au centre de ce travail.

La Génétique Médicale utilise une grande variété de symptômes: symptômes dysmorphiques, symptômes d'anomalies fonctionnelles... Nous considérerons également comme symptôme, dans quelques cas exceptionnels, un véritable syndrome. Il arrive en effet qu'un syndrome contienne, dans la liste des symptômes qui le décrit, un autre syndrome.

### II.5.1.2 NIVEAUX D'OBSERVATIONS - PARTITION ORDONNÉE DE L'ENSEMBLE DES SYMPTÔMES

Dans le cas de notre application, les symptômes se subdivisent en quatre catégories correspondant chacune à un niveau croissant de difficulté d'observation ou de coût. Aussi, chaque symptôme est affecté d'un indice compris entre 1 et 4, dont la signification est la suivante :

- 1: **Clinique immédiat**: symptôme visible ou identifiable sans instrument médiateur (observation visuelle, palpation, etc.).
- 2: **Clinique approfondi**: symptôme identifiable au moyen d'un instrument médiateur (appareil de mesure, auscultation, etc.), sur le lieu d'observation du malade.

3: **Para-clinique standard**: symptôme révélé par un examen courant (radiographie, numération formule sanguine, antibiogramme, etc.), hors du lieu d'observation du malade.

4: **Para-clinique approfondi**: symptôme mis en évidence par un examen paraclinique long et coûteux (biopsie, scanner, RMN, etc.).

Cette partition réalise un ordre partiel sur l'ensemble des symptômes.

**NB**: le coût d'observation d'un fait susceptible d'être demandé par le programme est une notion importante en médecine: SCHILD (SCHILD 78) par exemple propose le classement des faits en deux catégories: ceux qui ont une signification diagnostique sans nécessiter un coût élevé, et ceux qui encourent un coût ou un risque élevés.

### II.5.1.3 INDICE FREQUENTIEL D'APPARITION

Chaque symptôme de la liste est, en outre, affecté d'un coefficient ordinal correspondant au "poids" estimé du symptôme dans le syndrome. Ce coefficient prend ses valeurs dans l'ensemble (Q, S, T, I) défini de la façon suivante :

Le fait est présent quelquefois : Q,  
souvent : S,  
toujours : T,  
de manière indéterminée : I, chez un patient  
atteint de la maladie correspondant au syndrome.

Ces coefficients sont évalués dans l'intervalle [0, 1] des nombres réels, de la façon suivante :

Q -> 0,2  
S -> 0,7  
T -> 1  
I -> 0,5

Nous n'avons pas, dans l'exemple traité en II.6 et II.7, ce type de pondération, néanmoins, il peut être utilisé pour pondérer les coordonnées des syndromes sur les axes de l'analyse des données et permettre ainsi une meilleure séparation.

**N.B.** : La valeur I de l'Indice Fréquentiel d'Apparition, correspond à une lacune d'information dans la littérature médicale. Il est bien évident que cette valeur doit être considérée comme temporaire. Elle est le reflet, au niveau de la représentation des connaissances, du manque de connaissance épidémiologique qui existe en Génétique Médicale (§ II.5.4.3).

#### **II.5.1.4 ARBORESCENCE DES SYMPTOMES**

Comme nous le verrons au paragraphe suivant, les symptômes correspondent à des valeurs d'entités symptomatiques. L'arborescence des symptômes se superpose donc à l'arborescence des entités symptomatiques (§ II.5.2.3).

La notion de hiérarchie dans les symptômes intervient dans le calcul du **coefficient de précision descriptive** des syndromes (§ II.5.4.4).

#### **II.5.2 PREMIER NIVEAU DE STRUCTURATION : LES ENTITES SYMPTOMATIQUES**

##### **II.5.2.1 DEFINITION**

Une anomalie, qu'elle soit morphologique ou fonctionnelle, spatiale ou temporelle, est toujours l'expression d'un écart (un accroissement, une diminution, une détérioration, etc.) par rapport à un état de référence admis comme normal. Il est donc aisé de reconnaître, dans chaque symptôme, l'existence d'une **entité symptomatique (e.s.)** assortie d'une **valeur** qui la qualifie (en général

dans un sens pathologique).

C'est ainsi que, lorsqu'on parle d'un "front étroit" ou d'une "asymétrie partielle de la face", on attribue à l'entité symptomatique "taille du front" la valeur "étroite" et à l'entité symptomatique "asymétrie de la face" la valeur "partielle".

L'entité symptomatique représente donc la localisation spatio-temporelle d'un symptôme.

Un symptôme  $s$  correspond alors à la réalisation de l'événement: "l'entité symptomatique  $e$  prend la valeur  $v$ ":

$$s = \{ e_s = v \}$$

La notion d'entité symptomatique est voisine de celle d'entité présentée par M. Fieschi (FIESCHI 83 p.60). Elle est cependant plus précise en ce qu'elle définit une classe particulière d'objets de raisonnement: les localisations anatomiques ou fonctionnelles, désignées, pour une ou plusieurs maladies, comme lieu possible d'anomalies.

L'entité symptomatique correspond grosso modo à la définition donnée par (JOUBERT 81, p. 65) d'une **variable sémantique** puisant ses valeurs dans un ensemble de valeurs possibles appelé **catégorie sémantique**.

Par exemple, l'e.s. "durée de la grossesse" peut avoir pour catégorie sémantique: {diminuée, normale, augmentée} ; pour l'e.s. "télorisme", on aura { hypo, normal, hyper}, etc.

Il est fréquent qu'une catégorie sémantique ait plusieurs formulations possibles, eu égard au degré de précision requis.

Ainsi, la durée de la grossesse, vue précédemment, pourra être (inférieure à 36 semaines, normale, supérieure à 42 semaines).

La définition de toutes les entités symptomatiques et de leurs catégories sémantiques respectives constitue un travail extrêmement long et minutieux pour le médecin. Il est à noter que ce travail constitue, d'une manière assez inattendue, un apport bénéfique important de l'informatique à la sémiologie médicale par l'effort parfois extrême de précision qu'il requiert dans chaque définition (DEROUESNE 80).

Quoi qu'il en soit, cette structuration du symptôme en un couple est à la base même de certains travaux, notamment les **analyses de données en génétique médicale** (Mc GIBBON 79), (PREUS 83), (RIALLE 85). On lira en annexe 2 un fragment de la liste des entités symptomatiques utilisées dans la Base de Données.

### II.5.2.2 AXE ORIENTE ASSOCIE A UNE ENTITE SYMPTOMATIQUE

Une entité symptomatique correspond à un axe sémantique orienté (appelé parfois par les sémioticiens espace axiologique (KERGOSIEN 84)), sur lequel s'ordonnent les différentes valeurs possibles des symptômes (symboles ordinaux).

Nous pouvons alors définir, pour chacune des entités symptomatiques, une **bijection** entre son axe sémantique et l'espace vectoriel  $R$  en attribuant, à chaque valeur ordinale, une valeur numérique dans un intervalle choisi **a priori**.

**Exemple** : tétorisme



en effet à toutes les localisations anatomiques, fonctionnelles, organiques, etc, du corps humain sur lesquelles ont été constatées des anomalies. Ces entités symptomatiques s'organisent d'une manière logique en une vaste **arborescence** dont la racine (stade 0) est l'entité "corps humain" avec, pour espace axiologique, l'espace paradigmatique binaire de référence (normal, anormal). Notons ici qu'il existe un grand nombre d'espaces axiologiques en relation paradigmatique avec celui-ci :

{bonne santé, mauvaise santé},  
{"ça va", "ça ne va pas"},  
{0,1},  
etc.

La racine "corps humain" se décompose ensuite en quinze branches ou noeuds (stade 1) figurant les quinze chapitres de la symptomatologie telle qu'elle a été définie précédemment (§ II.3)

Chaque noeud-chapitre du stade 1 de l'arborescence possède à son tour un espace axiologique banal (normal, anormal) et une subdivision en plusieurs branches conduisant aux noeuds du stade 2.

A partir du stade 2 de l'arborescence, les noeuds peuvent être, soit des entités symptomatiques terminales munies de leur espace axiologique propre, soit des entités symptomatiques non terminales, par conséquent munies à leur tour d'un espace axiologique banal et d'une décomposition en plusieurs branches.

La figure suivante donne une représentation simplifiée de l'arborescence théorique de la symptomatologie complète.



être possibles.

Dans l'application qui nous intéresse ici, ces deux conditions ne sont pas remplies en raison, d'une part, du faible nombre de niveaux des variables (en général égal à trois) et, d'autre part, du nombre souvent plus important d'entités symptomatiques que de syndromes candidats au diagnostic à discriminer.

Pour pallier ces deux difficultés, nous proposons une réduction des entités symptomatiques en super-entités symptomatiques de deux manières possibles :

1°) en établissant des **scores** obtenus en sommant les valeurs d'entités symptomatiques correspondant à une même fonction physiologique, à une même localisation anatomique, à un même stade d'examen, plus généralement ayant une "parenté" dans l'observation médicale.

Cette parenté, jointe à une signification commune de mesure d'intensité ou d'importance d'une grandeur clinique (intensité de la coloration des téguments, importance d'un volume d'organe,...) permet la sommation, donc la définition de la super-entité.

2°) Par une Analyse des Données préalable, principalement une Analyse en Composantes Principales, pouvant permettre de retenir des super-entités symptomatiques (celles du premier plan factoriel par exemple) ayant une signification médicale et obtenues par une combinaison linéaire des entités symptomatiques élémentaires.

Si la première composante principale ajoute 0,5 fois la tension artérielle à 2,8 fois l'écartement des yeux, on peut douter de sa signification en tant que super-entité, mais il n'est pas exclu que

## Chapitre II

l'analyse mathématique fournisse (au prix de l'élimination de certaines variables "peu chargées" dans les combinaisons linéaires des composantes principales) des super-entités acceptables sur le plan sémantique.

L'exemple que nous avons traité en II.6 et II.7 ne nous a pas permis de mettre à jour de telles combinaisons linéaires (en raison principalement des limitations de notre base de données). Nous restons néanmoins persuadés que des super-entités symptomatiques de ce type sont envisageables sur une base de données plus complète.

La considération de ces super-variables permet de vérifier les deux conditions énoncées ci-dessus :

- a) elles possèdent beaucoup plus de niveaux que les variables qui les composent,
- b) elles sont en nombre plus restreint que les variables primitives, et donc peuvent être en nombre inférieur à celui des syndromes étudiés.

Il faut noter ici que les projections des point-syndromes se font généralement dans les sous-espaces des entités de niveau 1, 2, 3 ou 4 d'observation (§ II.5.1.2), ce qui augmente d'autant le rapport: nombre de points/nombre d'axes.

### **Exemples de scores couramment utilisés en Médecine Infantile:**

- score d'APGAR
- score de SILVERMAN

### **II.5.3.3 CHOIX DES SUPER-ENTITES DE TYPE SCORE**

Il s'agit de répondre à la question : sur quel critère choisir

les entités symptomatiques à regrouper dans un même score ? En d'autres termes, comment repérer les localisations anatomiques, les fonctions physiologiques, etc, qui peuvent donner lieu à un score ?

L'un des intérêts majeurs des scores est la possibilité d'établir des liens transversaux (latéraux) à l'intérieur de l'arborescence des entités symptomatiques.

En effet, les liens de cette arborescence sont essentiellement des liens de parenté ascendante et descendante. Il existe pourtant, dans cette arborescence, nombre d'entités symptomatiques terminales de même finalité, telles que :

- la mobilité
- le développement
- la forme
- le nombre
- etc.

Les entités symptomatiques ne sont autres que des attributs affectant des localisations anatomiques ou fonctionnelles différentes.

Il est donc intéressant, du point de vue de l'analyse générale des données, de regrouper, au sein d'une même variable (le score), l'ensemble des entités symptomatiques reliées par un lien non plus de parenté, mais de **similarité** (lien paradigmatique).

Un score peut notamment mettre en relief une **fonction** au sein d'une arborescence d'entités morphologiques, telle que la fonction "articulation" (exemple 4 ci-après). Il peut également mettre en relief un concept dysmorphique.

C'est ainsi que, dans la branche des membres (niveau 1, sixième branche, ou chapitre VI de la nomenclature des entités

## Chapitre II

symptomatiques, § annexe 2) ont été regroupées, au sein du score référencé VI.D (axe numéro 61 de l'A.D., § annexe 3), toutes les entités symptomatiques relatives au concept d'**amputation** (exemple 1 ci-après).

### Exemple 1:

score VI.D            chapitre VI            concept: amputation

entités symptomatiques regroupées:

JAMBE-AMPUTATION  
ORTEIL-AMPUTATION  
MAIN-DOIGT-AMPUTATION  
NIVEAU NON PRECISE-AMPUTATION  
MEMBRE-NIVEAU NON PRECISE-AMPUTATION  
AVANT BRAS-AMPUTATION  
METACARPE-AMPUTATION

### Exemple 2:

score XIII.A        chapitre XIII

concept: indice de maturation osseuse

entités symptomatiques regroupées:

OS-QUALITE-SOLIDITE  
OS-QUALITE-OSTEOPOROSE  
OS-QUALITE-CALCIFICATION  
OS-MATURATION  
OS-CORTICALE

### Exemple 3:

score XIV.D        chapitre XIV

concept: fonction cérébrale

entités symptomatiques regroupées:

FONCTION CEREBRALE-LANGAGE  
FONCTION CEREBRALE-MOTRICITE  
FONCTION CEREBRALE-COMPORTEMENT  
FONCTION CEREBRALE-DEVELOPPEMENT PSYCHOMOTEUR  
FONCTION CEREBRALE-MOUVEMENT  
FONCTION CEREBRALE-NIVEAU MENTAL  
FONCTION CEREBRALE-SENSIBILITE

FONCTION CEREBRALE-TONUS MUSCULAIRE  
SYSTEME NERVEUX CENTRAL-FONCTION CEREBRALE

**Exemple 4:**

score XIII.I      chapitre XIII

concept: articulation

entités symptomatiques regroupées:

ARTICULATION-RADIOCUBITALE  
ARTICULATION-FEMOROTIBIALE  
ARTICULATION-COXOFEMORALE  
ARTICULATION-OMOHUMERALE  
ARTICULATION-RADIOHUMERALE

**II.5.3.4 PROBLEMES INHERENTS AU CHOIX D'UN SCORE ET SOLUTION  
RETENUE**

Le choix d'une s.e.s. de score est soumis à une contrainte non négligeable: la valeur nulle du score doit correspondre à un état non pathologique de la s.e.s.

Voici le problème: supposons que, dans un score sur l'état des membres supérieurs, nous ayons l'humérus raccourci (valeur -1), le radius normal (valeur 0), et le cubitus allongé (valeur +1), le score serait alors nul:  $((-1) + 0 + (+1))$ , malgré une pathologie évidente.

On voit donc se poser ici un problème aigu, inhérent au principe d'un score sur une échelle ordinale de valeurs qualitatives: la sommation des entiers relatifs images, sur l'axe vectoriel, des différentes valeurs qualitatives, peut conduire au score nul; ceci est en contradiction avec le principe du caractère non pathologique associé au score nul.

Ce problème est assez délicat à résoudre d'une manière satisfaisante. Une solution à écarter d'emblée consiste à choisir

des valeurs positives pour tous les symptômes (caractère normal -> 0, caractère diminué -> + 1, caractère augmenté -> 2). Dans ce cas, le score nul est bien réservé au caractère normal, mais ceci au prix de la perte complète du "sens" des entités symptomatiques: il devient en effet impossible de reconnaître si une valeur non nulle d'une e.s. témoigne d'une augmentation ou d'une diminution. On ne peut donc plus parler d'un axe **orienté** associé à l'e.s.

Pour conserver la signification de l'axe orienté, il est donc indispensable d'avoir des images numériques des symptômes de part et d'autre du zéro.

Une solution possible consiste à poser comme contrainte au choix d'une s.e.s. que **toutes les entités symptomatiques varient dans le même sens**, c'est-à-dire que les valeurs non nulles, dans un syndrome, soient ou toutes positives, ou toutes négatives.

Cette condition apparaît beaucoup moins contraignante lorsqu'on considère l'origine génique des dysmorphies ou des anomalies fonctionnelles : un gène délétère provoque en effet, le plus souvent, des malformations analogues dans des localisations anatomiques différentes. Le paradigme de la diminution par exemple affectera une catégorie précise d'os.

La solution qui a été retenue n'a pas été le score additif pur, dont l'élaboration aurait nécessité un délai supplémentaire important.

Nous nous sommes orientés vers le score additif des **valeurs absolues** des entités symptomatiques. Ce dernier score ne donne pas d'information sur le sens, positif ou négatif, de la sommation. Il

donne simplement une information sur le degré de "pathologicité" c'est-à-dire d'excentricité par rapport à l'état normal du patient. Son intérêt est d'être très facile à élaborer.

Le choix d'un score doit être validé par une analyse approfondie portant sur un ensemble de malades. Nous allons en décrire brièvement la méthode.

### II.5.3.5 LA ROC-ANALYSE

L'objectif d'un score est de permettre une prise de décision. Le score représente en cela un outil à part entière d'aide à la décision et justifie, à ce titre, la place de plus en plus importante qu'il occupe dans les travaux sur l'aide au diagnostic médical (DAPHY 84), (LABBE 85), (ROBERT 86).

Un score devient véritablement un outil de décision à partir du moment où il est doté d'un ou plusieurs **seuils** de discrimination.

Le seuil est une valeur limite entre deux zones qualitativement disjointes, telles que la maladie ou la non-maladie (diagnostic), le décès ou la guérison, des complications ou l'amélioration de l'état de santé (pronostic).

Ainsi, si  $S$  représente la valeur d'un score pour un individu donné,  $\alpha$  le seuil, et  $Z1$  et  $Z2$  les zones délimitées par ce seuil (cas d'un seul seuil), on a alors la règle alternative suivante:

Si  $S > \alpha$  alors  $Z1$  sinon  $Z2$

La qualité d'un score dépend de deux caractères principaux:

- son **pouvoir prédictif**
- son **seuil** (éventuellement plusieurs seuils)

Le pouvoir prédictif d'un score est analysé de la même manière que pour un examen biologique ou un simple symptôme (§ I.2.2).

Nous rappelons brièvement la méthode, adaptée à un score. Soit la matrice de décision suivante:

	Z1	Z2
S > α	a (P)	b (FP)
S < α	c (FN)	d (N)

Supposons, pour fixer les idées, que Z1 corresponde à une maladie et Z2 l'absence de cette maladie.

6 paramètres se dégagent de ce tableau:

1 - Le taux des "vrais positifs": sujets pour lesquels la valeur du score est en accord avec la maladie.

$$TVP = \frac{a}{a + c} = \text{sensibilité du score}$$

2 - Le taux des "vrais négatifs": sujets pour lesquels la valeur du score est en accord avec l'absence de maladie.

$$TVN = \frac{d}{b + d} = \text{spécificité du score}$$

3 - Le taux des "faux positifs": sujets pour lesquels la valeur du score indique à tort la maladie.

$$TFP = \frac{b}{b + d}$$

4 - Le taux des "faux négatifs" (sujets pour lesquels la valeur

du score indiquait à tort l'absence de maladie).

$$TFN = \frac{c}{c + a}$$

5 - La valeur prédictive positive du score.

$$VP+ = \frac{a}{a + b}$$

6 - La valeur prédictive négative.

$$VP- = \frac{d}{c + d}$$

Le pouvoir prédictif d'un score dépend à la fois de sa sensibilité et de sa spécificité. Ce pouvoir sera d'autant plus élevé que les sensibilité et spécificité seront proches de 1.

- Le seuil d'un score est déterminé par l'analyse de la courbe ROC (Receiver Operating Characteristic) (Mc NEIL 75) (METZ 78).

Les étapes de l'analyse sont les suivantes:

a) construction des histogrammes des scores: non malades (courbe A) et malades (courbe B). Voir la figure suivante: \_

## Répartitions

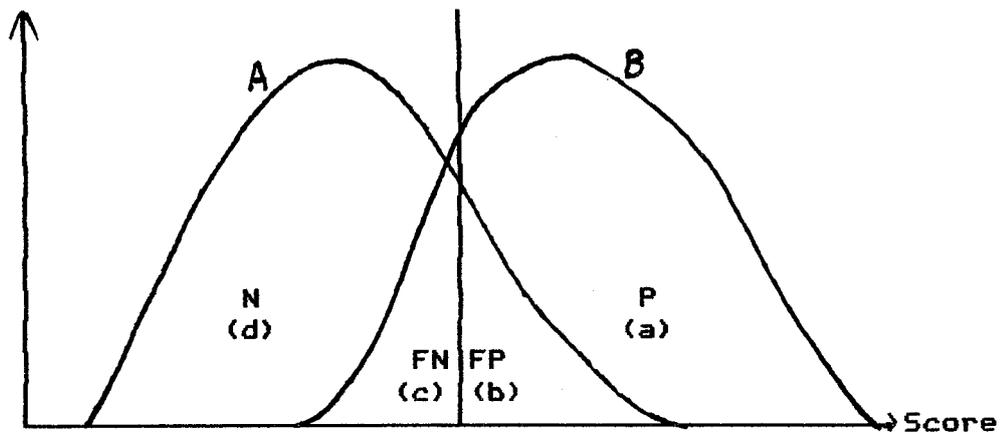


figure II.3

### b) construction et analyse de la courbe ROC

La courbe ROC est obtenue par balayage du double histogramme précédent par un seuil fictif, depuis la valeur minimale jusqu'à la valeur maximale du score.

Pour chaque valeur de ce seuil, on relève le pourcentage des faux positifs et des vrais positifs. On reporte ce point dans des axes de coordonnées (FP en abscisse, VP en ordonnée).

La courbe ROC est constituée par l'ensemble de ces points. Le schéma suivant en donne une illustration:

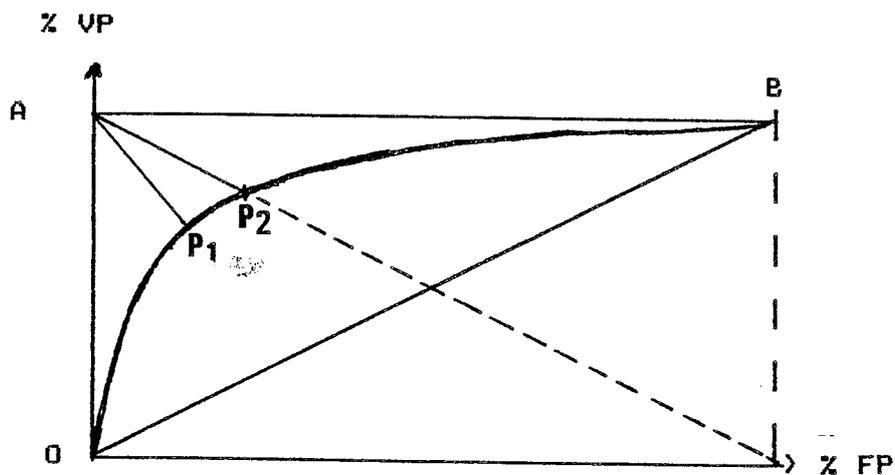


Figure II.4

- La courbe OAB représente la limite idéale de la courbe ROC. Elle correspond à une disjonction complète des 2 histogrammes avec un seuil situé entre les deux (pouvoir prédictif maximum du score).
- Le segment OB représente la limite inverse, c'est-à-dire celle pour laquelle le score n'a aucun pouvoir prédictif.
- La courbe ROC se situe entre ces deux extrêmes.

Le critère de sélection du seuil optimum, sur cette courbe, dépend du contexte dans lequel se situe le score.

D'une manière classique, étant donné que la courbe OAB représente un idéal, on retient pour seuil le point P1 (figure II.4) pour lequel la distance euclidienne à A est minimum.

On peut également retenir le point P2 correspondant à l'égalité entre les faux positifs et les faux négatifs. Dans le cas où le risque lié au choix positif est très grand (opération risquée par exemple) on choisira un seuil qui augmente sensiblement le taux des faux négatifs.

Dans le cas contraire (risque important lié au choix négatif), on choisira un seuil qui augmente le taux des faux positifs.

En conclusion, nous pouvons dire que la ROC-Analyse a pour objectif l'étude et la validation d'un score. La technique des scores et la ROC-Analyse constituent un domaine important de l'aide au diagnostic.

## II.5.4 BUTS DIAGNOSTIQUES : LES SYNDROMES

### II.5.4.1 LE CONCEPT DE SYNDROME

Un syndrome est une association de symptômes. Le dictionnaire médical nous en fournit la définition précise suivante: "Réunion d'un groupe de symptômes qui se reproduisent en même temps dans un certain nombre de maladies" (GARNIER 80). Le terme de syndrome est donc voisin du terme de maladie.

Une association de symptômes est reconnue comme étant la manifestation d'un syndrome lorsqu'il s'est dégagé dans la littérature un consensus pour individualiser cette association comme une nouvelle entité nosologique.

Nous rappelons (§ II.5.1.2) que l'ensemble des syndromes est subdivisé en quatre classes selon leur origine :

**Classe 1** : syndromes d'origine génique, maladies ou malformations entraînées par un gène délétère.

**Classe 2** : syndromes d'origine chromosomique, maladies ou malformations entraînées par une anomalie chromosomique.

**Classe 3 :** syndromes d'origine environnementale, accident survenu en cours de grossesse par exemple.

**Classe 4 :** syndromes d'origine inconnue.

#### II.5.4.2 DOUBLE REPRESENTATION D'UN SYNDROME

La problématique de la représentation tient au fait qu'un syndrome doit être manipulé par deux systèmes d'analyse différents: le médecin d'une part, le système d'aide au diagnostic d'autre part. Ces deux systèmes "raisonnent" selon des modalités différentes, et s'appuient chacun sur leur propre modèle de représentation des données. Ceci conduit aux deux représentations parallèles du syndrome suivantes:

##### II.5.4.2.1 REPRESENTATION SEMIOLOGIQUE

Cette représentation reflète la vision du médecin, dans laquelle un syndrome s'exprime sous la forme d'un ensemble de symptômes :

$$sd = \{s_1, \dots, s_n\}$$

Elle est caractéristique du champ médical dans lequel: \_

1°) toute maladie est caractérisée par un certain nombre de manifestations, soit directement observables, soit décelables au moyen d'un instrument médiateur.

2°) le diagnostic s'établit par analogie entre une série d'observations et une série de symptômes admis comme étant une représentation nosologique de la maladie.

#### II.5.4.2.2 REPRESENTATION VECTORIELLE

Cette représentation, adaptée à l'analyse numérique (module d'extraction de connaissances), rend compte des coordonnées d'un point-syndrome dans un espace vectoriel dont les axes de coordonnées sont constitués par les entités et les super-entités symptomatiques.

Un syndrome est donc un vecteur à m composantes correspondant aux m e.s. ou s.e.s. de l'espace.

$$sd = \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_m \end{bmatrix} = x_1 \vec{u}_1 + x_2 \vec{u}_2 + \dots + x_m \vec{u}_m$$

Afin d'homogénéiser l'expression du raisonnement, nous considérons que chaque axe représente une super-entité symptomatique. Sachant en effet qu'une s.e.s. est un regroupement significatif d'entités symptomatiques élémentaires, il n'est pas préjudiciable au raisonnement global de considérer qu'une e.s. est une s.e.s. composée uniquement d'elle-même.

On peut alors dire que  $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m$  constituent respectivement les vecteurs unitaires des axes  $ses_1, ses_2, \dots, ses_m$ .

#### II.5.4.3 RESOLUTION DES INFORMATIONS INCOMPLETES, COEFFICIENT DE PRECISION DESCRIPTIVE D'UN SYNDROME

La définition des syndromes est obtenue à partir de la littérature médicale. Ces définitions sont souvent assez vagues ou laissent une place importante à l'interprétation du médecin. Il

n'est pas rare en effet de voir apparaître, dans la définition sémiologique d'un syndrome, un symptôme non terminal. Nous avons vu qu'un symptôme non terminal correspond à la valeur **anormal** dans l'espace axiologique banal ({normal, anormal}) de l'entité symptomatique correspondante. Il est alors difficile, à partir d'une information aussi peu précise, de positionner un syndrome sur un axe réel.

Cette remarque est à l'origine du **coefficient de précision descriptive** associé à un syndrome, que nous allons décrire maintenant.

L'incertitude attachée à la définition d'un syndrome est un élément important dans la prise de décision : les syndromes dont la définition sémiologique est précise et complète devront être privilégiés par rapport à ceux dont la définition est imprécise, cette imprécision étant induite par la présence de symptômes non terminaux dans leur liste de symptômes.

Le **Coefficient de Précision Descriptive (CPD)** d'un syndrome est calculé à partir du nombre **t** de symptômes terminaux et du nombre **n** total de symptômes du syndrome. Sa formule est la suivante :

$$\text{CPD} = \frac{t}{n}$$

Le CPD varie donc entre 0 et 1 :

- il sera nul lorsque tous les symptômes seront non terminaux (définition totalement imprécise du syndrome),
- il sera égal à 1 lorsque tous les symptômes seront terminaux (définition totalement précise du syndrome par rapport au contexte

des symptômes).

Le CDP interviendra dans le choix des syndromes candidats (§ II.7.3). Ce poids aura donc une répercussion sur le choix de l'entité symptomatique à proposer à l'observation.

Enfin, précisons qu'un syndrome peut avoir plusieurs Coefficients de Précision Descriptive étant donné que le CPD dépend du niveau d'observation des symptômes. Il y aura donc, pour chaque analyse dans lequel le syndrome intervient, un CPD calculé en préalable à cette analyse. Cette dernière précision reflète le fait qu'un syndrome peut être plus ou moins bien décrit selon le niveau auquel on l'observe.

#### II.5.4.4 PASSAGE DE LA REPRESENTATION SEMIOLOGIQUE A LA REPRESENTATION VECTORIELLE D'UN SYNDROME

Le problème est le suivant : étant donné un syndrome défini par sa liste de symptômes, comment déduire les coordonnées de son point-syndrome dans l'espace complet des s.e.s.?

$$sd = \{ s_1, \dots, s_k \} \longrightarrow sd = \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_m \end{bmatrix}$$

Le passage d'une représentation à l'autre s'effectue d'une manière univoque: à une représentation sémiologique correspond une représentation vectorielle unique.

Le modèle vectoriel de représentation d'un syndrome, s'il rompt apparemment assez radicalement avec le modèle sémiologique, reste néanmoins en relation étroite avec l'analyse médicale des

symptômes.

La constitution d'abord des entités symptomatiques, ensuite des super-entités est, à l'évidence, régie par des critères médicaux:

- les diverses valeurs prises par les entités symptomatiques représentent des symptômes (SII.5.2),

- les regroupements d'e.s. en s.e.s. sont médicalement significatifs. S'il en était autrement, l'analyse des données perdrait son sens.

La solution consiste en un processus de 3 phases:

1 - établir la liste des couples (entité symptomatique, valeur) images des symptômes.

```
s1 ---) (es1, v1)
      .
      .
      .
sk ---) (esk, vk)
```

2 - rechercher les super-entités symptomatiques à partir des e.s. Cette opération consiste à regrouper les e.s. qui appartiennent à la même s.e.s.

```
ses1 = (esa1, [esb1],....)
      .
      .
      .
sesq = (esaq, [esbq],...)
```

(les [] signifient "éventuellement" et apparaissent à partir de la seconde e.s. de chaque s.e.s.).

3 - calculer les scores réalisés par chaque s.e.s. de type "score" dans le syndrome.

Le calcul consiste à faire la somme des valeurs absolues des

e.s. composantes de la s.e.s.

Exemple : ses1  $\rightarrow$  (es1, es2, es3) = (v1, v2, v3)

=> Score (ses1) = |v1| + |v2| + |v3|

## II.5.5 ESPACES DE TRAVAIL

### II.5.5.1 DEFINITIONS : ESPACE COMPLET - ESPACES CONJECTURAUX

L'espace de travail est un hyper espace à m dimensions, constitué par l'ensemble des m entités et super-entités symptomatiques impliquées dans les syndromes.

On distinguera deux types d'espaces de travail :

- l'espace complet
- les espaces conjecturaux

L'espace complet est constitué par la totalité des entités et super-entités symptomatiques incluses dans la Base de Données.

Cet espace servira essentiellement à la **pré-étude statique** du domaine complet (§ II.6).

Les **espaces conjecturaux** sont les sous-espaces de l'espace complet, mis en jeu au cours d'une recherche de diagnostic particulière. Ils correspondent chacun à un état, à un instant donné, du couple d'ensembles : (BASE D'OBSERVATIONS, SYNDROMES CANDIDATS).

### II.5.5.2 CONSTRUCTION DES ESPACES DE TRAVAIL

La méthode de construction d'un espace de travail est unique, quel que soit le type d'espace : complet ou conjectural. Elle est composée des étapes suivantes :

1 - Détermination de l'ensemble ESC des syndromes candidats à partir de l'ensemble 0 des observations. C'est l'opération R1 :

$$0 \xrightarrow{R1} ESC$$

avec  $ESC = \{sd / 0 \text{ inclus dans } sd\}$

où  $sd$  représente un ensemble de type "syndrome", à savoir une liste de symptômes.

2 - Détermination de l'ensemble SES des entités et super-entités symptomatiques mises en jeu par les syndromes candidats. C'est l'opération R2 :

$$ESC \xrightarrow{R2} SES$$

avec  $SES = \{ses / \text{il existe } sd \in ESC / (ses, v) \in sd\}$ .

Ces super-entités constituent l'ensemble des axes de l'espace de travail.

3 - Détermination des coordonnées des points-syndromes

Chaque syndrome candidat occupe, dans l'espace de travail, un point particulier dont les coordonnées sont les valeurs  $v$  des couples  $(ses, v)$  ou zéro.

$$sd = \begin{bmatrix} x1 \\ . \\ . \\ . \\ xm \end{bmatrix}, \text{ avec } xi = \begin{cases} 0 & \text{s'il n'existe pas } sesi \\ & / (sesi, xi) \in sd \\ v & \text{s'il existe } sesi \\ & / (sesi, v) \in sd \end{cases}$$

## II.6 PRE-ETUDE STATIQUE DU DOMAINE

La pré-étude statique (RIALLE 85) consiste en une exploration du domaine située hors du processus dynamique d'apprentissage, et en préalable à celui-ci.

Cette pré-étude a pour but la construction d'une partie de la connaissance directement déductible de la base de données.

Elle correspond à une analyse des données en cinq étapes, sur les cinq espaces suivants :

1ère étape : espace complet des entités symptomatiques

2ème étape : sous-espace de niveau d'observation 1

3ème étape : sous-espace de niveau d'observation 2

4ème étape : sous-espace de niveau d'observation 3

5ème étape : sous-espace de niveau d'observation 4

Pour chacune de ces étapes, nous procédons de la façon suivante:

1 - Constitution de la matrice (syndromes  $\times$  entités symptomatiques). Cette matrice possède:

- en ligne: tous les syndromes de la base qui ont des composantes dans le niveau d'observation considéré,

- en colonne: les coordonnées des syndromes dans le sous-espace de niveau considéré (sauf pour la première étape correspondant à l'espace complet).

2 - Construction de la matrice de variances-covariances correspondante.

3 - Diagonalisation de cette matrice (nous avons utilisé l'algorithme de Jacobi) avec l'objectif suivant:

## Chapitre II

4 - Recueil et analyse des valeurs propres et des vecteurs propres de cette matrice.

Chaque analyse fournit:

a) La liste des valeurs propres et des pourcentages simples et cumulés d'inertie (§ annexe 9).

b) La liste des principaux vecteurs propres normés correspondant à au moins 80 % de l'inertie totale. Un vecteur propre est exprimé par la liste de ses composantes sur les axes du niveau d'observation considéré. Appelons **contributions** les valeurs des composantes de chaque vecteur propre.

c) Pour chaque vecteur propre, la liste ordonnée des axes qui ont la plus grande contribution.

Le critère de sélection de ces axes est le suivant:

- les axes sont ordonnés (Tri rapide ou par permutation par exemple) de façon décroissante sur leur contribution dans le vecteur propre (ces axes sont nommés A, suivi du numéro de l'axe, exemple: A50, A41, etc. La liste complète des axes se trouve en annexe 1).

- on retient les n premiers axes tel que la somme des carrés des contributions atteigne un seuil fixé expérimentalement (0,85 du carré de la norme euclidienne du vecteur propre, par exemple). Ce seuil permet de sélectionner, pour chaque vecteur propre, de 1 à 10 axes principaux. Nous ne dépassons pas les 10 premiers, étant donné la faiblesse des contributions rencontrées au-delà de ce chiffre. Nous appelons **axes pertinents** les axes retenus par cette

## Chapitre II

méthode.

d) La sélection des axes les plus pertinents s'obtient par un tri de type calcul de scores sur les puissances de 2 (ordre induit par la numération binaire) dont le principe est le suivant:

Soit la configuration suivante des listes ordonnées de contributions pour chaque vecteur propre:

<b>V1</b>	<b>V2</b>	<b>V3</b>	<b>... etc.</b>
$\begin{bmatrix} \cdot \\ \cdot \\ \mathbf{A5} \\ \cdot \\ \cdot \\ \mathbf{A1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{A2} \end{bmatrix}$	$\begin{bmatrix} \cdot \\ \cdot \\ \mathbf{A3} \\ \cdot \\ \cdot \\ \mathbf{A2} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{A4} \end{bmatrix}$	$\begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{A1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{A3} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{A2} \end{bmatrix}$	

AXE \ p.2	V1 7	V2 6	V3 5	...	V8 0
A1	x		x		
A2	x	x	x		
A3		x	x		
A4		x			
A5	x				

On calcule, pour chaque axe, la somme de puissances de 2 représentant son **pouvoir discriminant**. Soit, dans l'exemple (si l'on s'arrête aux 3 premiers vecteurs propres et aux 8 premières puissances de 2):

$$A1 = 2^7 + 2^5 = 150$$

## Chapitre II

$$A2 = 2^7 + 2^6 + 2^5 = 214$$

$$A3 = 2^6 + 2^5 = 96$$

$$A4 = 2^6 = 64$$

$$A5 = 2^7 = 128$$

d'où l'ordre suivant: A2, A1, A5, A3, A4.

La suite de la méthode est exposée en traitant notre propre exemple de base de données. Il faut rappeler ici que l'exemple est une simple illustration de la méthode, afin de montrer que celle-ci a été pratiquée concrètement. Néanmoins, nous insistons sur le fait que la base de données, malgré ses quelque 800 éléments syndromiques et symptomatiques, reste trop étroite pour fournir des résultats absolument probants. On s'attachera donc essentiellement à la méthodologie proposée.

Nous présentons l'analyse du niveau 1 d'observation en raison de son importance majeure par rapport aux autres niveaux.

### **Niveau d'observation 1:**

Les axes de ce niveau sont au nombre de 67 nommés A1, A2, ..., A67 (ce sont les 67 premiers axes de l'annexe 1). Les axes pertinents retenus au cours de l'analyse des 12 premiers vecteurs propres sont exposés dans le tableau suivant:

Chapitre II

VECTEUR PROPRE	% INERTIE	% INERTIE CUMULEE	AXES MIS EN EVIDENCE ET CONTRIBUTIONS	
1	39,80	39,80	A60	0,99
2	11,37	51,18	A62	0,74
			A59	0,44
			A63	0,35
3	8,36	60,04	A59	0,70
			A63	0,40
4	5,53	65,57	A40	0,46
			A34	0,44
			A46	0,34
			A33	0,29
			A63	0,23
5	4,19	69,75	A63	0,43
			A45	0,41
			A40	0,22
6	2,94	72,69	A54	0,35
			A42	0,33
			A45	0,25
			A44	0,20
7	2,69	75,39	A46	0,26
			A32	0,21
			A34	0,18
			A59	0,16
8	2,43	77,82	A46	0,40
			A63	0,26
			A51	0,23
9	2,22	80,04	A51	0,44
			A45	0,36
			A46	0,28
10	1,82	81,86	A42	0,37
11	1,80	83,66	A33	0,37
			A34	0,26
			A64	0,22
12	1,61	85,27	A40	0,35
			A46	0,22
			A54	0,21

L'algorithme de tri décroissant sur les puissances de 2 donne la liste suivante pour les axes pertinents de niveau 1:

Rang	Axe	Type	VECTEURS PROPRES											
			1	2	3	4	5	6	7	8	9	10	11	12
1	A60	S	x											
2	A63	S		x	x	x	x				x			
3	A59	S		x	x					x				
4	A62	S		x										
5	A40	S				x	x							x
6	A46	S				x				x	x	x		x
7	A34	S				x				x				x
8	A33	S				x								x
9	A45	S						x	x			x		
10	A42	S							x				x	
11	A54	S							x					x
12	A44	S							x					
13	A32	S								x				
14	A51	S									x	x		

Nous faisons un premier constat: l'analyse des 12 premiers vecteurs propres donnés par l'algorithme de Jacobi mettent en évidence, sur les 67 axes que contient le niveau 1, 14 axes pertinents de première importance.

**N.B.:**

1) le fait que ce soient des scores qui possèdent les meilleurs pouvoirs discriminants est logique puisqu'ils possèdent un plus grand nombre de valeurs naturelles ainsi que des valeurs absolues supérieures à 1 (et donc des variances) plus importantes que dans le cas d'entités symptomatiques isolées. Les variances importantes témoignent de l'étalement des valeurs des syndromes sur l'axe.

2) La mise en évidence de ces axes peut être vérifiée en reprenant les 12 premiers vecteurs propres de niveau 1 (annexe 10).

## Chapitre II

L'ensemble de ces axes constitue un espace vectoriel de bonne qualité pour la projection des syndromes de la base.

Nous proposons ensuite de pratiquer une exploration du nuage de syndromes en projection sur les plans déterminés par les axes pertinents pris deux à deux.

Les plans de projection sont donc les suivants:

PLAN	AXE ABSCISSE	AXE ORDONNEE	Syndromes à l'origine	Syndromes représentés
1	A60	A63	E11	E12
2	A59	A62	E21	E22
3	A40	A46	E31	E32
4	A34	A33	E41	E42
5	A45	A42	E51	E52
6	A54	A44	E61	E62
7	A32	A51	E71	E72

**TABLEAU DES PLANS DISCRIMINANTS**

Les deux dernières colonnes du tableau précédent donnent les dénominations des ensembles dichotomiques de syndromes. Ces ensembles seront utilisés dans les règles de dichotomie exposées plus loin.

Les graphiques des projections sont donnés en annexe 11, les ensembles dichotomiques sont donnés en annexe 17.

L'expert-médecin peut retirer, de ce mécanisme d'Analyse en Composantes Principales, un certain nombre de connaissances dont nous donnons quelques exemples à titre indicatif et sous forme de règles de production :

Si **score 60** > 9 alors syndrome de ptérygium multiple

Si **score 63** > 5 alors syndrome de Charlie M.

## Chapitre II

Si **score 60** et **score 63** nuls alors fente palatine

(vraisemblance = 0.7)

Si **score 59** > 6 alors syndrome de Charlie M.

Si **score 62** > 9 alors syndrome d'Apert ou ptérygium poplité

Si **score 63** = 0 alors anomalie chromosomique

(vraisemblance = 0.6)

etc.

D'une manière plus générale, nous pouvons dire que :

1°- un score permet de dégager un ou plusieurs syndromes isolés pour lesquels il présente une valeur maximum,

2°- un couple de scores, s'il est nul, permet d'isoler un groupe important de syndromes (45 dans le cas du couple d'axes 60 et 63, § annexe 11). Un tel couple permet par conséquent une approche souvent dichotomique excluant d'emblée une moitié des syndromes. La répétition de cette procédure dichotomique pour la succession de couples de scores dans l'ordre décroissant des pouvoirs discriminants (correspondant aux couples de valeurs propres décroissantes) permet une convergence rapide du processus de restriction progressive de l'éventail des syndromes restant candidats.

Cette approche dichotomique, exposée en II.7.3, sera effectuée au moyen de **règles de dichotomie**.

Nous exposons ci-après les 7 règles de dichotomie (RD1, ..., RD7) issues des 7 plans de projection précédents.

**Préalables** : désignons dans la suite par :

## Chapitre II

- 1 - score ( X, Y ), la fonction qui attribue à l'ensemble Y de symptômes, le score ou la valeur obtenue sur l'axe X.
- 2 - BF, l'ensemble des symptômes de la base de faits.
- 3 - ESC, l'ensembles des syndromes candidats.
- 4 - Eij, l'ensemble de syndromes du plan i dans la situation j (voir tableau des plans discriminants précédent).

### Règle RD1 :

```
si score ( A60, BF ) = 0 et score ( A63, BF ) = 0
alors ESC <-- E11
sinon ESC <-- E12
```

### Règle RD2 :

```
si score ( A59, BF ) = 0 et score ( A62, BF ) = 0
alors ESC <-- ESC  $\cap$  E21
sinon ESC <-- ESC  $\cap$  E22
```

### Règle RD3 :

```
si score ( A40, BF ) = 0 et score ( A46, BF ) = 0
alors ESC <-- ESC  $\cap$  E31
sinon ESC <-- ESC  $\cap$  E32
```

### Règle RD4 :

```
si score ( A34, BF ) = 0 et score ( A33, BF ) = 0
alors ESC <-- ESC  $\cap$  E41
sinon ESC <-- ESC  $\cap$  E42
```

### Règle RD5 :

```
si score ( A45, BF ) = 0 et score ( A42, BF ) = 0
alors ESC <-- ESC  $\cap$  E51
```

## Chapitre II

**sinon** ESC <-- ESC n E52

**Règle RD6 :**

**si** score ( A54, BF ) = 0 **et** score ( A44, BF ) = 0

**alors** ESC <-- ESC n E61

**sinon** ESC <-- ESC n E62

**Règle RD7 :**

**si** score ( A32, BF ) = 0 **et** score ( A51, BF ) = 0

**alors** ESC <-- ESC n E71

**sinon**-ESC <-- ESC n E72

En général, la partition ainsi opérée n'a malheureusement pas de sens évident pour le généticien, et ne peut donner très souvent des règles qui augmentent son savoir et qui permettent par exemple de rassembler sous une même étiquette (correspondant à un même stade embryologique des malformations, à une même étiologie, à une anomalie du même chromosome, etc...) les syndromes éliminés à chaque étape.

Nous avons donné ci-dessus quelques exemples de règles nouvelles apparues grâce à l'Analyse en Composantes Principales. Une étude exhaustive du domaine à l'aide de cette analyse permettrait d'accroître cet ensemble de règles.

### II.7 PRODUCTION DYNAMIQUE DE CONNAISSANCES NOUVELLES ET APPRENTISSAGE

#### II.7.1 POSTULAT DE BASE

La production dynamique de connaissances (RIALLE 86) s'effectue

## Chapitre II

au cours du processus d'interrogation et de recueil des données sémiologiques du patient.

Chaque étape de l'observation (nouvelle configuration de l'ensemble des symptômes - base de faits) donne lieu à l'interrogation interne suivante, de la part du système: la configuration actuelle de la base de fait est-elle totalement nouvelle, ou a-t-elle été observée au moins une fois dans le passé ?

La méta-règle appliquée, constituant le postulat de base, est la suivante:

Si la configuration actuelle de la base de faits est inconnue du système  
Alors déclencher une analyse des données  
Sinon déclencher la connaissance correspondante

Cette méta-règle correspond, du point de vue du système, à l'idée suivante:

- un tableau symptomatique se présente;
- si l'on a déjà rencontré ce tableau,
- alors on sait ce qu'il faut faire et l'on applique la conduite à tenir,
- sinon on l'étudie en profondeur, on applique la solution trouvée, on mémorise le couple (base de faits, solution).

**N.B.:** le couple (base de faits, conséquence) est soit un couple (base de faits, e.s. à observer), soit un couple (base de faits, syndrome(s)).

Cette règle traduit une conduite antropomorphe bien connue: agir si l'on sait, apprendre si l'on ne sait pas.

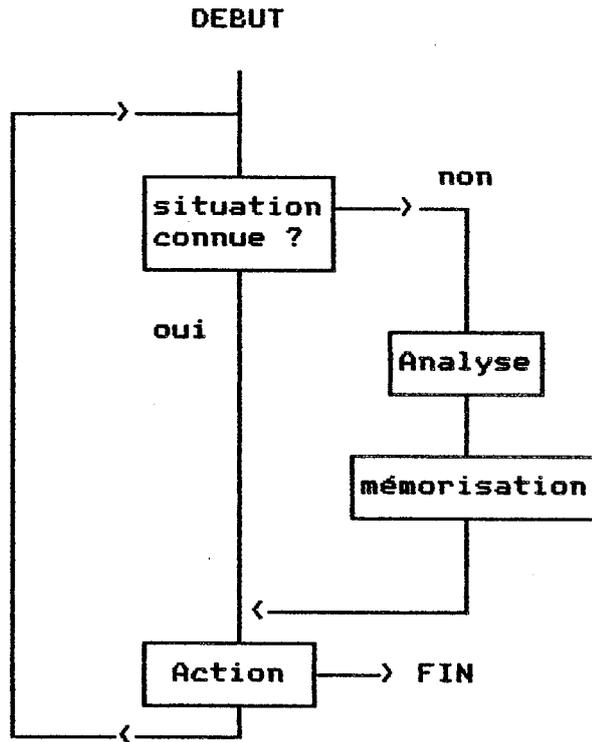
### II.7.2 CYCLE THEORIQUE

Le cycle complet et idéal correspond à l'étude approfondie d'un

## Chapitre II

malade, depuis l'observation initiale jusqu'à la conclusion diagnostique.

L'organigramme simplifié du cycle théorique est le suivant :



Le cycle reste théorique et idéal en ce qu'il constitue une proposition de modèle d'un processus très complexe de décision.

La branche médiane du cycle (situation connue) correspond à la démarche classique d'application de connaissances appropriées à des situations maîtrisées par un expert.

La branche droite du cycle (situation non connue) correspond au processus de découverte et d'apprentissage.

- La découverte est effectuée au moyen d'une analyse approfondie de la situation. Toute méthode d'analyse d'une situation, en relation

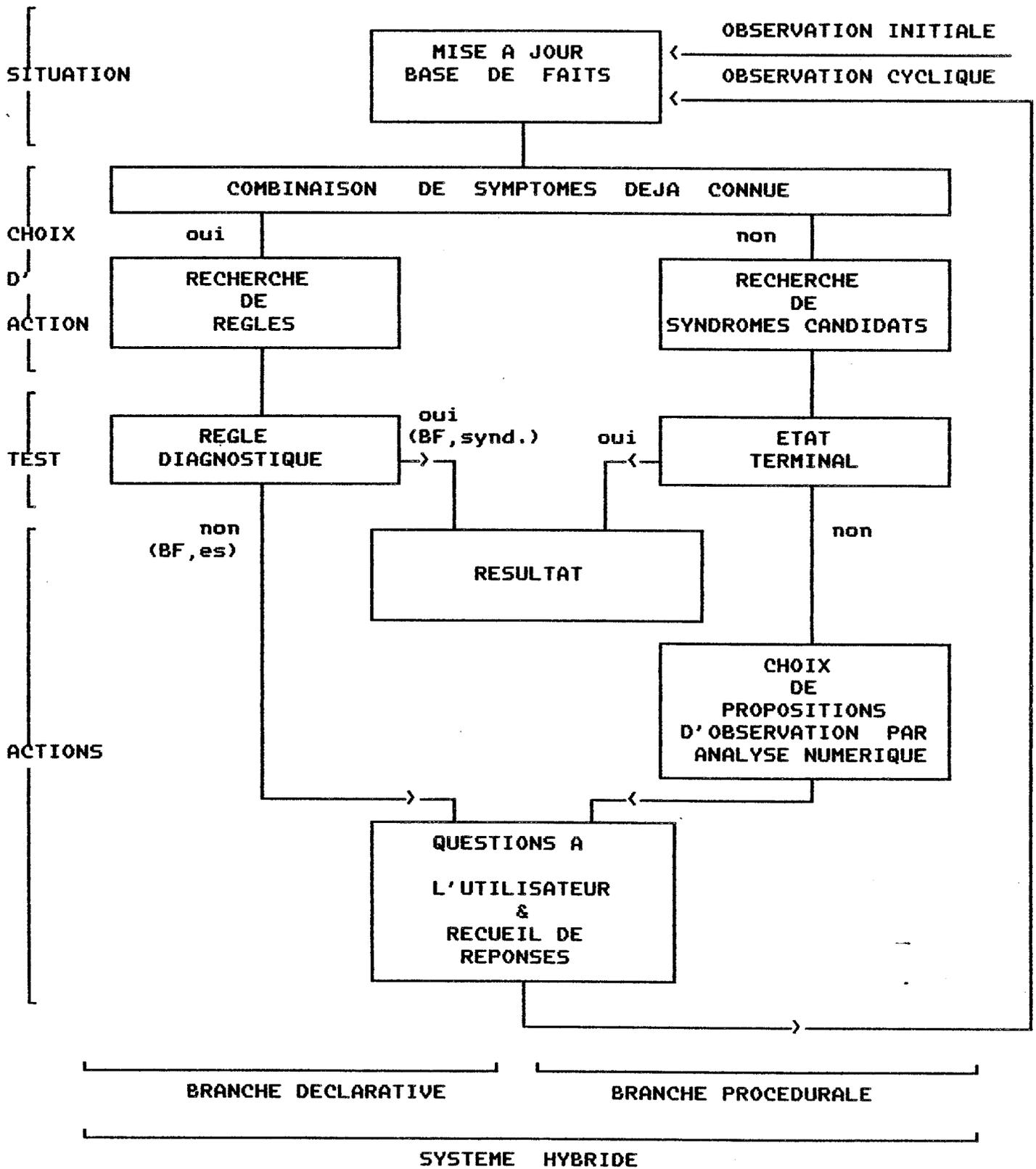
## Chapitre II

avec le domaine d'application, peut avoir sa place à cet endroit. Nous avons retenu une méthode mathématique exposée et justifiée en II.4. Quelle qu'elle soit, l'analyse doit avoir pour but la détermination de l'action qui répondra le mieux à la situation analysée, selon le but recherché.

- L'apprentissage correspond à la mémorisation d'un couple (situation donnée, action appropriée).

Ainsi, un organigramme plus détaillé du système est le suivant:

Chapitre II



II.7.3 METHODOLOGIE PROSPECTIVE

## Chapitre II

La méthodologie exploratoire en situation d'apprentissage qui va être exposée s'applique, rappelons-le, dans l'hypothèse d'une base de données qui soit la plus complète possible, c'est-à-dire comportant plusieurs milliers de syndromes et de symptômes. Les exemples cités ont une valeur purement illustrative de la méthode. Celle-ci comporte un aspect procédural et un aspect déclaratif.

L'analyse des projections des syndromes sur les couples d'axes pertinents a révélé, dans la pré-étude statique du domaine, l'existence de 2 groupes remarquables de syndromes sur chaque plan:

- a) les syndromes situés sur l'origine des deux axes - coordonnées (0,0) -
- b) les syndromes qui ont l'une au moins de leurs coordonnées non nulle (syndromes représentés sur l'un ou l'autre des axes, ou sur les deux).

Partant de cette constatation, nous proposons une approche dichotomique pour la réduction de l'éventail des syndromes candidats, selon la démarche suivante:

**PREMIERE PHASE:** analyse de niveau 1 d'observation (clinique immédiat) - génération d'hypothèses diagnostiques.

### 1) Initialisation de la base de faits :

- le médecin transmet au système la liste des symptômes qu'il a observés sur le malade,
- le système complète cette liste par les symptômes situés hiérarchiquement au-dessus des symptômes observés (voir l'arborescence générale des symptômes en II.5.1.4). L'ajout de tels symptômes - moins précis que les symptômes terminaux de l'arborescence - est indispensables pour une sélection exhaustive des syndromes candi-

## Chapitre II

dats. En effet, nous rappelons (§ II.5.4.3) que certains syndromes sont définis, dans la littérature, d'une manière imprécise, et que cette imprécision se traduit par des symptômes non terminaux dans l'arbre.

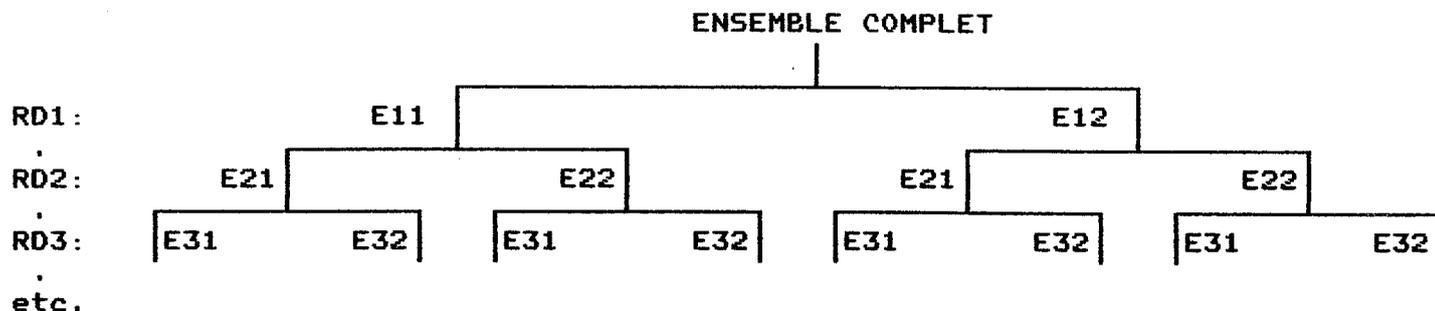
- Les axes mis en évidence par l'analyse a priori des vecteurs propres de l'espace complet (§ II.6) ont un caractère particulièrement discriminant dans la recherche du diagnostic. On a donc intérêt à renforcer l'observation du malade en fonction de ces axes. Le système propose donc au médecin une série d'observations complémentaires - les symptômes portés par les 14 axes pertinents de niveau 1, dans notre exemple - de manière à s'assurer que les scores correspondants ont été pris en compte.

**N.B.:** du point de vue ergonomique - sur le poste de travail - la réponse à cette demande de compléments d'observation pourra être facilitée par le déplacement manuel d'un curseur mobile (de type "souris" par exemple) sur une liste de symptômes, évitant ainsi la frappe au clavier de symptômes longs à transcrire.

**2) réduction de l'éventail des syndromes candidats par déclenchement des règles de dichotomie:**

à la suite de l'initialisation, les règles de dichotomie, élaborées lors de la pré-étude du domaine, sont appliquées dans l'ordre décroissant des pouvoirs discriminants des axes. Il y a donc, dans notre exemple, déclenchement systématique et ordonné des règles RD1 à RD7.

Cette enchaînement de règles réalise les intersections successives des ensembles de syndromes des plans de discrimination. L'arborescence des intersections peut être symbolisée par le schéma suivant:



### ARBRE DES INTERSECTIONS

Le résultat de cette opération est un ensemble E de syndromes tel que :

$$E = E1a \cap E2b \cap E3c \cap E4d \cap E5e \cap E6f \cap E7g$$

les lettres a, b, c, d, e, f et g étant mises pour les chiffres 1 ou 2 et représentent donc une des deux possibilités de cheminement à chaque embranchement de l'arbre décisionnel.

IL existe, théoriquement, 128 ensembles-intersection possibles. Pratiquement, ce nombre doit être diminué à cause des occurrences possibles d'intersections vides dans l'arbre des intersections.

La chaîne des règles RD1, ..., RD7 correspond au cheminement guidé par la base de faits dans l'espace de recherche d'intersections créé par l'Analyse des Données.

Il existe trois types de chemins depuis la racine (ensemble complet des syndromes) jusqu'aux feuilles :

1 - les chemins conduisant à une intersection comportant de nombreux syndromes.

2 - les chemins conduisant à un singleton (ou à un ensemble de syndromes jugé satisfaisant pour être proposé comme résultat de l'investigation): le syndrome correspondant peut alors être proposé comme diagnostic,

3 - les chemins conduisant à une intersection vide: la base de

faits ne peut conduire à aucun syndrome de la Base de Données. Ce cas peut correspondre à l'occurrence d'un nouveau syndrome.

Les types 2 et 3 conduisent à un arrêt du processus, alors que le type 1, plus courant, donne lieu à un éventail réduit de syndromes transmis à la deuxième phase de l'analyse.

**3) Recherche complémentaire de syndromes candidats** par analyse de proximités :

L'opération suivante est une analyse des proximités entre les points-syndromes et le point-malade.

A la suite de cette analyse, sont retenus comme "candidats", les points-syndromes dont la distance avec le point-malade est inférieure à un seuil fixé expérimentalement.

Ce seuil correspond au rayon de la "sphère", centrée sur le point-malade, enveloppant les syndromes candidats; étant donné son importance et la difficulté de fixer sa valeur *a priori*, le système offrira la possibilité de faire varier interactivement ce seuil afin de donner une idée précise de l'environnement syndromique du point-malade (base de faits). Le système attribue, à titre indicatif, une valeur par défaut égale au tiers de la plus grande distance existant entre point-malade et syndromes.

Un tel seuil peut être comparé à une lunette d'observatoire qui regarde plus ou moins loin dans l'espace des syndromes environnants.

La distance retenue est la distance euclidienne (racine carrée de la somme des carrés des distances des projections sur chaque axe).

Les syndromes retenus seront classés dans l'ordre décroissant

## Chapitre II

de leur distance au point malade multipliée par leur Coefficient de Précision Descriptive.

L'ensemble A1 des syndromes candidats, déduits par proximité avec la base de faits, peut être complété par l'ensemble A2 des syndromes sélectionnés directement par coïncidence avec la base de faits. Cette opération consiste à ne retenir que les syndromes proches, qui possèdent dans leur liste de symptômes ceux de la base de faits.

Si l'ensemble  $A2 \setminus A1$  est non vide, on déclenchera une procédure d'examen des axes (entités isolées ou scores) qui sont très chargés par les syndromes situés dans cet ensemble.

La recherche de syndromes candidats par proximité, ainsi que la recherche par coïncidence avec la base de faits, sont deux méthodes qui se renforcent mutuellement. Dans ce sens, il est possible d'envisager d'autres méthodes éventuellement utilisables sous certaines conditions. Il n'est donc pas exclu de faire intervenir dans ce processus algorithmique un processus complémentaire de type déclaratif pour renforcer la phase de génération d'hypothèses diagnostiques.

Nous obtenons donc, au terme de cette première phase d'analyse, une sélection complète de syndromes candidats. Cette sélection peut être vide ou n'avoir que très peu de syndromes. Dans ce cas, le système propose au médecin les résultats de son investigation. Il y a alors production d'une connaissance nouvelle, dans la mesure où l'observation conduit à un état terminal du processus (voir organigramme en II.7.2), c'est-à-dire l'échec ou la proposition de diagnostics. Le médecin décide ensuite de retenir ou non cette nouvelle connaissance.

## Chapitre II

Voici, sur un exemple, une réduction de l'éventail des syndromes candidats :

soit le tableau clinique suivant :

symptômes observés:	axe d'appartenance:
PIED BOT	60
RACINE DU NEZ LARGE	40
EXTREMITE DU NEZ LARGE	40
OREILLES BASSES	34
MICROGNATHIE	34
ENOPHTALMIE	45
MICROCEPHALIE	32

Soit ESC l'ensemble des syndromes candidats, les syndromes sont donnés par leur numéro, leur nom étant consultable en annexe 12.

règle : RD1

choix : E12 (car axe 60 chargé)

ESC :

1	2	4	9	10	11	19	20	21	22	25	26	27	28	29	30	32	33
36	37	38	39	43	44	45	47	50	52	53	57	58	60	62	69	70	71
72	74	76	77	78	79	80	82	83	85	86	87	89	92	93	94	95	96
100	101	102															

effectif : 57

règle : RD2

choix : E21 (car aucun des deux axes 59 et 62 n'est chargé)

intersection : E12  $\cap$  E21

ESC :

1	9	10	19	20	21	22	26	30	32	36	38	43	44	45	47	50	57
60	62	69	70	76	77	78	79	82	83	85	86	87	89	92	94	95	96
100	102																

effectif: 38

## Chapitre II

règle : RD3

choix : E32 (car axe 40 chargé)

intersection : E12  $\cap$  E21  $\cap$  E32

ESC :

9	19	22	26	30	32	36	38	44	45	50	57	69	70	76	77	78	79
82	83	85	87	89	94	95	96	100									

effectif : 27

règle : RD4

choix : E42 (car axe 34 chargé)

intersection : E12  $\cap$  E21  $\cap$  E32  $\cap$  E42

ESC :

9	19	26	30	32	44	45	57	69	77	78	79	82	85	87	89	94	95
96	100																

effectif : 20

règle : RD5

choix : E52 (car axe 45 chargé)

intersection : E12  $\cap$  E21  $\cap$  E32  $\cap$  E42  $\cap$  E52

ESC :

9	19	26	30	32	44	45	57	77	78	79	82	85	87	89	94	95	96
100																	

effectif : 19

règle : RD6

choix : E61 (car aucun des deux axes 54 et 44 n'est chargé)

intersection : E12  $\cap$  E21  $\cap$  E32  $\cap$  E42  $\cap$  E52  $\cap$  E61

ESC :

30	32	44	57	77	78	82	85	94	95	96	100						
----	----	----	----	----	----	----	----	----	----	----	-----	--	--	--	--	--	--

effectif : 12

règle : RD7

choix : E72 (car axe 32 chargé)

intersection : E12  $\cap$  E21  $\cap$  E32  $\cap$  E42  $\cap$  E52  $\cap$  E61  $\cap$  E72

ESC :  
32 77 78 82 95 96 100

effectif : 7

## DEUXIEME PHASE

La seconde phase de cette investigation a pour but d'analyser, sous l'angle de vue du second niveau d'observation (clinique approfondi), la sélection de syndromes résultant de la première phase.

Cette étude est centrée sur une Analyse en Composantes Principales des syndromes candidats, dans le sous-espace vectoriel des axes du second niveau d'observation. Les deux premiers vecteurs propres de l'ACP canonique (§ II.4.2) vont mettre en évidence les axes les plus discriminants (axes pertinents). La méthode de recherche des axes pertinents consiste à retenir, en nombre inférieur à 4 ou 5, les axes qui ont les plus fortes contributions dans le premier vecteur propre, puis dans le second vecteur propre. Cette sélection d'axes pertinents conduit à une demande d'informations complémentaires - confirmation ou infirmation des symptômes portés par ces axes - en direction du médecin.

Dans l'exemple proposé, l'Analyse en Composantes Principales met en évidence la super-entité symptomatique de type score correspondant à l'axe numéro 87. Ce score donne en effet la contribution la plus importante dans le premier vecteur propre (il correspond également à la plus forte contribution dans les second et troisième vecteur propres). On trouvera en annexe 19 le compte rendu de cette A.C.P.

L'axe numéro 87 correspond au score XIV.D : FONCTION CEREBRALE.

## Chapitre II

La liste des symptômes appartenant à ce score (§ annexe 20) est donc proposée pour un examen approfondi au médecin. Celui-ci confirmera ou infirmera chacun des symptômes. Un nouvel état de la base de faits en résultera qui conduira soit à un syndrome unique (ou à un ensemble de syndromes jugé satisfaisant pour être proposé comme résultat de l'investigation) soit à un nouvel éventail de syndromes candidats transmis à l'analyse du niveau 3 d'observation (paraclinique standard). Dans ce dernier cas, une analyse identique à celle du niveau 2 est alors pratiquée. Le processus a ainsi la possibilité d'être itéré jusqu'au niveau 4 d'observation (paraclinique approfondi).

Chapitre II

## CHAPITRE III

### SYSTEME HYBRIDE PROCEDURAL-DECLARATIF D'APPRENTISSAGE ET D'AIDE AU DIAGNOSTIC - PROJET GENESYS

#### III.1 PRINCIPE ET SCHEMA GENERAL DU SYSTEME

Le principe du système est le suivant:

- a - en Génétique Médicale, une connaissance complète du sujet, de type déclaratif (a) n'est pas envisageable (§ II), d'où :
- b - la nécessité d'une méthodologie d'apprentissage et de production de connaissances, de type procédural (b), telle qu'on l'a définie dans le chapitre II.

Le projet proposé consiste à associer dans une seule et même démarche diagnostique les connaissances de type a et celles de type b.

L'idée sous-jacente à ce système est que la pertinence de la connaissance est indépendante de son mode d'élaboration: mode intuitif, mode expérimental, mode mathématique, etc.

Dès l'instant qu'une connaissance est validée, elle devient aussitôt exploitable, et susceptible de concourir à la solution d'un problème nouveau. Ce problème produira à son tour des connaissances nouvelles et ainsi de suite.

D'autre part une connaissance élaborée à partir d'une analyse numérique, dès lors qu'elle est acceptée en tant que connaissance experte, est soumise aux règles de cette catégorie de

confrontation avec d'autres connaissances du même type ou d'autres types, acceptation ou rejet par d'autres experts, etc.

Ainsi, le schéma général du système est le suivant :

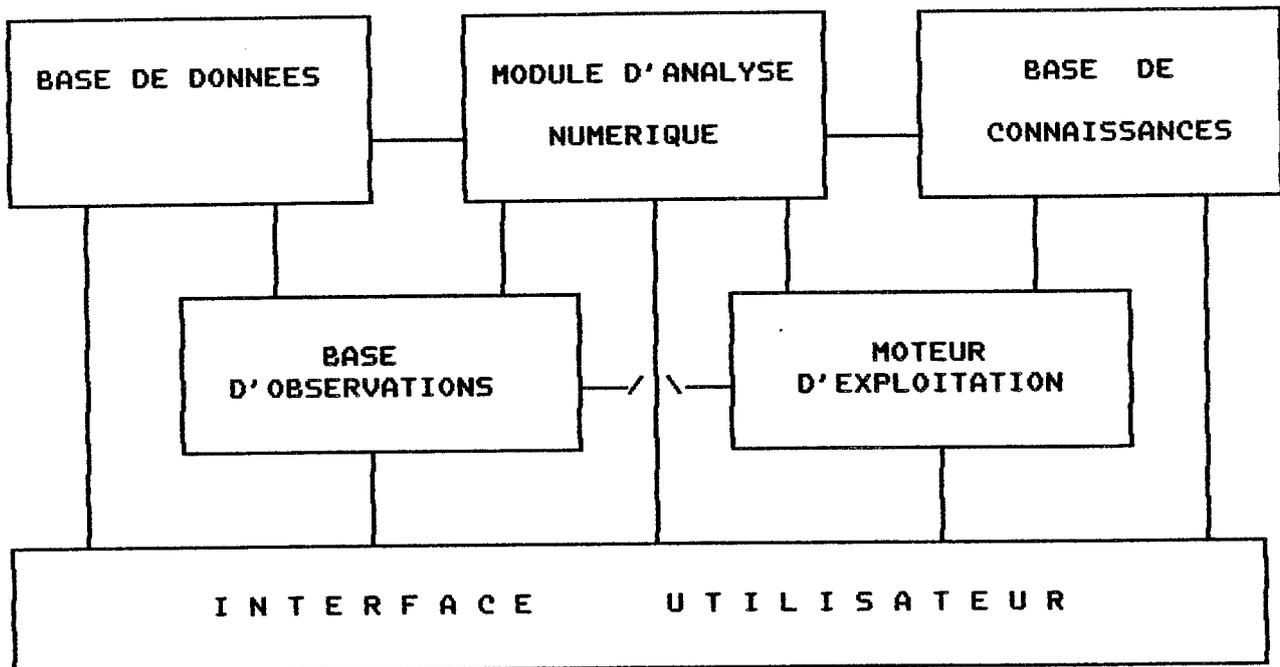


figure III.1

Le projet GENESYS a pour but la réalisation complète du système. Cette réalisation met en jeu aussi bien le versant médical que le versant informatique du système.

A / En ce qui concerne le versant informatique, l'avancement actuel des travaux concerne :

- 1°- La définition complète de la Base de Données et la réalisation d'un certain nombre de programmes utilitaires (§ III.2);
- 2°- La réalisation des principales procédures du Module d'Analyse Numérique (§ III.3).
- 3°-La réalisation d'une importante partie de la branche

procédurale du système hybride (schéma en II.7). Le programme correspondant, non terminé, n'a pu être porté en annexe en raison d'une part de la longueur de la liste des instructions, d'autre part de son état encore inachevé.

B / En ce qui concerne la partie médicale:

1°- Plus d'un millier de syndromes restent à ajouter à la Base de Données.

2°- Un travail d'affinement des scores (définition des entités symptomatiques composantes, recherches de seuils de discrimination) est en cours ou reste à faire.

### III.2 BASE DE DONNEES DYNAMIQUE ASSOCIEE AU SYSTEME

#### III.2.1 DEFINITION DU MODELE RELATIONNEL - FORMALISME Zo -

Nous proposons une définition théorique des relations liant les diverses entités de la base. La définition de ces relations utilise le Modèle Zo (ABRIAL 74) (ABRIAL 80) (DELOBEL 82). Ce modèle constitue un cadre souple et rigoureux de définition des entités (objets concrets ou abstraits) d'un domaine d'application particulier, ainsi que la définition des relations binaires existant entre ces objets.

Voici le tableau définissant les principaux objets et relations:

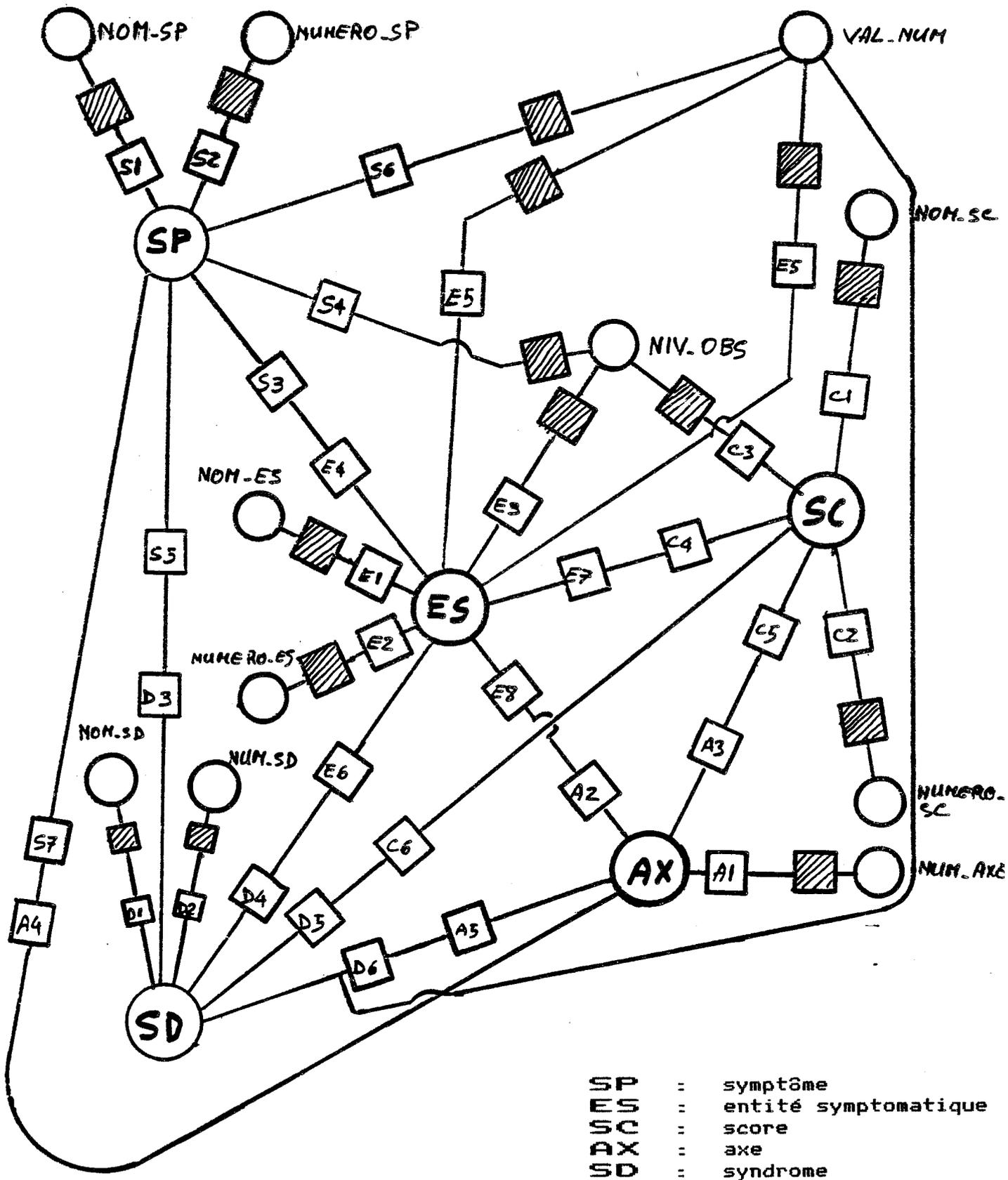
(la **cardinalité** donne le couple (minimum, maximum) d'éléments de la relation)

Chapitre III

OBJET	RELIE A	PAR LA RELATION	DONT LA DEFINITION EST:	DONT LA CARDINALITE
SYMPTOME	NOM - SP	S1	a pour nom	(1,1)
	NUMERO -SP	S2	a pour numéro	(1,1)
	ENTITE-SYMPT	S3	appartient à	(1,1)
	NIV - OBS	S4	a pour niveau d'observation	(1,1)
	SYNDROME	S5	apparaît dans	(0,n)
	VAL-NUM	S6	a pour valeur numérique	(1,1)
	AXE	S7	se trouve sur	(1,1)
ENTIT-SYMPT (entité symptomatique)	NOM-ES	E1	a pour nom	(1,1)
	NUMERO-ES	E2	a pour numéro	(1,1)
	NIV-OBS	E3	a pour niveau d'observation	(1,1)
	SYMPTOME	E4	a pour valeur naturelle	(1,n)
	VAL-NUM	E5	a pour ensemble de valeurs	(1,n)
	SYNDROME	E6	composante de	(1,n)
	SCORE	E7	composante de	(1,1)
	AXE	E8	représente l'axe vectoriel	(0,1)
SCORE	NOM-SC	C1	a pour nom	(1,1)
	NUMERO-SC	C2	a pour numéro	(1,1)
	NIV-OBS	C3	a pour niveau d'observation	(1,1)
	ENTIT-SYMPT	C4	a pour composante	(1,n)
	AXE	C5	représente l'axe vectoriel	(1,1)
	SYNDROME	C6	composante de	(1,n)
AXE	NUM-AXE	A1	a pour numéro	(1,1)
	ENTIT-SYMP	A2	représente	(0,1)
	SCORE	A3	représente	(0,1)
	SYMPTOME	A4	est le lieu vectoriel de	(1,n)
	SYNDROME	A5	est composante de	(1,n)
SYNDROME	NOM-SD	D1	a pour nom	(1,1)
	NUM-SD	D2	a pour numéro	(1,1)
	SYMPTOME	D3	définit sémiologiquement par	(1,n)
	ENTIT-SYMPT	D4	composé de	(1,n)
	SCORE	D5	composé de	(1,n)
	AXE	D6	a pour coordonnée	(n,n)

- voir le diagramme correspondant au modèle page suivante -

Le modèle partiel de base de donnée relationnelle qui vient d'être esquissé représente une forme idéale de référence pour cette base. La concrétisation de ce modèle dans notre application a dû se plier à un certain nombre de contraintes matérielles, logicielles et temporelles qui le défigure passablement, l'objectif de notre travail étant la réalisation d'un système utilisant la base de données et non pas la réalisation même de la base.



### III.2.2 MISE EN OEUVRE

La Base de Données est structurée de manière à répondre à deux objectifs différents :

1 - autoriser toutes les actions élémentaires que l'on est en droit d'attendre d'un Système de Gestion de Base de Données (S.G.B.D),

2 - être adaptée à la fonction de calcul numérique s'exerçant sur un nombre éventuellement grand de données réparties sur toute la base. Le point névralgique de cette fonction est naturellement le temps d'accès pour le rassemblement des données éparses et l'initialisation des traitements numériques.

En ce qui concerne le premier objectif, la solution qui nous est apparue la plus efficace a consisté à utiliser au mieux ce qui existait déjà, au moment de ce travail, en matière de S.G.B.D. sur un micro-ordinateur de type "compatible IBM-PC". Le choix s'est porté sur les systèmes DBASE II, puis DBASE III en raison tout d'abord de leur disponibilité sur le site, en raison également de la possibilité, que nous avons élaborée, d'accéder en lecture ou écriture à ces fichiers d'une manière totalement extérieure au S.G.B.D.

Le second objectif requiert en effet, comme nous allons le voir, un sous-système de gestion totalement spécifique et hors du cahier des charges de DBASE III.

La base de données doit couvrir la totalité des symptômes et des syndromes du domaine étudié. L'application a été faite (§ II.3) sur un sous-ensemble de la Génétique Médicale, à savoir :

l'ensemble des symptômes et des syndromes mis en jeu à partir d'une dysmorphie faciale caractéristique: la **fente labiale** ou **labio-palatine**; cette dysmorphie est caractéristique d'un ensemble de syndromes, au nombre de 102 au moment de l'expérimentation du système, le nombre de symptômes constatés dans les différents syndromes étant d'environ 675.

Il faut rappeler ici que ces chiffres n'ont rien de définitif puisque la Génétique Médicale évolue très rapidement, comme nous l'avons déjà annoncé. Nous pouvons seulement dire qu'ils sont relatifs à l'état d'avancement du domaine au moment de la réalisation du travail de N.GOMES et Ch.BAUDINO (BAUDINO 85).

Nous présenterons ensuite l'état actuel des travaux concernant le Module d'Analyse Numérique.

### III.2.3 STRUCTURATION DE LA BASE

#### III.2.3.1 - SCHEMA FONCTIONNEL - GENERATION EN CASCADE DES FICHIERS F1 à F8

Compte tenu des nécessités évoquées en introduction, nous avons structuré la base de données en 8 fichiers principaux:

a - Deux fichiers consacrés aux symptômes:

- . Un fichier de renseignements de longueur fixe: **SYMPTOME.DBF** (F1)
- . Un fichier de listes de syndromes attachés à chaque symptôme de F1: **FSYNDROM** (F2)

b - Trois fichiers consacrés aux syndromes:

- . Un fichier de renseignements de longueur fixe: **SYNDROME.DBF** (F3)
- . Un fichier de listes de symptômes attachés à chaque syndrome de F3: **FSYMPOM** (F4)

- Un fichier de listes de couples (entité symptomatique, valeur) attachés à chaque syndrome de F3: FES (F5).
- c - Un fichier d'informations concernant les entités symptomatiques: ES.DBF (F6)
- d - Un fichier d'informations sur les scores: LSTSCORE.DBF (F7)
- e - Un fichier de définition des axes pour les analyses multifactorielles: AXES.DBF (F8)

N.B.: le suffixe .DBF associé aux fichiers F1, F3, F7, F8 est caractéristique des fichiers gérés par le S.G.B.D. dbase III.

**BASE DE DONNEES**

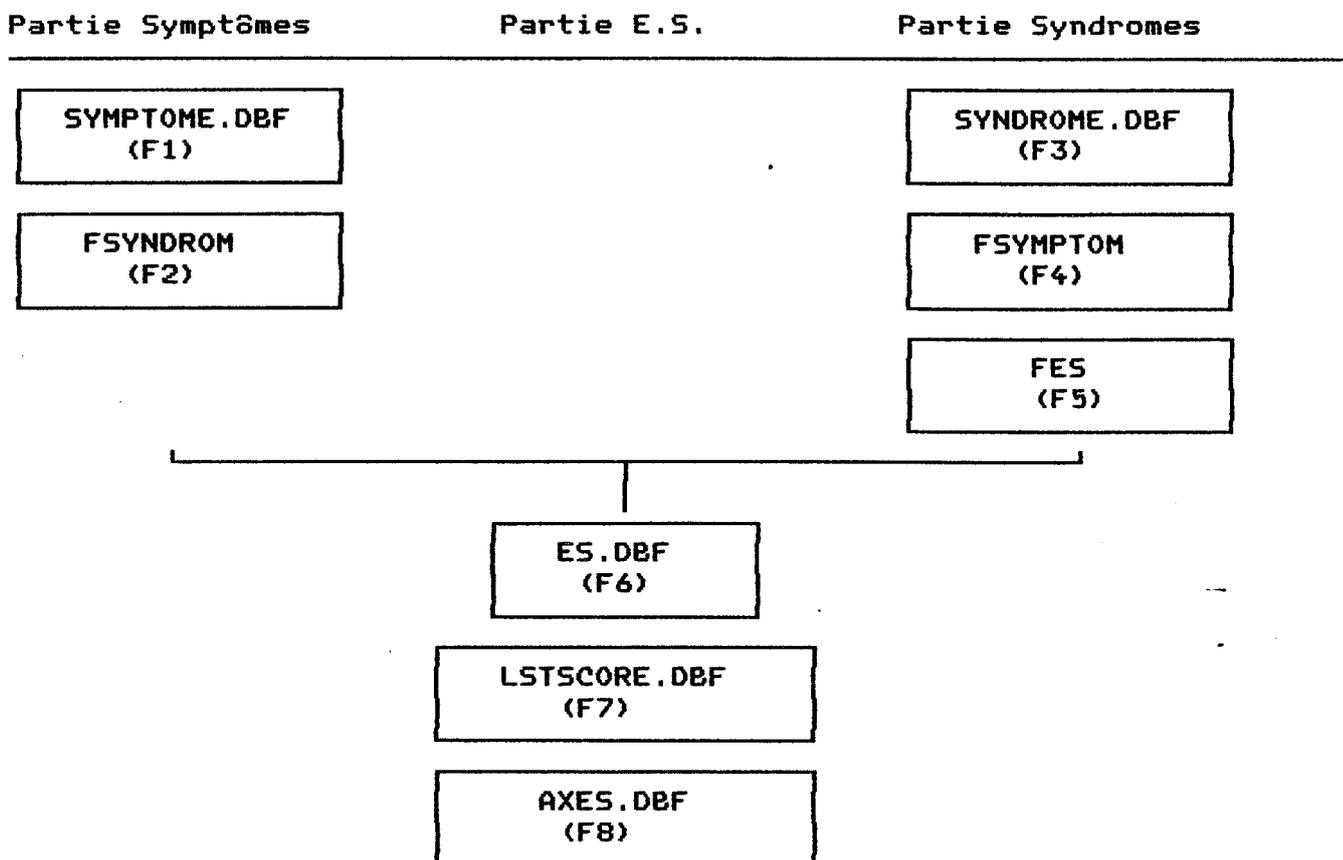


figure III.2

La base de données est générée **en cascade** à partir des fichiers

F1, F2 et F3 ou F4. Les deux fichiers F1 et F3 sont indispensables, F2 pouvant être remplacé par son dual F4 ; les autres fichiers sont créés à partir des 3 fichiers du noyau initial.

La chaîne de génération des fichiers est contenue dans un unique programme écrit en langage dbase III. C'est le programme CREAX.PRG (annexe 5) dont la finalité est de déduire les axes (F8) à partir des symptômes (F1), ainsi que les coordonnées (F5), sur ces axes, des syndromes.

La base de données est donc un ensemble de fichiers indépendants, à régénérer d'une manière simple (par un programme approprié), dans les cas d'ajout ou de modification d'un symptôme (F1), d'une liste de syndromes associée à un symptôme (F2) ou d'un syndrome (F3). Un certain nombre de champs de F1 et F2, destinés aux liens entre fichiers et aux numérotations d'axes ou d'entités symptomatiques, sont réservés au système.

La base de données possède un caractère éminemment mobile, adapté à un contexte de système expert et d'aide à l'apprentissage et à la production de connaissances. Elle s'intègre dans le modèle global de représentation des connaissances.

N.B.: le suffixe .DBF associé aux fichiers F1, F3, F7, F8 est caractéristique des fichiers gérés par le S.G.B.D.

### III.2.3.2 FICHIERS RELATIFS AUX SYMPTOMES

Les informations indispensables, associées à chaque symptôme se répartissent en deux catégories et, par conséquent, dans deux fichiers différents. Il en sera de même, nous le verrons plus loin,

pour les syndromes.

#### III.2.3.2.1 INFORMATIONS DE LONGUEUR FIXE (F1)

Elles sont regroupées dans le fichier SYMPTOME.DBF ou F1 géré en majeure partie par DBASE III.

Les divers champs de ce fichier sont constitués par :

- . l'expression en langage naturel du symptôme
- . le code numérique du symptôme
- . l'entité symptomatique (E.S)
- . le code numérique de l'entité symptomatique
- . la valeur naturelle de l'entité symptomatique
- . la valeur numérique de l'entité symptomatique
- . l'objet
- . l'attribut
- . les liens: code numérique de son ascendant, de son premier descendant, de son latéral droit
- . le niveau d'observation
- . le nombre de syndromes associés
- . le lien avec la liste des syndromes dans F2

On constatera, dans la structure type de l'enregistrement (donnée en annexe 3), que certains champs ont dû être ajoutés (L1, L2, ..., L18). Ces champs ont une valeur purement technique et sont destinés à rendre possible un accès au fichier par une procédure externe à DBASE III de lecture d'enregistrement et de modification de n'importe quel champ. Cette procédure, interne au système pascal, est utilisée par le programme de gestion des listes de syndromes ainsi que par le module de production de connaissances (§ III.1 et III.2.3.1).

#### III.2.3.2.2 INFORMATIONS DE LONGUEUR VARIABLE (F2)

Ces informations, gérées par un programme spécifique, GFSYND (annexe 6), indépendant de DBASE III, sont constituées par les listes de syndromes dans lesquels apparaissent les symptômes.

Ces listes constituent le fichier FSYNDROM ou F2.



fichier FSYNDROM.

Cette liste est constituée de 4 types d'éléments:

- a - le numéro du symptôme (lien avec F1)
- b - le nombre de syndromes associés (existe également dans F1 ; cette redondance est justifiée par la rapidité d'accès).
- c - les chaînages vers les symptômes PERE, FILS, FRERE (arborescence des symptômes, § II.4.1.4)
- d - la liste proprement dite de numéros de syndromes (lien avec F3 et F4), avec pour chacun d'eux, le coefficient fréquentiel d'apparition du symptôme (Q,S,T,I) (§ II.5.1.3).

Chacun des éléments (a,b,c,d) occupe un nombre variable de **cellules de base**. La cellule de base de la liste a été étudiée pour occuper le plus petit espace possible de mémoire, en l'occurrence 16 bits ( un objet de type "integer" en pascal sur micro-ordinateur).

Chaque cellule possède une adresse. L'adresse de la première cellule, correspondant à l'adresse de la liste dans le fichier, figure dans F1.

Il peut paraître surprenant de voir :

- 1°) se répéter le numéro du symptôme concerné ( 1ère cellule de la liste) : il est déjà dans le fichier F1,
- 2°) apparaître des champs qui auraient tout aussi bien pu se trouver au niveau du fichier F1 (liens avec les symptômes père, fils, frère)

La réponse à ces deux remarques tient au fait que l'un des critères primordiaux est celui de la rapidité d'exécution et donc

## Chapitre III

de l'optimisation du nombre d'accès aux fichiers.

Dans la mesure où F2 est lu une seule fois et d'un seul bloc (§ III.2.3.2), il devenait intéressant d'y mettre des informations non essentielles mais dont la connaissance aurait nécessité des accès à F1.

Par exemple, lors de la mise à jour de la base de faits (observation d'un malade), il est nécessaire de rechercher les "symptômes pères" de chaque symptôme observé, afin de ne pas "oublier" de syndromes candidats. Cette recherche est plus rapide au niveau de F2 que de F1.

Voici donc le schéma d'une structure de liste de syndromes:

(QSTI étant le nom du coefficient fréquentiel d'apparition du symptôme dans le syndrome)

### Chapitre III

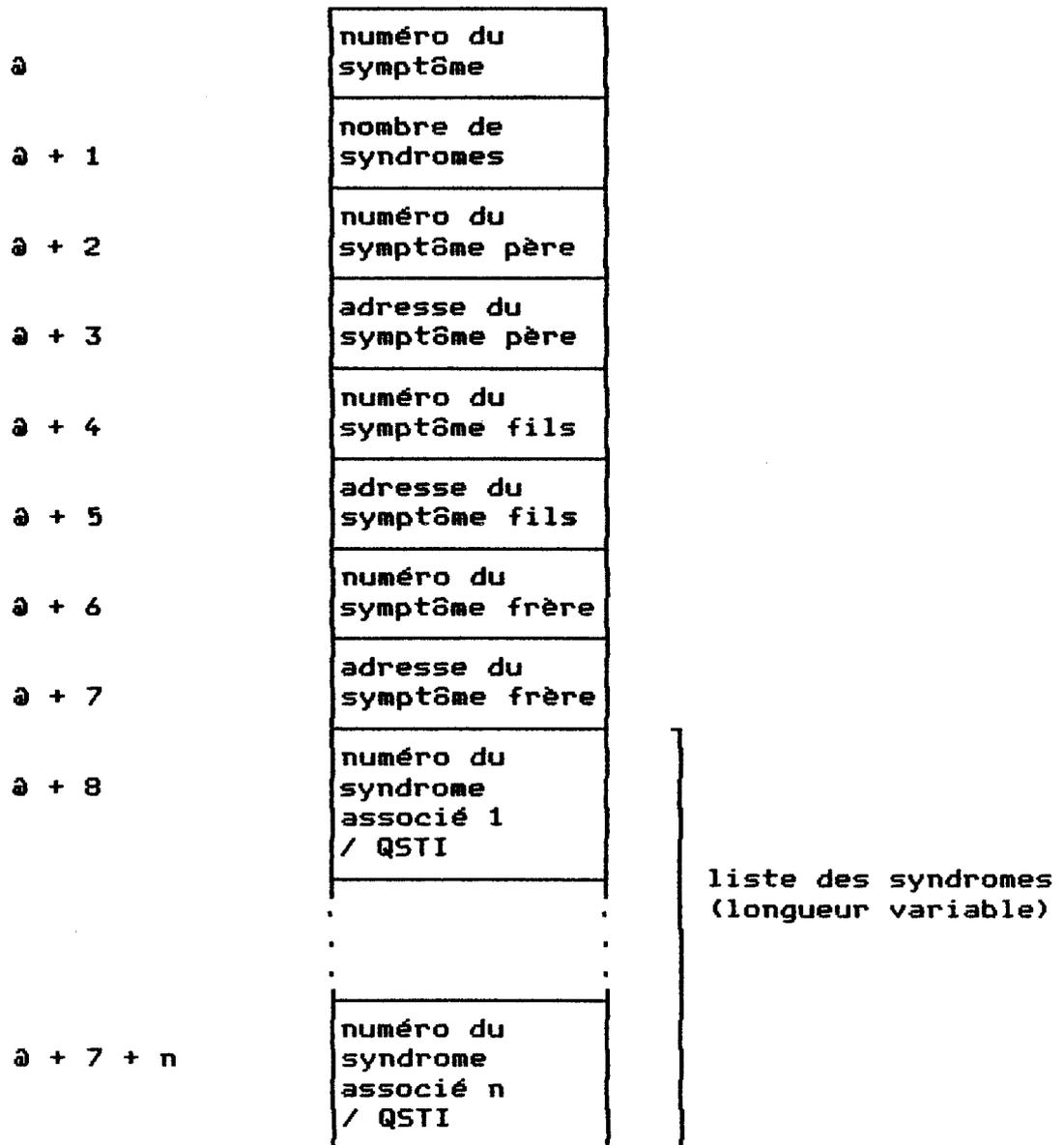


figure III.4

Le calcul de l'adresse de la liste de syndromes associés au symptôme suivant du fichier F1 est effectué simplement par calcul de la longueur:  $a + 8 + \text{contenu de } (a + 1)$ .

#### III.2.3.2.4 CODAGE DES CELLULES DE SYNDROMES

Le codage des informations a été établi de manière à offrir une

certaine possibilité de reconnaissance du type du contenu d'une cellule:

C'est ainsi que les codes des numéros:

- de symptômes de référence,
- de symptômes père, fils, frère,

sont des nombres négatifs compris entre -257 et -32768 obtenus par la transformation affine  $f(x) = -(x+256)$  du numéro  $x$  du symptôme.

Le code du nombre de syndromes est un chiffre négatif compris entre -1 et -255, obtenu par multiplication par -1 du nombre réel de syndromes.

Pour chaque symptôme, un seul entier positif compris entre 0 et 32767 représente à la fois le numéro d'un syndrome le possédant et le coefficient fréquentiel du symptôme dans le syndrome. Soit  $n$  le numéro d'un syndrome, et  $q$  le nombre représentant l'indice fréquentiel (QSTI) du symptôme dans le syndrome  $n$ ;  $q$  est pris dans l'ensemble:  $\{ 0,1,2,3 \}$  représentant respectivement les symboles Q,S,T,I. Ces chiffres représentent de simples codages. Ils n'ont rien à voir avec les évaluations des indices fréquentsiels (§ II.5.1.3).

Le codage est donné par la formule:  $c = n * 4 + q$

Le décodage s'obtient par l'opération inverse:

$$\begin{aligned} n &= c \text{ DIV } 4 && (\text{DIV: division entière}) \\ q &= c \text{ MOD } 4 && (\text{MOD: modulo}) \end{aligned}$$

Ainsi codées, les informations du fichier FSYMPTOM occupent une place minimum. Ceci permet une lecture en bloc de ce fichier en mémoire centrale. Cette lecture globale et unique du fichier permet de diminuer considérablement les temps de recherches des syndromes

candidats, recherches se faisant par intersection des listes de syndromes associés aux symptômes observés.

D'autre part, également pour une question de gain de rapidité, le coefficient fréquentiel QSTI apparaît de manière redondante dans les deux fichiers duaux FSYNDROM et FSYMPTOM.

### III.2.3.3 FICHIERS RELATIFS AUX SYNDROMES

A l'instar de celles concernant les symptômes, les informations associées à chaque syndrome se répartissent en deux catégories: d'une part les renseignements de longueur fixe (fichier SYNDROME, DBF ou F3); d'autre part, les renseignements de longueur variable (fichier FSYNDROM ou F4).

#### III.2.3.3.1 INFORMATIONS DE LONGUEUR FIXE (F3)

Ce sont les informations gérées par le S.G.B.D. DBASE III. Un enregistrement du fichier est composé des champs suivants:

- . le nom en clair du syndrome
- . le code numérique du syndrome
- . le nombre de symptômes composants
- . des renseignements divers
- . le coefficient de prévalence
- . le lien avec la liste des symptômes dans F4
- . le lien avec la liste des couples (es, v) dans F5

De même que pour les symptômes, on constatera, dans la structure complète donnée en annexe 3, que différents champs (L9, ..., L8), d'un octet chacun, ont été rajoutés afin de permettre une compatibilité avec le traitement en pascal des chaînes de caractères.

**III.2.3.3.2. INFORMATIONS DE LONGUEUR VARIABLE - LISTE DE SYMPTOMES (F4) ET DE COUPLES (es, v) (F5)**

Le fichier F4 ou FSYMPTOM est analogue au fichier F2 en ce qu'il fournit une liste chaînée de listes de symptômes. Bien qu'il puisse être géré individuellement (ajouts, suppressions, etc.), une telle gestion est déconseillée. En effet, ce fichier doit rester constamment le "dual" du fichier FSYNDROM: à partir de l'ensemble des listes de syndromes associés à chaque symptôme, on doit pouvoir reconstituer intégralement la liste des symptômes de chaque syndrome.

C'est la solution qui a été adoptée: un programme spécial, écrit en pascal et nommé GENFSYMP (annexe 7) est chargé de créer entièrement le fichier F4 et de mettre à jour les champs: "nombre de symptômes" et "lien avec la liste des symptômes" dans le fichier F1.

Le fichier ainsi créé a donc la structure indiquée par la figure suivante:

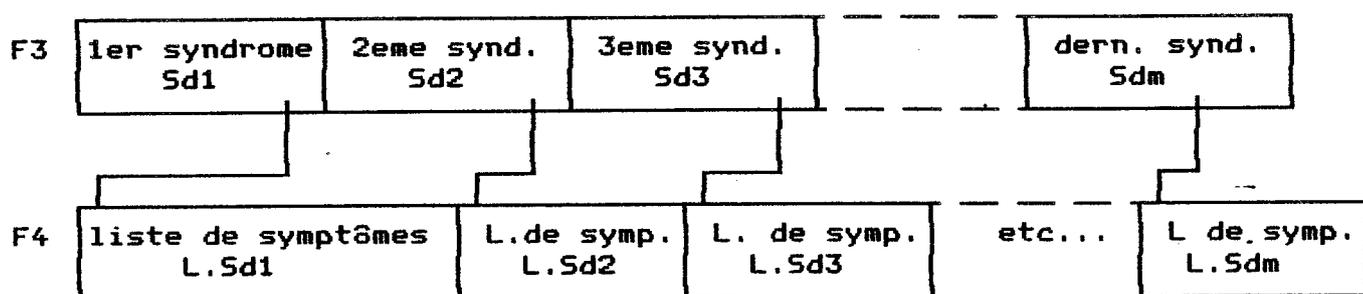


figure III.5

D'un point de vue "base de données", l'association des deux fichiers F3 et F4 réalise (entre autre) la relation de possession d'un syndrome vis-à-vis de son ensemble de symptômes.

La relation POSSEDE (SD,SP) se résout de la façon suivante:

Le syndrome SD est recherché dans le fichier F3. S'il existe dans ce fichier, alors, dans l'enregistrement correspondant, le champ "lien avec la liste des symptômes" fournit l'adresse, dans le fichier F4, de la liste dans laquelle il conviendra de rechercher SP.

### III.2.3.3.3 STRUCTURE DES LISTES DE SYMPTÔMES

Comme pour les listes de syndromes de F2, les listes de symptômes de F4 sont composées de cellules. Nous n'aurons ici que trois types d'éléments composant la liste:

- a - le numéro du syndrome
- b - le nombre de symptômes
- c - la liste proprement dite des symptômes, chacun affecté du coefficient fréquentiel d'apparition dans le syndrome concerné. Cette liste est de longueur variable.

Voici le schéma d'une structure de liste de symptômes:

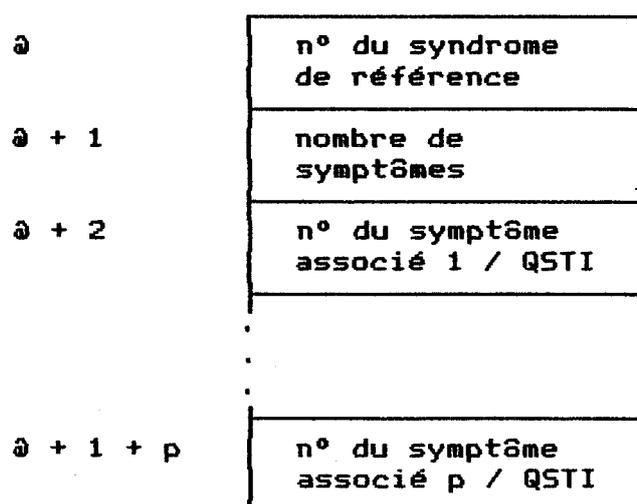


figure III.6

Chaque cellule de la liste occupe 2 octets représentant un objet de type entier.

Le calcul de l'adresse de la liste suivante, dans F3, de symptômes définissant un syndrome s'effectue simplement par calcul de la longueur de la liste:  $\emptyset + 2 + \text{contenu de } (\emptyset + 1)$ .

#### III.2.3.3.4 CODAGE DES CELLULES DE SYMPTOMES

Le numéro de syndrome est un nombre négatif compris entre -257 et -32768.

Il est obtenu par la transformation affine suivante:

$$f(x) = x + 256; \quad x \text{ étant le numéro de code du syndrome.}$$

Le code du nombre de symptômes est un entier négatif obtenu par multiplication par -1 du nombre réel de symptômes.

Les numéros de symptômes ainsi que leur coefficient fréquentiel sont représentés par un nombre positif compris entre 0 et 32767.

Ce nombre est obtenu de la même manière que pour le fichier F2.

#### III.2.3.3.5 STRUCTURE DES LISTES DE COUPLES (ENTITE SYMPTOMATIQUE, VALEUR)

Le fichier FES (F5) est similaire au fichier F4 des listes de symptômes. Il fournit pour chaque syndrome du fichier F3 sa liste de couple (es,v). Etant donné qu'un tel couple représente un symptôme, il existe pour un syndrome donné autant de couples (es,v) que de symptômes.

A l'instar des listes des syndromes (F2) et de symptômes (F4), les listes de couples (es,v) sont composées de 3 types d'éléments:

a - le numéro du syndrome

b - le nombre de couples de la liste

c - la liste des couples, avec:

1ère valeur: le numéro de l'es.

2ème valeur: la valeur numérique du symptôme.

Voici le schéma d'une structure de liste de couples (es,v):

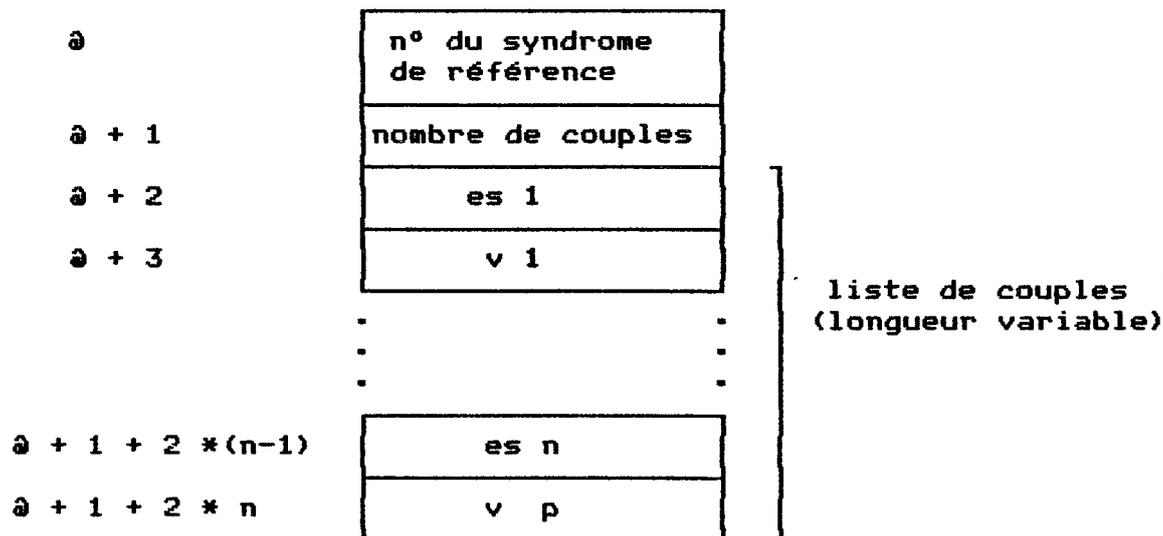


figure III.7

Le fichier FES est créé par le programme GENFES (annexe 8) écrit en pascal. Ce programme réalise également les chaînages avec le fichier des syndromes (F3).

#### III.2.3.4 FICHIERS RELATIFS AUX ENTITES SYMPTOMATIQUES (F6)

Il s'est avéré nécessaire, au cours de la phase d'élaboration du fichier F1, de construire un fichier spécifique des entités symptomatiques. En effet, le couple (entité symptomatique, valeur) doit recevoir un codage pour être utilisable par le module de calcul.

Or, il n'existait a priori aucune liste d'e.s. à partir de laquelle établir un numérotage: les e.s. ainsi que leurs valeurs

possibles sont distribuées parmi les champs "entité symptomatique", "valeur naturelle" et "valeur numérique" des enregistrements de symptômes (fichier F1).

Nous avons donc:

1°) défini un fichier d'entités symptomatiques, nommé ES.DBF, structuré de la manière suivante:

- . le nom de l'entité symptomatique
- . le numéro de code
- . le numéro de chapitre
- . le nombre de valeurs
- . 25 champs de 4 octets chacun, nommés S1, S2, ..., S25, dans lesquels figurent les numéros de symptômes.

Les différentes valeurs naturelles d'une entité symptomatique doivent donc être lues dans les enregistrements du fichier SYNDROME.DBF.

2°) rempli automatiquement les enregistrements de ce fichier par analyse du fichier F1. Cette tâche a été réalisée par un programme spécial, écrit en langage de commande Dbase III, nommé CRES.PRG (§ annexe 4) (le suffixe .PRG est spécifique des programmes gérés par le S.G.B.D.).

3°) affecté automatiquement un numéro de code à chacune des e.s., et inscrit ce numéro dans le champ correspondant des enregistrements de F1. Cette tâche a été également réalisée par le programme CREAX.PRG (§ III.2.3.1).

Le nombre actuel d'entités symptomatiques est égal à 469.

#### III.2.3.5 FICHIER RELATIF AUX SCORES (F7)

Le fichier des scores, nommé LSTSCORE.DBF, est un fichier de

structure dbase III, entièrement créé à partir du fichier F6 des entités symptomatiques. Il fournit une liste détaillée des scores de la base de données. Ainsi, chaque score est décrit par les champs suivants (structure de l'enregistrement):

- numéro de l'axe correspondant
- niveau d'observation du score
- numéro du chapitre d'appartenance
- nom du score
- nombre d'entités symptomatiques composantes
- liste des entités composant le score.

Ce fichier est généré en même temps que le fichier suivant F8.

Le nombre actuel de scores est égal à 116. La liste des scores peut être consultée en annexe 1 (axes de type S). Des exemples de scores sont donnés au paragraphe II.5.3.3.

La structure d'un enregistrement de score figure en annexe 2.

#### III.2.3.6 FICHER RELATIF AUX AXES DE L'ANALYSE DES DONNEES (F8)

Le fichier des axes de l'analyse des données constitue, à la suite du fichier des scores, le dernier fichier de la chaîne de génération de la base de données. Il fournit une liste détaillée des axes. C'est au cours de sa création que sont attribués les numéros des axes auxquels appartiennent les symptômes. Par conséquent, à la suite de la création de F8, un retour à F1 est effectué afin de mettre à jour le champ "numéro de l'axe" de chaque symptôme. Chaque axe de l'analyse des données est décrit par les champs suivants :

- numéro de l'axe
- niveau d'observation
- chapitre
- type d'axe: E= élémentaire, S= score
- numéro de l'entité symptomatique (si type = E)
- nom du score (si type = S)
- nombre de symptômes disposés sur cet axe

## Chapitre III

- liste des numéros des sympt8mes.

Les axes sont actuellement au nombre de 174. Ils sont distribués de la façon suivante :

niv.ob	I	II	III	IV
axes E	31	6	14	14
scores	36	16	18	39
TOTAL	67	22	32	53

**NOMBRE TOTAL D'AXES: 174**

### III.2.4 GESTION DE LA BASE - METHODES D'ACCES AUX DIFFERENTS FICHIERS

Nous présentons ici, d'une manière succincte, les différentes actions pratiquées sur les données de la base ainsi que les méthodes employées.

#### III.2.4.1 FICHIERS DBASE III

En ce qui concerne les fichiers SYMPTOME.DBF et SYNDROME.DBF, nous avons à notre disposition toutes les fonctions du S.G.B.D.: AJOUT, MODIFICATION, VISUALISATION, RECHERCHE MULTICRITERE, IMPRESSION, etc.

Il faut simplement retenir qu'il existe, dans chaque fichier, des champs gérés "de l'extérieur", et qui, par conséquent, ne doivent pas être modifiés à partir de DBASE III.

Ces champs, nous le rappelons, sont :

- pour SYMPTOME.DBF:
  - , le nombre de syndromes associés

## Chapitre III

- . le lien avec la liste des syndromes

- pour SYNDROME. DBF:

- . le nombre de symptômes définissant le syndrome
- . le lien avec la liste des symptômes
- . le lien avec la liste des couples (es, v).

### III.2.4.2 FICHIERS PASCAL

Comme nous l'avons indiqué brièvement, ces trois fichiers, à cause de leur structure de listes de longueurs variables, ne peuvent être gérés par le S.G.B.D. retenu. Nous avons donc d'abord créé un programme particulier de gestion du fichier FSYNDROM (F3), puis un programme de génération à partir de FSYNDROM du fichier "dual" FSYMPTOM (F4, § III.2.3.3.2). Enfin, nous avons réalisé un programme de création du fichier FES (F5) des couples, (entité, valeur) définissant les syndromes.

Le programme de gestion de FSYNDROM permet à l'utilisateur (le médecin ou le gestionnaire de la base), de garder la maîtrise des données de longueur variable: ajout, modification, suppression de symptômes dans des syndromes.

### III.3 MODULE D'ANALYSE NUMERIQUE

Le module de calcul regroupe un ensemble de programmes centrés sur l'Analyse des Données de la base. Les parties principales de ce module sont constituées par :

- le programme de création de la matrice (syndrome, axes), point de départ de l'analyse numérique. Ce programme (annexe 13) utilise le fichier FES (F5) des listes de couples (es,v) pour calculer les coordonnées des syndromes impliqués dans une recherche de diagnostic, pour un niveau d'observation donné.

**N.B.:** toutes les matrices créées ou utilisées par ce module sont contenues dans des fichiers.

- Le programme de calcul de la matrice de variances-covariances (matrice symétrique, définie positive) utile pour l'Analyse en Composantes Principales et la recherche des axes pertinents (§ II.6 et II.7) aux divers niveaux d'observation. Ce programme a comme principal paramètre d'entrée une matrice (syndromes x axes) donnant les coordonnées des syndromes (lignes) sur les axes (colonnes). Syndromes et axes sont déterminés par l'espace de travail (espace complet ou espace conjectural, § II.5.5).

- La procédure de recherche des valeurs et des vecteurs propres à partir de la matrice de variances-covariances. Nous avons utilisé l'algorithme de Jacobi (diagonalisation d'une matrice symétrique, définie-positive). Cette procédure de diagonalisation a été empruntée à la Bibliothèque d'Analyse Numérique de Grenoble (B.A.N.G.- Institut de Mathématiques Appliquées de Grenoble, Université de Grenoble I). Elle donne pour résultat la liste des valeurs propres ordonnées par valeurs décroissantes, ainsi que la

### Chapitre III

liste des vecteurs propres correspondants (pour ces deux dernières procédures, voir annexe 14).

- Le programme graphique d'Analyse en Composantes Principales. Ce programme trace sur l'écran graphique (640 x 400 points) les représentations cartésiennes des syndromes dans les axes de coordonnées souhaités (§ annexe 15)

- Le programme de recherche des axes pertinents à partir des vecteurs propres, dont le principe algorithmique est exposé en II.6 (annexe 16).

Ces programmes sont encore à un stade de développement et seront, à terme, intégrés dans le logiciel complet.



## CHAPITRE IV

### CONCLUSION

Le travail que nous avons présenté avait, comme il a été annoncé, le double objectif d'étudier d'une manière précise et critique les diverses méthodes employées jusqu'à maintenant en matière d'aide au diagnostic médical, et de proposer, pour le domaine très particulier de la génétique médicale, une méthodologie nouvelle et adaptée aux caractéristiques particulières de ce domaine.

En ce qui concerne le premier objectif, un tour d'horizon des théories et des techniques existantes a été réalisé, avec, pour corollaire, l'analyse plus ou moins approfondie d'un certain nombre de systèmes, experts ou non, en relation avec les méthodes exposées. Les objectifs et les limites de ces méthodes ont également été analysés.

En épilogue à cette analyse, une place assez importante a été accordée aux directions de recherches d'avant-garde telles que l'association Intelligence Artificielle - Bases de Données ou les Systèmes Experts hybrides (procéduraux-déclaratifs). Cette dernière phase de l'étude a tenté d'apporter quelques éléments de prospective scientifique en matière d'aide au diagnostic médical.

La transition avec le second objectif de la thèse s'est faite au moyen de cette prospective. Nous avons en effet proposé ensuite un important travail méthodologique d'analyse et de conception d'un système d'aide au diagnostic dans un domaine vaste, incertain, incomplet et éminemment évolutif.

## Chapitre IV

Nous avons, en préliminaire, cité et brièvement exposé quelques systèmes qui intègrent des outils d'aide à l'expression et à l'affinement de connaissances expertes.

Compte tenu des caractéristiques particulières du domaine considéré, nous avons proposé une méthode mixte d'aide au diagnostic et de production de connaissances nouvelles apparentée à un processus d'apprentissage.

Cette méthode se distingue des méthodes existantes par deux aspects principaux:

- elle s'en distingue d'une part par le modèle de représentation des données: les diagnostics (ou buts) du système sont des points répartis dans un espace multidimensionnel d'entités symptomatiques. Ces entités symptomatiques représentent les lieux (anatomiques, fonctionnels, temporels, etc.) dans lesquels les syndromes ont une incidence et au travers desquels ils peuvent être décelés. Ces diverses entités constituent le support (les axes) d'une représentation spatiale des syndromes. Une telle représentation, située dans un espace à plusieurs dizaines de dimensions, est "rendue visible" sur un plan au moyen de l'Analyse en Composantes Principales. Une telle analyse produit un certain nombre de connaissances sur le domaine: elle permet de mettre en évidence les axes qui ont le meilleur pouvoir décisionnel, c'est-à-dire ceux sur lesquels les projections des syndromes sont les plus espacées.

- elle s'en distingue d'autre part par l'approche hybride - à la fois déclarative et procédurale - de l'aide au diagnostic. Cette approche a l'ambition d'intégrer au sein d'un même système, des formalismes aussi différents que ceux de l'Intelligence Artificiel-

le d'un côté, de l'Analyse des Données d'un autre côté. Nous avons, à ce propos, tenté de démontrer qu'une solution concrète et opérationnelle au problème précis d'aide au diagnostic considéré, ne pouvait être entièrement assumé par un formalisme unidimensionnel. La complexité - au moins structurelle - de la solution répond à la complexité du problème qu'il nous incombait de résoudre.

La dimension d'aide à l'apprentissage fait partie intégrante de cette structure complexe. Nous nous sommes en effet très vite aperçus que le processus d'aide au diagnostic élaboré n'était autre qu'un processus d'investigation débouchant sur des connaissances nouvelles et réutilisables ultérieurement. Dès lors, les connaissances produites par l'analyse des données de la totalité de la base, puis celles qui sont accumulées au cours des sessions de traitement individuel de malades, peuvent être mémorisées et enrichir ainsi une base de connaissances utilisables par un moteur d'exploitation.

La méthode a été illustrée par un exemple partiel de traitement sur une Base de Données restreinte à environ 675 symptômes et 102 syndromes. La valeur seulement exemplaire de ce traitement est due au fait que la méthodologie exposée s'accorde et s'applique aux très grands nombres de données, ce qui est le cas pour la Génétique Médicale, avec ses plusieurs milliers de symptômes et de syndromes. Un certain nombre de problèmes ont été volontairement laissés dans l'ombre :

1°) Le problème des disjonctions exclusives de symptômes dans la définition d'un syndrome.

Exemple : le syndrome X possède les symptômes A ou B exclusifs, par exemple : A = hypertélorisme, B = hypotélorisme.

Ce problème peut être résolu si l'on accepte que le syndrome X possède plusieurs représentations spatiales, ce qui vraisemblablement ne devrait pas poser de problèmes sur le plan théorique.

2°) Les scores additifs sur les valeurs absolues séparent mal les syndromes : pour un même score, certains syndromes ayant des projections différentes dans l'espace des axes du score peuvent avoir des valeurs de score identiques.

Exemple sur un score à deux axes :

X ( 1, -1)      score (X) = 2

Y (-1, 1)      score (Y) = 2

Cet inconvénient explique le chargement de certains points dans les graphes de projection (annexe 10).

Une solution consiste par exemple à trouver un mécanisme de calcul de score qui sépare totalement les syndromes (comme l'addition de puissances de deux).

3°) La validation et l'affinement des scores passent par des études épidémiologiques approfondies qui dépassent probablement le cadre d'un seul C.H.U. Compte tenu des caractéristiques de la Génétique Médicale (§ II.3), il n'est pas exagéré de dire que ces enquêtes devraient être effectuées à l'échelle mondiale.

Nous n'avons pu dépasser l'échelle d'une Base de Données réduite en raison du délai très important, sans rapport avec celui d'une seule thèse, que nécessitera la réalisation complète de la base. Une telle réalisation s'inscrit dans un projet à plus long terme de réalisation du système complet d'aide au diagnostic et à la production de nouvelles connaissances, nommé GENESYS, dont nous avons décrit la démarche globale.

## Chapitre IV

En ce qui concerne le versant informatique du projet, l'avancement actuel des travaux concerne :

1°- La définition complète de la Base de Données et la réalisation d'un certain nombre de programmes utilitaires.

2°- La réalisation des principales procédures du Module d'Analyse Numérique.

3°-La réalisation d'une importante partie de la branche procédurale du système hybride.

Le calendrier des prochains travaux comporte:

- l'étude de faisabilité de la partie déclarative et de son interfaçage avec la partie procédurale, au moyen d'outils avancés de production de systèmes experts (environnement et langage orienté objet).

- la réalisation de cette partie.

B / En ce qui concerne la partie médicale:

1°- Plus d'un millier de syndromes restent à ajouter à la Base de Données.

2°- Un travail d'affinement des scores (définition des entités symptomatiques composantes, recherches de seuils de discrimination) est en cours ou reste à faire.

Si notre base peut paraître restreinte par rapport à l'ampleur de la Génétique Médicale, sa réalisation, au cours de trois années de travail avec le Docteur A. JOANNARD - Clinique Médicale Infantile, CHU de Grenoble - a connu, en ce qui concerne la connaissance médicale, un certain nombre d'épiphénomènes: questionnement en profondeur sur la notion d'entité symptomatique, mise à jour de relations particulières ou de liaisons entre ces entités,

## Chapitre IV

étude de scores, etc. Il faut noter, au travers et au-delà de ces épiphénomènes, la mise en place des bases d'une recherche à long terme.

Nous avons insisté tout au long de ce travail sur le nombre important de diagnostics et de symptômes considérés. Un tel système de production de connaissances ne peut donc conduire qu'à un nombre également très élevé de règles de production .

Le système sera alors confronté à la problématique des grandes bases de connaissances. Il faut noter ici que la question des bases de connaissances très chargées est encore à l'heure actuelle peu abordée en Intelligence Artificielle : la spécialisation des bases de connaissances sur des domaines restreints est une pratique quasi générale. Néanmoins, ce caractère restreint des domaines d'application étant plus souvent lié à des problèmes pratiques qu'à des questions théoriques, il est vraisemblable que nous assisterons dans un proche avenir au développement des grandes bases de connaissances.

Dans de tels systèmes, et dans notre application en particulier, une avancée théorique et pratique sera indispensable pour rendre la base de connaissances effectivement opérationnelle. Il est vraisemblable que cette avancée se fera dans le domaine de la métaconnaissance et de la construction de règles de connaissances heuristiques.

Cette métaconnaissance proviendra à la fois des experts eux-mêmes, dans la mesure où la méthodologie et la pratique de la construction des bases seront mieux maîtrisées ; elle proviendra également vraisemblablement de méthodes **d'acquisition automatique de métaconnaissances.**

B I B L I O G R A P H I E

- 1 ( ABRIAL 74 )  
ABRIAL J.R.  
"Data semantics"  
in Data Based Management, Klimbie and Koffeman Eds., North-Holland,  
pp. 1-59.
- 2 ( ABRIAL 81 )  
ABRIAL J.R.  
"Méthodes de spécification et construction"  
Proc. 3ème Jour. Fran. Inform. Genève, Jan. 81
- 3 ( ADAM 76 )  
ADAM J.B.  
"A probability model of medical reasoning and the MYCIN model"  
Math. Biosciences 32, 1976, pp. 177-186.
- 4 ( ADLASSNIG 80 )  
ADLASSNIG K.P.  
"A fuzzy logical model of computer-assisted medical diagnostic"  
Methods of Information in Medicine, 18, 3 (1980), pp. 141-148.
- 5 ( AYME 81 )  
AYME S., GOUVERNET J.  
"Informatique et Génétique Médicale : de la gestion d'un centre à  
l'aide au diagnostic..."  
in "Génétique Médicale, acquisitions et perspectives", ouvrage coll.  
ed. FEINGOLD J., INSERM/FLAMARION, 1981, pp. 283-311.
- 6 ( BAUDINO 85 )  
BAUDINO Ch., GOMES N.  
"Contribution à l'élaboration d'un système informatique d'Aide au  
Diagnostic des Syndromes Malformatifs"  
Thèse de doctorat de Médecine, Faculté de Médecine de Grenoble, 30  
mai 1985.
- 7 ( BAYES 1763 )  
BAYES Th.  
"An essay towards solving a problem in the doctrine of chances"  
Phil. Trans. Roy. Soc., 1763, pp. 53-370, reproduit dans Biométrie,  
1958, pp. 45-296, cité par ( GREMY 69 ).
- 8 ( BENSÂID 86 )  
BENSÂID A., RECHENMANN F., SIMONET A., VIGNARD P.  
"Mécanisme d'Inférence et d'Exploitation dans des Bases de Connais-  
sances Centrées - Objets"  
Proc. 8e Jour. Francophones sur l'Inf.: "Bases de Données et Bases  
de Connaissances", Grenoble, 14 & 15 Janv. 86, pp. 15-29.

## Bibliographie

- 9 ( BENZECRI 81 )  
BENZECRI J.P. & Coll.  
"Pratique de l'Analyse des Données"  
Ed. DUNOD, tomes 1,2,3, Paris, 1981.
- 10 ( BERGMAN 84 )  
BERGMAN D.A., PANTELL R.M.  
"The art and science of medical decision making"  
J. of Pediatrics, vol 104, n°5, pp. 649-656.
- 11 ( BERGSMA 79 )  
BERGSMA D.  
"Birth defects comendium"  
Alan R. LISS- 2nd ed., New-York, 1979.
- 12 ( BERNARD 77 )  
BERNARD J.  
"La Décision médicale"  
Cahiers Médicaux, 1977, 3 (10), pp. 529-532.
- 13 ( BERNARD 85 )  
BERNARD P., FABRIES J.P., DEMONGEOT J.D.  
"Le score pronostique de la chirurgie réparatrice tubaire distale"  
GYNECOLOGIE, vol. 36, n° 1, 1985, pp. 35-41.
- 14 ( BLUM 82 )  
BLUM R.L.  
"Discovery, Confirmation and Incorporation of Causal Relationships  
from a large time-oriented Clinical Data Base : the RX Project"  
Comp. Biom. Res., 15, 1982, pp. 164-187.
- 15 ( BONAITI 82 )  
BONAITI C., BRIARD M.L., FEINSGOLD J., PAVY B., PSAUME J.,  
MINGUE-TRUFFERAUD G., KAPLAN J.  
"An epidemiological and genetic study of facial clefting in France -  
I Epidemiology and frequency in relatives"  
J. Med. Gen., vol 19, n°1, 1982, pp. 8-15.
- 16 ( BONNET 82 )  
BONNET A., HARRY J., GANASCIA J.G.  
"LITHO : un système expert inférant la géologie du sous-sol"  
TSI, 5(1), 1982, pp. 393-402.
- 17 ( BRIARD 83 )  
BRIARD M.L., BONAITI C., FREZAL J.  
"Facteurs épidémiologiques et génétiques des fentes labiales et pa-  
latines"  
in Proc. 40e Cong. Soc. Fr. Chir. Ped., 1983 (24), pp. 228-230.
- 18 ( CARD 73 )  
CARD W.I.  
"Computer-assisted diagnosis and pattern recognition : the computing  
approach to clinical diagnosis"  
Proceedings of the Royal Society, London, B, 184, p. 421.

## Bibliographie

- 19 ( CHALES 81 )  
CHALES G., PEDRONO R., BURGEVIN P., INGELS PH., MONNIER E., LENOIR P.  
"Utilisation de l'ordinateur dans un système d'aide à l'enseignement du diagnostic médical (AEDM)"  
Med. Inform., vol 6, n° 1, 1981, pp. 41-49.
- 20 ( CHEESEMAN 83 )  
CHEESEMAN P.  
"A Method of Computing Generalised Bayesian Probability"  
Proc. IJCAI - 1983, Karlsruhe, West-Germany, pp. 198-202.
- 21 ( CINQUIN 84 )  
CINQUIN Ph.  
"Réflexions Préliminaires sur le Développement d'un Système d'Information Hospitalière Intégré"  
Thèse doc. med., Fac. Med. de Grenoble, 12-10-1984.
- 22 ( CLANCEY 84 )  
CLANCEY W.J., SHORTLIFFE E.H.  
"Readings in Medical Artificial Intelligence : the First decade"  
Addison-Wesley, Reading, 1984
- 23 ( CORDIER 84 )  
CORDIER M.O.  
"Les systèmes experts"  
La Recherche, Janvier 1984.
- 24 ( CROFT 72 )  
CROFT D.J.  
"Is Computerized Diagnosis Possible ?"  
Comput. Biomed. Res., 5, 1972, p. 351.
- 25 ( CROFT 74 )  
CROFT J.D., MACHOL R.E.  
"Mathematical methods in medical diagnosis"  
Annals of Biomedical Engineering, 2, 1974, p. 69.
- 26 ( CUMBERBATCH 76 )  
CUMBERBATCH J., HEAPS H.S.  
"A disease-conscious method for sequential diagnosis by use of disease probabilities without assumption of symptom independance"  
Int. J. Biomed. computing, 4, 1976, pp. 61-78.
- 27 ( DAPHY 84 )  
DAPHY N.  
"Le score pronostique de la chirurgie tubaire réparatrice distale"  
Thèse Méd., Grenoble, 1984.
- 28 ( DAVIS 79 )  
DAVIS R.  
"Interactive Transfert of Expertise : Acquisition fo New Inference Rules"  
Art. Int., 12 (1979), pp. 121-157.

## Bibliographie

29 { DE LAGARDE 83 }

DE LAGARDE J.  
"Initiation à l'Analyse des Données"  
Ed. DUNOD, Paris, 1983.

30 { DELOBEL 82 }

DELOBEL C., ADIBA M.  
"Bases de données et systèmes relationnels"  
ed. Dunod Infor., Bordas, Paris, 1982.

31 { DEMPSTER 67 }

DEMPSTER A.P.  
"Upper and lower probabilities induced by a multivalued mapping"  
Ann. Math. Stat., 39, pp. 325-339.

32 { DEROUESNE 79 }

DEROUESNE C.  
"Peut-on simplifier l'enseignement de la Neurologie ? Une expérience pédagogique testée par ordinateur"  
Rev. Fse. d'Ed. Méd, II, 1, 1979, pp. 13-17.

33 { DOSTERT 82 }

DOSTERT A.F.  
"Approche Comparative de l'Intelligence Humaine et d'une Intelligence Artificielle"  
Thèse de Médecine, Strasbourg, 1982.

34 { DREYFUS 84 }

DREYFUS H.  
"L'Intelligence Artificielle - Mythes et Limites"  
ed. Flammarion, Paris, 1984.

35 { DUBOIS 83 }

DUBOIS D.  
"Modèles Mathématiques de l'Imprécis et de l'Incertain en Vue d'Applications aux Techniques d'Aide à la Décision"  
Thèse d'Etat, Université Scientifique et Médicale de Grenoble, 1983.

36 { DUBOIS 83 }

DUBOIS D., PRADE H.  
"Théorie des Possibilités"  
Masson ed., Paris, 1985, p. 248.

37 { DUDA 76 }

DUDA R.O., HART P.E., NILSON N.J.  
"Subjective Bayesian methods for rule-based inference systems"  
Proc. 1976 National Computer Conference, AFIPS Press, Juin 1976.

38 { DURBEC 76 }

DURBEC J.P., CORNEE J.  
"Application de l'Analyse des Correspondances à l'étude de la Sémiologie Clinique en Pathologie Pancréatique"  
C.R. Acad. Sci, Paris, 283, 1381.

## Bibliographie

39 ( DURRMEYER 78 )

DURRMEYER J.F.

"Fentes labio-vélo-palatines-contribution à l'établissement du système A.D.M"

Thèse médicale n° 179, Rennes, 1978.

40 ( EDWARDS 86 )

EDWARDS C.N., BUYSE M.L.

"Building a computerized international network for medical expertise : the birth defects information system"

Proc. IA-BIOMED 86, Art. Intel. and its impacts in biology and med., C.R.I.M. Montpellier (France), 29 sept.-2 oct. 86, pp. 115-119.

41 ( ELSTEIN 78 )

ELSTEIN A., SHULMAN L., SPRAFKA S.

"Medical Problem Solving - An Analysis of Clinical Reasoning"

Harvard University Press, 1978.

42 ( FARRENY 84 )

FARRENY H., PRADE M.

"A possibility theory-based approach to default and inexact reasoning"

Proc. Colloque International d'Intelligence Artificielle, Marseille 24-27 octobre 1984, T.2, pp. 47-56.

43 ( FARRENY 85 )

FARRENY H.

"Systèmes-Experts et médecine"

Nouv. Rev. Méd. Toulouse - avril 85, pp.17.

44 ( FEIGENBAUM 84 )

FEIGENBAUM

"La cinquième génération - le pari de l'intelligence artificielle à l'aube du 21ème siècle"

Inter. Edition - Paris 1984.

45 ( FIESCHI 83 )

FIESCHI M.

"SPHINX: Un Système d'Aide à la Décision en Médecine"

Thèse de doc. d'Et. en biol. hum., Fac-de Med. MARSEILLE, 23-03-1983

46 ( FIESCHI 84 a )

FIESCHI M.

"Intelligence artificielle en médecine"

ed. MASSON, Paris, 1984.

47 ( FIESCHI 84 b )

FIESCHI M., JOUBERT M.

"Quelques Réflexions sur l'évaluation des Systèmes Experts en Médecine"

Proc. Coll. MEDECINE INFORMATIQUE, Marseille, 30 nov. - 1er dec. 1984.

## Bibliographie

48 { FLEHINGER 75 }

FLEHINGER B.J., ENGLE Jr. R.L.

"HEME : A Self-Improving Computer Program for Diagnosis-Oriented Analysis of Hematologic Diseases"

IBM J. Res. DEVELOP., 19, 1975, p. 557.

49 { FOX 77 }

FOX J., ALVEY P.

"Computer Assisted Medical Decision Making"

British Medical Journal, vol 287 du 10/09/83.

50 { FOX 80 }

FOX J., BARBER D., BARDHAN K.D.

"Alternatives to Bayes ? A Quantitative Comparison with Rule-based Diagnostic Inference"

Meth. Inform. Med, vol 19, n°4, 1980, pp. 210-215.

51 { FRUTIGER 77 }

FRUTIGER P., ROSSIER Ph., SCHERRER J.R.

"Automatic ICD, SNOMED codification of symptoms, signs, and diagnostics using tree branching logic"

North-Hol. Publ. Comp., pp. 267-270.

52 { FRYBACK 78 }

FRYBACK D.G.

"Bayes' theorem and conditional non independence of data in medical diagnosis"

Computer and Biomedical Research 11, 1978, pp. 423-434.

53 { GALEN 75 }

GALEN R.S., GAMBINO S.R.

"Beyond Normality : The prediction value and efficiency of medical diagnosis"

J. Wiley, New-York, London, 1975, 235 pages.

54 { GARDARIN 86 }

GARDARIN & Coll.

"Compte rendu de débat: les systèmes-experts sont-ils plus que des interfaces évolués des systèmes de Bases de Données et de Connaissances"

Proc. 8e Journées Francophones. sur l'Inf: "Bases de Données et Bases de connaissances", Grenoble, 14 & 15 Jan. 86, pp. 211-223.

55 { GARNIER 80 }

GARNIER M., DELAMARE V.

"Dictionnaire des Termes Techniques de Médecine"

Maloine ed., Paris, 1980.

56 { GASCHING 79 }

GASCHING J.

"Preliminary performance analysis of the prospector consultant system for mineral exploration"

Proceed. Sixth. I.J.L.A.I. Tokyo, 1979, pp. 308-310.

## Bibliographie

- 57 { GASCUEL 81 }  
GASCUEL O.  
"Un Systeme Expert dans le Domaine Médical"  
Thési 3e cy., Univ. P et M CURIE, publi. GR 22. - CNRS, nov. 81.
- 58 { GORRY 68 }  
GORRY G.A.  
"Stratégie for computer-aided diagnosis"  
Math. Biosciences, 2, 1968, pp. 293-318.
- 59 { GOUVERNET 79 }  
GOUVERNET J.  
"Apport des Méthodes de Classification en Génétique Médicale -  
Application aux maladies osseuses constitutionnelles"  
Thèse de Médecine, Fac. de Méd. de Marseille, 25/09/79.
- 60 { GOUVERNET 85 a }  
GOUVERNET J. Ayme S.  
"GENDIAG : Système d'aide au diagnostic en Génétique Médicale"  
Médecine Informatique, n° 15-16, pp. 7-8.
- 61 { GOUVERNET 85 b }  
GOUVERNET J., CARABOEUF M., Ayme S.  
"GENDIAG : A Computer-assisted Facility in Medical Genetics Based on  
Belief Functions"  
Math. Inform. Med. 24 (1985) pp. 177-180.
- 62 { GREMY 69 }  
GREMY F., SALMON D., JOLY M.  
"Quelques aspects de la théorie de la décision en médecine, appli-  
cations du théorème de Bayes"  
Revue Franc. d'Etudes Clin. et Biol., 1969, XIV, 88-95.
- 63 { GRENIER 83 }  
GRENIER B., DRUCKER J.  
"Les valeurs prédictives, bases fondamentales de la décision médi-  
cale"  
Revue d'Education Médicale, 1983, tome VI, n° 4.
- 64 { HELSER 78 }  
HELSEr J.F., BROOKS R.E. BALLARD J.P.  
"Progress Report : computerized psycho-pharmacology advisor"  
11th Collegium Int., Neuropsychopharmacologieum, Vienne, 1978.
- 65 { HUDSON 84 }  
HUDSON D.L., ESTRIN T.  
"EMERGE : a data-driven medical decision making aid"  
PAMI, vol 6, n° 1, janvier 1984.
- 66 { JOUBERT 81 }  
JOUBERT M.  
"Etude et réalisation d'un programme de dialogue homme-machine  
d'aide à la décision en médecine"  
Thèse de Doctorat d'Etat en Biologie Humaine, Fac. Méd. Marseille,  
1981.

## Bibliographie

- 67 ( KAISER 84 )  
KAISER D.  
"Examen des Diverses Méthodes utilisées en Représentation des Connaissances"  
Proc. IVe congrès Rec. des formes et I.A, AFCET, Paris, jan. 84, pp. 115-144.
- 68 ( KASSIRER 78 )  
KASSIRER J., GORRY G.  
"Clinical Problem Solving - A Behavioral Analysis"  
Ann. Int. Med., 89 (1978), pp. 245-255.
- 69 ( KAUFMAN 75 )  
KAUFMAN A.  
"Introduction to the theory of fuzzy subsets"  
Vol. 1, Fundamental Theoretical Elements, Academic Press, 1975, New-York.
- 70 ( KERGOSIEN 85 )  
KERGOSIEN Y.L.  
"Sémiotique de la nature"  
in IVe Séminaire de l'Ec. de Biol. Théo., Ed. CNRS, Paris, 1985, pp. 11-26.
- 71 ( KOHLER 84 )  
KOHLER F., MONCHOVET S., GROUSSIN M., MARTIN J.  
"SELF, un générateur simple d'aide à la décision médicale - Application de quelques outils logiques"  
Journées Informatique et Médecine, Marseille, 30/11/84 et V12/84.
- 72 ( LABBE 85 )  
LABBE J.  
"Proposition d'un score physiologique de risque péri-opératoire"  
Thèse Méd., Grenoble, 1985.
- 73 ( LACAINÉ 78 )  
LACAINÉ F., HUGIER M., REMY F.  
"L'Efficacité d'un examen à pour but diagnostique : de la donnée à la décision médicale"  
Nouvelle Presse Médicale, 1978, 7 (11), pp. 1451-1453.
- 74 ( LAURENT 83 )  
LAURENT J.P., MORLAYE B.  
"L'Intelligence Artificielle, la démarche, les domaines"  
Gazette des Mathématiciens, sept.83.
- 75 ( LAURENT 84 )  
LAURENT J.P.  
"La structure de contrôle dans les systèmes experts"  
Technique et Science Informatiques, vol 3, n° 3, 1984.
- 76 ( LENAT 83 )  
LENAT D.B.  
"EURISKO : a program that learns new heuristics and domain concepts"  
Art. Int., vol 21, 1983, pp. 61-98.

## Bibliographie

- 77 ( LENOIR 80 a )  
LENOIR P., CHALES G.  
"Pourquoi et comment aider les médecins à porter des diagnostics"  
Med. Inform., vol 5, n° 4, 1980, pp. 281-289.
- 78 ( LENOIR 80 b )  
LENOIR P., BOUREL M., ROGER M., CHALES G.  
"Systèmes d'aide au diagnostic médical : méthodes utilisées"  
Med. Inform., vol 5, n° 4, 1980, pp. 291-307.
- 79 ( LENOIR 80 c )  
LENOIR P., CHALES G.  
"Efficacité de l'ADM sur les succès et les coûts du diagnostic"  
Med. Inform., vol 5, n° 4, 1980, pp. 309-315.
- 80 ( LENOIR 81 a )  
LENOIR P., CHALES G., ROGER M.  
"Dérivés du système ADM"  
Med. Inform., vol 6, n° 1, 1981, pp. 33-40.
- 81 ( LENOIR 81 b )  
LENOIR P., ROGER M., FRANGEUL C., CHALES G.  
"Réalisation, développement et maintenance de la base de données ADM"  
Med. Inform., vol 6, n° 1, 1981, pp. 51-56.
- 82 ( LESMO 82 )  
LESMO L., SAITTA L., TORRASSO L.  
"Learning of Fuzzy Production Rules for Medical Diagnosis"  
Approximate Reasoning in Decision Analysis, North-Holland  
Amsterdam, (1982), pp. 249-260.
- 83 ( MAC GIBBON 79 )  
MACGIBBON B., PREUS M.  
"The Distorted Shell Method of Clustering for Syndrome Classification"  
Am. J. Hum. Genet., vol 31, 1979, pp. 498-507.
- 84 ( MEDLAB )  
MEDLAB  
"HELP/HCOM USER's reference manual"  
MEDLAB COMPANY, a division of Control Data Corporation.
- 85 ( METZ 78 )  
METZ Ch. E.  
"Basic Principles of ROC Analysis"  
Seminar in Nuclear Med., vol 8(4), 1978, pp. 283-298.
- 86 ( MICHALSKI 83 )  
MICHALSKI R.S.  
"A Theory and Methodology of Inductive Learning"  
Art. Int., 22, 1983, pp. 111-161.

## Bibliographie

87 ( MITTAL 84 )

MITTAL S., CHANDRASEKARAN B., STICKLEN J.  
"Patrec : A knowledge - Directed Database for a Diagnostic Expert System"  
IEEE-Computein, sept. 84, pp. 51-58.

88 ( Mc NEIL 75 )

Mc NEIL B.J., KEELER E., ADELSTEIN S.J.  
"Primer on certain elements of Medical Decision Making"  
New Engl. J. of. Med., vol. 293(5), pp. 211-215.

89 ( NESTOR 84 )

NESTOR C.G.  
"A medical decision support that integrates causal, temporal and probabilistic knowledge"  
Ph. D., Computer Sci. Dept., Stanford U., 1984.

90 ( NORUSIS 75 )

NORUSIS M.J., JACQUEZ J.A.  
"Diagnosis, I. Symptom Non dependence in Mathematical Models of Diagnosis"  
Comput. Biomed. Res. 8, 156 (1975).

91 ( PAUKER 76 a )

PAUKER S.G., et al.  
"Towards the Simulation of Clinical Cognition"  
Am. J. Med., 60 (1976), pp. 981-996.

92 ( PAUKER 76 b )

PAUKER S.G., GORRY G.A., KASSIRER J.P., SCHWARTZ W.B.  
"Toward the simulation of Clinical Cognition. Tacking a Present Illness by Computer"  
American Journal of Medicine, Vol. 60 (1976); pp. 981-996.

93 ( PEDRONO M.J. 81 )

PEDRONO M.J.  
"Présentation d'un système expert d'aide au diagnostic différentiel dans le domaine médical"  
Thèse doctorat de 3e cycle d'informatique, Univ. de Rennes I, 1981.

94 ( PEDRONO R. 79 )

PEDRONO R.  
"L'A.E.M.I un système d'aide à l'enseignement médical"  
Thèse doctorat de 3e cycle d'informatique, Univ. de Rennes I, 1979.

95 ( POLITAKIS 82 )

POLITAKIS P.G.  
"Using Empirical Analysis to Refine Expert System Knowledge Bases"  
Technical Report CBM-TR-130, Dept. of Comp. Sci, Rutgers Univ., oct. 82.

96 ( POLITAKIS 84 )

POLITAKIS P.G., WEISS S.L.  
"Using Empirical Analysis to Refine Expert System Knowledge Bases"  
Art. Int., 22, 1984, pp. 23-48.

## Bibliographie

- 97 ( POPLE 75 a )  
POPLE H.E., MYERS J.D., MILLER R.A.  
"Dialog : a Model of Diagnostic Logic in Internal Medicine"  
Proc. 4th IJCAI, Tbilissi, USSR, 1985.
- 98 ( POPLE 75 b )  
POPLE H.E.  
"The Formation of Composite Hypotheses in Diagnostic Problem Solving : an Exercice in Symthetic Reasoning"  
Proc. 5th IJCAI, Cambridge, Mass., USA, 1977.
- 99 ( PRADE 82 )  
PRADE H.  
"Modèles Mathématiques de l'Imprécis et de l'Incertain en Vue d'Applications au Raisonnement Naturel"  
Thèse d'Etat n° 1048, Université Paul SABATIER, Toulouse, 1982.
- 100 ( PRADE 83 )  
PRADE H.  
"A Synthetic View of Approximate Reasoning Techniques"  
Proceedings of the 8th IJCAI, pp. 130-136, Karlsruhe Aout 1983.
- 101 ( PRADE 84 )  
PRADE H.  
"Modèles de Raisonnement Approché pour les Systèmes Experts"  
Proc. 4ème Congrès Reconnaissance des Formes et Intelligence Artificielle - Paris 25-27 janvier 1984. Tome 2, pp. 355-373.
- 102 ( PREUS 83 )  
PREUS M., AYME S.  
"Formal Analysis of Dysmorphism : objective methods of syndrom definition"  
Clini. Gen., 1983, vol. 23, pp. 1-16.
- 103 ( PRYOR 82 )  
PRYOR T.A., GARDNER R.M., CLAYTON P.D., WARNER H.R.  
"The HELP System"  
Proceedings of the 6th Annual Symposium on Computer Applications in Medical Care, octobre-novembre 1982.
- 104 ( REALE 68 )  
REALE A., MACCACCARA A., ROCCA E. et al.  
"Computer Diagnosis of Congenital Heart Disease"  
Comput. Biomed. Res., vol 1, 1968, p. 533.
- 105 ( RECHENMANN 85 )  
RECHENMANN F., CHENIN P. & al.  
"Représentation des Connaissances dans les logiciels de Calcul Scientifique"  
Proc. Congrès " Le Calcul Demain ", Wiley et Masson eds., Grenoble, dec. 75.

## Bibliographie

106 { REGGIA 82 }

REGGIA

"Computer Assisted Medical Decision Making"

Applications of Computers in Medicine, M. Schwartz, ed., IEEE Press, 1982, pp. 198-213.

107 { RIALLE 85 }

RIALLE V.

"Génération de Connaissances Expertes - Application à la Génétique Médicale"

Proc. Ve Int. Workshop. on Exp. Synt., Avignon, 13-15/5/85, vol II, pp. 901-913.

108 { RIALLE 86 }

RIALLE V.

"Aide au diagnostic et à l'apprentissage dans un domaine médical incertain, incomplet et évolutif"

Proc IA-BIOMED 86, C.R.I.M., Montpellier, pp. 224-233.

109 { ROBERT 86 }

ROBERT C., ZARSKI M., DEMONGEOT J.

"Utilisation de la notion de score comme outil décisionnel en médecine, proposition pour la fabrication automatique de règles de production"

Proc IA-BIOMED 86, C.R.I.M., Montpellier, pp. 234-244.

110 { ROUSSEAU 86 }

ROUSSEAU B., RECHENMANN F., LAUDAU M.

"Edora projet: Artificial Intelligence Approach and Work Station Concept to aid Dynamic Modeling in Biology and Ecology"

1986 Summer Computer Simulation Conf., Reno, Nevada (U.S.A.), 28-30 juillet 1986.

111 { SALAMON 76 }

SALAMON R., BERNARDET M., SAMSON M., DEROUESNE C., GREMY F.

"Bayesian Method Applied to Decision Making in Neurology-Methodological Considerations"

Methods of Information in Medicine, 15(1976), pp. 174-179.

112 { SANCHEZ 79 }

SANCHEZ E.

"Medical Diagnosis and composite Fuzzy Relations"

Advances in Fuzzy Set Theory and Applications, GUPTA M.M., RAGADE R.K. and YAGER R.R. (Eds), North-Holland, Amsterdam 437-44.

113 { SCHILD 78 }

SCHILD W., LUNENFELD B., GAVISH B.

"Computer-Aided Diagnosis with an Application to Endocrinology"

IBM J. of Res. and. dev., vol 22, n° 5, 1978, pp. 518-532.

114 { SCHORDERET 85 }

SCHORDERET D., AEBISCHER P.

"SYNDROC : microcomputer based differential diagnosis of malformation patterns"

Archives of Dism. Child., 1985, vol 60, pp. 248-251.

## Bibliographie

115 { SHAFER 76 }

SHAFER G.

"A Mathematical Theory of Evidence"

Pr. University Press, 1976.

116 { SHAPIRO 76 }

SHAPIRO A.R., PORTA J.

"Comparison of prediction algorithms : physician versus computer"

Presented at Ninth Biometric Conference, Boston, August 23, 1976.

117 { SHAPIRO 77 }

SHAPIRO A.R.

"The evaluation of clinical prediction"

The New England Journal of Medicine, Vol. 296, 26 (1977), pp. 1509-1514.

118 { SHIZUKA 81 }

SHIZUKA M., FU K.S., YAO J.T.P.

"A Theoretical Treatment of Certainty Factor in Production Systems"

Structural Eng. Report, Purdue University, CE-STR-81-6, 1981.

119 { SHORTLIFFE 75 }

SHORTLIFFE 75 } SHORTLIFFE E.H., BUCHANAN B.G.

"A Model of Inexact Reasoning in Medicine"

Mathematical Biosciences, 23 (1975), pp. 351-379.

120 { SHORTLIFFE 76 }

SHORTLIFFE E.H.

"MYCIN : Computer-Based Medical Consultation"

American Elsevier, New-York, 1976.

121 { SMETS 78 }

SMETS Ph.

"Un modèle mathématico-statistique simulant le processus du diagnostic médical"

Mémoire Agreg. ens. sup., Université Libre, Bruxelles, 1978.

122 { SNOMED 75 }

College of American Pathologists

"Systematized Nomenclature of Medicine (SNOMED)"

Committee on Nomenclature and Classification of Diseases, Chairman R.A. Cote, M.D., trial version, 1975.

123 { SNOP 65 }

College of American Pathologists

"Systematized Nomenclature of Pathology (SNOP)"

Chicago, Illinois (USA), 1965.

124 { SOULA 81 }

SOULA G.

"Aide à la décision en logique floue : Application en médecine"

Thèse d'Etat en Biologie Humaine, Faculté de Médecine, Marseille, 1981.

## Bibliographie

125 { SOULA 82 }

SOULA G., GOVERNET J., BARRE A., SAN MARCO J.L.  
"Application of fuzzy relations to medical decision making"  
MEDINFO 80, Tokyo, Proc. North-Holland II, 1980, pp. 844-848.

126 { SOULA 83 }

SOULA G., VIALETES B., SAN MARCO J.L.  
"Protis : a fuzzy deduction-rules system : application to the  
treatment of diabetes"  
MEDINFO 83, Amsterdam.

127 { SULTAN 85 }

SULTAN C.  
"Les Systèmes Experts en médecine : un atout indispensable"  
Inform. de France, oct. 85, p. 13.

128 { SULTAN 86 }

SULTAN C.  
"Généralités sur les systèmes d'aide à la décision"  
Proceed. 1er Coll. I.A., soc. NDV, Paris, 21 fév. 86.

129 { SZOLOVITS 78 }

SZOLOVITS P., PAUKER S.G.  
"Categorical and Probabilistic Reasoning in Medical Diagnosis"  
Artificial Intelligence, 11 (1978), pp. 115-144.

130 { SZOLOVITS 82 }

SZOLOVITS P.  
"Artificial Intelligence in Medicine"  
AAAS Selected Symposia Series, Westview Press Inc., Colorado (USA),  
1982.

131 { VALOIS 81 }

VALOIS J.  
"La santé en fiches. Chroniques médicale et hospitalière"  
Ed. Fures Impression, P.U.F., 1981.

132 { VELOSO 86 }

VELOSO M.L.M., FEIJOO M.  
"DYSMOR : Computer Based Differential Diagnosis of Dysmorphic  
Syndroms"  
MEDINFO 86, Elsevier Sci. Publ. B.V., North-Holland, pp. 216-218.

133 { VENOT 79 }

VENOT A., GRANDJOUANS, STEINMER J.L.  
"Improvement of cholestyramine therapy through mathematical modelling of  
T.C. 99 in diethyl IDA pharmacokinetics"  
Proc. in Med. Imaging, INSERM, vol 88, 1979, pp. 459-468.

134 { WALKER 86 }

WALKER M.G., BLUM R.L.  
"Towards Automated Discovery from Clinical Data Base : the RADIX  
Project"  
MEDINFO 86, Elsevier Sci. Publ. B.V., North-Holland, pp. 32-36.

## Bibliographie

135 { WARNER 61 }

WARNER H.R., TORONTO A.F., VEASEY L.G., et al.  
"A mathematical approach to medical diagnosis : appliation to  
congenital heart disease"  
JAMA 177 1981, pp. 177-183.

136 { WARNER 72 a }

WARNER H.R., OLMSTEAD C.M., RUTHERFORD B.D.  
"HELP - A program for Medical Decision Making"  
Comput. Biomed. Res, 5, 65 (1982).

137 { WARNER 72 b }

WARNER H.R., RUTHERFORD B.D., HOUTCHENS B.  
"A sequential bayesian approach to history tacking and diagnosis"  
Comput. Biomed. Res, 5, 1972, pp. 256-262.

138 { WARNER 79 }

WARNER H.R.  
"Computer-Assisted Medical Decision-Making"  
Academic Press, New-York, p. 190.

139 { WECHSLER 76 }

WECHSLER W.  
"A fuzzy approach to medical diagnosis"  
International Journal of Bio-Medical Computing. (1976) pp. 191-203.

140 { WINKLER 67 }

WINKLER R.L.  
"The quantification of judgment : some methodological suggestions"  
J. Am. Statist. Assoc. 62 (1967), pp. 1105-1120.

141 { WOODBURY 78 }

WOODBURY M.A., CLIVE J., GARSON A.  
"Mathematical typology : a grade of membership technique for obtain-  
ing disease definition"  
Computer and Biomedical Research 11, 1978, pp. 277-298.

142 { ZADEH 75 a }

ZADEH L.A.  
"Fuzzy Logic and Approximate Reasoning"  
Synthese, Vol. 30, pp. 407-425, 1975.

143 { ZADEH 75 b }

ZADEH L.A.  
"The Concept of Linguistic Variable and its Applications to  
Approximate Reasoning"  
Information Science, Vol. 8, pp. 199-249 et pp. 301-357 et Vol. 9,  
pp. 43-80.

144 { ZADEH 78 }

ZADEH L.A.  
"Fuzzy sets as a basis for a theory of possibility"  
Fuzzy sets and systems, vol. 1, 1978, pp. 3-28.



## ANNEXES

## ANNEXE 1

## LISTE DES AXES:

1ere colonne : numéro de l'axe.  
 2ème colonne : niveau d'observation.  
 3ème colonne : type : E : entité symptomatique élémentaire,  
                   S : score.  
 4ème colonne : nom de l'axe.

1	1	E	APPAREIL GENITAL EXTERNE --> AMBIGUITE
2	1	E	APPAREIL GENITAL EXTERNE --> TESTICULE
3	1	E	ARTICULATION-GENOU --> DEFORMATION
4	1	E	CROISSANCE --> LONGEVITE
5	1	E	ETAT NEONATAL --> ETAT RESPIRATOIRE
6	1	E	GROSSESSE NAISSANCE --> DECES NEONATAL
7	1	E	GROSSESSE NAISSANCE --> ETAT NEONATAL
8	1	E	MAIN-DOIGT-POUCE --> HYPOPLASIE
9	1	E	MEMBRE --> MEMBRE SUPERIEUR
10	1	E	MEMBRE SUPERIEUR-MAIN --> DOIGT
11	1	E	MEMBRE SUPERIEUR-OMOPLATE --> POSITION
12	1	E	MORPHOLOGIE CRANIO-FACIALE --> CRANE
13	1	E	MORPHOLOGIE CRANIO-FACIALE --> NEZ
14	1	E	MORPHOLOGIE CRANIO-FACIALE --> OEIL
15	1	E	MORPHOLOGIE CRANIO-FACIALE --> PAVILLON DE L' OREILLE
16	1	E	NEZ --> AILE
17	1	E	NEZ --> POSITION
18	1	E	OEIL --> DEVELOPPEMENT
19	1	E	OEIL --> TAILLE
20	1	E	PAVILLON DE L' OREILLE --> CONDUIT AUDITIF EXTERNE
21	1	E	PAVILLON DE L' OREILLE --> TAILLE
22	1	E	PHANERE --> CILS
23	1	E	PHANERE --> SOURCIL
24	1	E	PUBERTE --> DEVELOPPEMENT
25	1	E	RACINE --> APPAREIL GENITAL EXTERNE
26	1	E	RACINE --> MEMBRE
27	1	E	RACINE --> MORPHOLOGIE CRANIO-FACIAL
28	1	E	SYSTEME NERVEUX CENTRAL --> DEVELOPPEMENT
29	1	E	THORAX --> OMOPLATE
30	1	E	THORAX --> REGION STERNALE
31	1	E	TRONC --> THORAX
32	1	S	III.A
33	1	S	III.C
34	1	S	III.D
35	1	S	III.E
36	1	S	III.F
37	1	S	III.G
38	1	S	III.H
39	1	S	III.I
40	1	S	III.J
41	1	S	III.K
42	1	S	III.L
43	1	S	III.M
44	1	S	III.N
45	1	S	III.P
46	1	S	III.R
47	1	S	III.S
48	1	S	III.U

## ANNEXES

49	1	S	IIII.B
50	1	S	IIII.D
51	1	S	IIII.F
52	1	S	IIII.G
53	1	S	IIII.H
54	1	S	V.A
55	1	S	V.B
56	1	S	V.D
57	1	S	VI.A
58	1	S	VI.B
59	1	S	VI.D
60	1	S	VI.F
61	1	S	VI.G
62	1	S	VI.H
63	1	S	VI.I
64	1	S	VI.J
65	1	S	VI.K
66	1	S	VII.A
67	1	S	VII.B
68	2	E	COEUR --> SOUFFLE
69	2	E	OEIL --> SEGMENT ANTERIEUR
70	2	E	OEIL --> TENSION
71	2	E	RACINE --> APPAREIL CARDIOVASCULAIRE
72	2	E	TRONC --> DEVELOPPEMENT
73	2	E	VAISSEAU-ARTERE --> PRESSION
74	2	S	I.A
75	2	S	I.B
76	2	S	I.C
77	2	S	II.A
78	2	S	II.B
79	2	S	III.B
80	2	S	III.Q
81	2	S	III.T
82	2	S	III.V
83	2	S	IIII.I
84	2	S	IIII.J
85	2	S	V.C
86	2	S	VIIII.A
87	2	S	XIIII.D
88	2	S	XIIII.E
89	2	S	XIIII.F
90	3	E	COEUR --> RYTHME
91	3	E	COLONNE --> DISRAPHIE
92	3	E	COLONNE VERTEBRALE-CERVICALE --> FORME
93	3	E	COTE --> NOMBRE
94	3	E	CRANE --> FAUX DU CERVEAU
95	3	E	MEMBRE INFERIEUR-BASSIN --> AILE ILIAQUE
96	3	E	METABOLISME-MINERAL --> CALCEMIE
97	3	E	PLEVRE --> EPANCHEMENT
98	3	E	RACINE --> SQUELETTE
99	3	E	SQUELETTE --> COLONNE VERTEBRALE
100	3	E	SQUELETTE --> COTE
101	3	E	SQUELETTE --> STERNUM
102	3	E	TRONC --> DIAPHRAGME
103	3	E	TUBE DIGESTIF --> OESOPHAGE
104	3	S	VI.E
105	3	S	VIIII.C
106	3	S	XII.A

## ANNEXES

107	3	S	XIII.A
108	3	S	XIII.C
109	3	S	XIII.D
110	3	S	XIII.E
111	3	S	XIII.F
112	3	S	XIII.H
113	3	S	XIII.I
114	3	S	XIII.J
115	3	S	XIII.K
116	3	S	XIII.L
117	3	S	XIII.M
118	3	S	XIII.N
119	3	S	XIII.O
120	3	S	XIII.P
121	3	S	XV.E
122	4	E	APPAREIL GENITO-URINAIRE --> APPAREIL URINAIRE
123	4	E	APPAREIL URINAIRE --> REIN
124	4	E	APPAREIL URINAIRE-REIN --> GLOMERULE
125	4	E	ENDOCRINE ONCOLOGIE METABOLISME --> ENDOCRINE
126	4	E	GLANDE ANNEXE-PANCREAS --> DEVELOPPEMENT
127	4	E	GROSSESSE NAISSANCE --> MONSTRE
128	4	E	METABOLISME-PROTEINE --> AMINOACIDEMIE
129	4	E	RACINE --> APPAREIL DIGESTIF
130	4	E	SYST. NERV. CENT. PERIPH. ET APP. SENSORIEL --> SYSTEME
131	4	E	SYSTEME NERVEUX CENTRAL --> TUMEUR
132	4	E	TUBE DIGESTIF-MESENTERE --> DEVELOPPEMENT
133	4	E	VAISSEAU --> CANAL ARTERIEL
134	4	E	VAISSEAU-ARTERE --> ARTERE CORONAIRE
135	4	E	VAISSEAU-VEINE --> CAVE SUPERIEURE GAUCHE
136	4	S	I.D
137	4	S	IIII.A
138	4	S	IIII.C
139	4	S	IIII.E
140	4	S	VIII.A
141	4	S	VIII.B
142	4	S	VIII.C
143	4	S	VIII.D
144	4	S	VIII.E
145	4	S	VIII.G
146	4	S	VIII.H
147	4	S	VIII.I
148	4	S	VIII.J
149	4	S	VIII.K
150	4	S	VIIII.B
151	4	S	VIIII.E
152	4	S	X.A
153	4	S	X.B
154	4	S	X.C
155	4	S	X.F
156	4	S	X.G
157	4	S	X.I
158	4	S	XI.A
159	4	S	XI.B
160	4	S	XI.C
161	4	S	XI.D
162	4	S	XII.B
163	4	S	XII.C
164	4	S	XIII.C

ANNEXES

165	4	S	XIIII.A
166	4	S	XIIII.B
167	4	S	XIIII.C
168	4	S	XIIII.G
169	4	S	XIIII.H
170	4	S	XIIII.I
171	4	S	XV.A
172	4	S	XV.B
173	4	S	XV.C
174	4	S	XV.D

## ANNEXES

## ANNEXE 2

## LISTE FRAGMENTAIRE DES ENTITES SYMPTOMATIQUES ET DE LEURS VALEURS

La liste est limitée aux deux premiers chapitres sur les quinze chapitres existants.

Pour chaque entité, sont donnés:

son numéro,

sa branche arborescente ascendante,

son nom proprement dit,

sa liste actuelle de valeurs (susceptible d'augmenter au fur et à mesure de l'enregistrement de nouveaux syndromes dans la base)  
cette liste comprend : .le numéro de symptôme ayant cette valeur,  
.le niveau d'observation,  
.l'expression naturelle de la valeur,  
.l'expression numérique de la valeur.

---

 C H A P I T R E 1

E.S. no. 1  
ETAT NEONATAL  
ETAT RESPIRATOIRE

462 1 EN DETRESSE -1

---

E.S. no. 2  
ETAT NEONATAL  
POIDS DE NAISSANCE

224 2 ELEVE 1  
315 2 FAIBLE -1

---

E.S. no. 3  
ETAT NEONATAL  
TONUS MUSCULAIRE

575 1 FAIBLE -1

---

E.S. no. 4  
GESTATION  
DUREE

336 1 SUPERIEURE A 42 SEM 1  
337 1 INFERIEURE A 36 SEM -1

---

E.S. no. 5  
GESTATION  
MOTRICITE FOETALE

## ANNEXES

573	2	FAIBLE	-1
-----	---	--------	----

---

E.S. no.	6	GESTATION NOMBRE	
----------	---	---------------------	--

496	4	GEMELLAIRE	1
-----	---	------------	---

---

E.S. no.	7	GESTATION PRESENTATION FOETALE	
----------	---	-----------------------------------	--

574	1	EN SIEGE	-1
-----	---	----------	----

---

E.S. no.	8	GESTATION-ANNEXE FOETALE ARTERE OMBILICALE	
----------	---	---	--

606	1	UNIQUE	-1
-----	---	--------	----

---

E.S. no.	9	GESTATION-ANNEXE FOETALE LIQUIDE AMNIOTIQUE	
----------	---	--	--

539	2	EN EXCES	1
-----	---	----------	---

---

E.S. no.	10	GESTATION-ANNEXE FOETALE PLACENTA	
----------	----	--------------------------------------	--

640	1	ANORMAL	-1
-----	---	---------	----

---

E.S. no.	11	GESTATION-RISQUE INTOX	
----------	----	---------------------------	--

548	2	AVEC INTOXICATION	1
-----	---	-------------------	---

---

E.S. no.	12	GESTATION-RISQUE TRAITEMENT	
----------	----	--------------------------------	--

544	2	AVEC TRAITEMENT	1
-----	---	-----------------	---

---

E.S. no.	13	GROSSESSE NAISSANCE DECES NEONATAL	
----------	----	---------------------------------------	--

106	1	NEONATAL	2
-----	---	----------	---

107	2	IN UTERO	1
-----	---	----------	---

108	1	AVANT 7 JOURS	3
-----	---	---------------	---

---

E.S. no.	14	GROSSESSE NAISSANCE ETAT NEONATAL	
----------	----	--------------------------------------	--

## ANNEXES

378	1	EN DETRESSE	-1
-----	---	-------------	----

---

E.S. no.	15	GROSSESSE NAISSANCE MONSTRE	
----------	----	--------------------------------	--

497	4	DOUBLE	1
-----	---	--------	---

---

C H A P I T R E            2

E.S. no.	16	CROISSANCE LONGEVITE	
----------	----	-------------------------	--

109	1	à MOINS DE 1 AN	1
110	1	à MOINS DE 15 ANS	2
403	1	NORMALE	0

---

E.S. no.	17	CROISSANCE POIDS	
----------	----	---------------------	--

386	2	INSUFFISANT	-1
-----	---	-------------	----

---

E.S. no.	18	CROISSANCE TAILLE	
----------	----	----------------------	--

54	2	GRANDE	1
55	2	PETITE	-1
56	2	NORMALE	0
282	1	NAINE	-2
653	2	GEANTE	2

---

E.S. no.	19	CROISSANCE VITESSE	
----------	----	-----------------------	--

338	2	LENTE	-1
-----	---	-------	----

---

E.S. no.	20	CROISSANCE HARMONIE SYMETRIE	
----------	----	---------------------------------	--

654	2	GEANTE ASYMETRIQUE	1
-----	---	--------------------	---

---

E.S. no.	21	CROISSANCE-LONGEVITE-PUBERTE CROISSANCE	
----------	----	--	--

360	2	NORMALE	0
-----	---	---------	---

---

E.S. no.	22	MENSURATION SEGMENTAIRE	
----------	----	-------------------------	--

## ANNEXES

## DYSHARMONIE

292	2	NAINES DYSHARMONIEUX	-1
-----	---	----------------------	----

---

E.S. no.	23	MENSURATION SEGMENTAIRE	
		RAPPORT SEG SUP/SEG INF	

57	2	INSUFFISANT	-1
58	2	EXCESSIF	1

---

E.S. no.	24	PUBERTE	
		DEVELOPPEMENT	

392	1	RETARDE	-1
-----	---	---------	----

---

## ANNEXES

## ANNEXE 3

## LISTE DES STRUCTURES DE FICHIERS DBASE III

## A / FICHER SYMPTOME.DBF :

LIST

Record#	FIELD_NAME	FIELD_TYPE	FIELD_LEN
1	L1	C	1
2	FORMENATUR	C	90
3	L2	C	1
4	NUMERO	N	4
5	L3	C	1
6	ENTITSYMP	C	45
7	L4	C	1
8	CODENTITSY	C	4
9	L5	C	1
10	VALNATUREL	C	25
11	L6	C	1
12	VALNUM	N	3
13	L7	C	1
14	NIVOSAXE	C	4
15	L8	C	1
16	ATTRIBUT	C	46
17	L9	C	1
18	PERE	N	4
19	L10	C	1
20	FILS	N	4
21	L11	C	1
22	FRERE	N	4
23	NIVOBSE	N	1
24	L12	C	1
25	NBDESYND	N	3
26	L13	C	1
27	ADLISTSYND	N	4
28	L14	C	1
29	CHAPITRE	N	2
30	L15	C	1
31	ESMERE	C	82
32	L16	C	1
33	AXE	C	57
34	L17	C	1
35	NUMAXE	N	3
36	L18	C	1
37	TYPAXE	C	2

## ANNEXES

**B / FICHER SYNDROME.DBF :**

LIST

Record#	FIELD_NAME	FIELD_TYPE	FIELD_LEN
1	L1	C	1
2	NOMSYND	C	120
3	L2	C	1
4	NUMERO	N	4
5	L3	C	1
6	CODESYND	C	3
7	L4	C	1
8	NBSYMP	C	4
9	L5	C	1
10	RENSEIGT	C	33
11	L6	C	1
12	PREVALENCE	N	1
13	L7	C	1
14	ADLISTSYMP	C	4
15	L8	C	1
16	ADLSTES	C	4

**C / FICHER ES.DBF :**

LIST

Record#	FIELD_NAME	FIELD_TYPE	FIELD_LEN
1	ESMERE	C	68
2	ENTITE	C	45
3	NUM	N	3
4	NIVOBS	N	1
5	CHAPITRE	N	2
6	NOMAXE	C	70
7	TYPE	C	1
8	CARDINAL	N	2
9	S1	N	4
10	S2	N	4
11	S3	N	4
12	S4	N	4
13	S5	N	4
14	S6	N	4
15	S7	N	4
16	S8	N	4
17	S9	N	4
18	S10	N	4
19	S11	N	4
20	S12	N	4

## ANNEXES

## D / FICHER LSTSCORE.DBF :

LIST

Record#	FIELD_NAME	FIELD_TYPE	FIELD_LEN
1	NUMAXE	N	3
2	NIVOBS	N	1
3	NIVOSAXE	N	1
4	CHAPITRE	N	2
5	TYPE	C	1
6	NUM	N	3
7	NOMAXE	C	70
8	CARDINAL	N	2
9	E1	N	3
10	E2	N	3
11	E3	N	3
12	E4	N	3
13	E5	N	3
14	E6	N	3
15	E7	N	3
16	E8	N	3
17	E9	N	3
18	E10	N	3
19	E11	N	3
20	E12	N	3
21	E13	N	3
22	E14	N	3
23	E15	N	3
24	E16	N	3
25	E17	N	3
26	E18	N	3
27	E19	N	3
28	E20	N	3

## ANNEXES

## E / FICHER AXES.DBF :

LIST

Record#	FIELD_NAME	FIELD_TYPE	FIELD_LEN
1	NUMAXE	N	3
2	NIV OBS	N	1
3	NIV OBS AXE	N	1
4	CHAPITRE	N	2
5	TYPE	C	1
6	NUM	N	3
7	NOM AXE	C	70
8	CARDINAL	N	2
9	S1	N	3
10	S2	N	3
11	S3	N	3
12	S4	N	3
13	S5	N	3
14	S6	N	3
15	S7	N	3
16	S8	N	3
17	S9	N	3
18	S10	N	3
19	S11	N	3
20	S12	N	3
21	S13	N	3
22	S14	N	3
23	S15	N	3
24	S16	N	3
25	S17	N	3
26	S18	N	3
27	S19	N	3
28	S20	N	3
29	S21	N	3
30	S22	N	3
31	S23	N	3
32	S24	N	3
33	S25	N	3
34	S26	N	3
35	S27	N	3
36	S28	N	3
37	S29	N	3
38	S30	N	3
39	S31	N	3
40	S32	N	3
41	S33	N	3
42	S34	N	3
43	S35	N	3

ANNEXES

ANNEXE 4

PROGRAMME CRES.PRG DE CREATION DU FICHIER ES.DBF DES ENTITES  
SYMPTOMATIQUES A PARTIR DU FICHIER SYMPTOME.DBF.  
ECRIT EN LANGAGE DE COMMANDE DBASE III.

```
set exact on
set talk off
USE SYMPTOME
INDEX ON STR(CHAPITRE,2) + SUBSTR(ESMERE,1,30) +SUBSTR(ENTITSYMPT,1,10) TO
INDEXSYMP
use ES
copy structure to TRAVAIL
select A
use SYMPTOME index INDEXSYMP

select B
use TRAVAIL
store "STR (CHAPITRE, 2)+SUBSTR(ESMERE,1,30)+substr (ENTITE,1,10)" to C
index on &C to ESINDEX
set index to ESINDEX

select A
store 0 to NBES
store 0 to M
do while .not. EOF()
if TYPAXE (<)'**'
store trim (ESMERE) to AUXESM
store trim (ENTITSYMPT) to AUXES
store NUMERO to AUXNO
store NIVOBSER to AUXNIV
store CHAPITRE to AUXCHAP
store substr (AXE,1,7) to AUXNOM
store substr (TYPAXE,2,1) to AUXTYP
store '"' +str (CHAPITRE, 2)+substr (ESMERE, 1, 30) +
substr (ENTITSYMPT, 1, 10)+'"' to IND

select B
find &IND
if EOF().OR. BOF()
append blank

if AUXTYP = 'E'
store substr (AUXESM,1,45)+' --)' +substr (AUXES,1,25) to AUXNOM
endif

replace ESMERE with AUXESM, ENTITE with AUXES, NIVOBS with AUXNIV,;
CHAPITRE with AUXCHAP, NOMAXE with AUXNOM, TYPE with AUXTYP,;
CARDINAL with 1, S1 with AUXNO

store NBES + 1 to NBES
else
replace CARDINAL with CARDINAL + 1
if CARDINAL < 10
store trim ( str (CARDINAL,1)) to D
else
store trim (str (CARDINAL, 2)) to D
endif
```

ANNEXES

```
    if CARDINAL > M
      store CARDINAL to M
      store AUXNO to NOSYMAX
    endif

    replace S&D with AUXNO

  endif

  select A
endif
  skip
enddo

select B
use
erase ES.DBF
rename TRAVAIL.DBF to ES.DBF
use ES
replace all NUM with RECNO()
set exact off
set talk on
```

ANNEXES

ANNEXE 5

PROGRAMME CREAX.PRG DE CRÉATION DU FICHER AXES.DBF CONTENANT LES FICHES DES AXES, DU PREMIER AU 4EME NIVEAU D'OBSERVATION.

Pour chaque niveau, les es élémentaires précèdent les scores.

Structure de la fiche :

NUMAXE	
NIVOBS	<--- ES.DBF
CHAPITRE	<--- ES.DBF
TYPE	<--- ES.DBF
NOMAXE	<--- ES.DBF
CARDINAL	

```
set talk off
set safety off
set echo on
do CRES
select a
use
select b
use
select a
use AXES
? '          PROGRAMME CREAX'
?
? ' NETTOYAGE DE AXES.DBF'
zap
? ' CHARGEMENT A PARTIR DE ES.DBF'
  append from ES
? ' COPIE -> TRAVAIL , SUP. SCORES'
  copy to TRAVAIL for TYPE = 'S'
  delete for TYPE = 'S'
  PACK
  replace all NIVOBSAXE with NIVOBS
  use TRAVAIL
? ' TRI DE TRAVAIL SUR NOMAXE -> SCORES.DBF'
  sort on NOMAXE to SCORES

* compactage de SCORES
? ' COMPACTAGE DE SCORES.DBF -> SCORESCO.DBF'
select A
use SCORES
select B
use
erase SCORESCO
use AXES
copy structure to SCORESCO
use SCORESCO
store 0 to MAXS
select A
set echo off
do while .not. eof()
  * nouvelle serie d'un même score
  store NIVOBS to NIVO
  store CHAPITRE to AUXCHAP
  store TYPE to AUXTYP
  store NOMAXE to AUXNOM
```

## ANNEXES

```

store NUM to AUXNUM
store CARDINAL to AUXCO

store 1 to I
do while I <= CARDINAL
  if I < 10
    store str (I,1) to IC
  else
    store str (I,2) to IC
  endif
  store S&IC to AUXS&IC
  store I + 1 to I
enddo

select b
append blank

replace CARDINAL with AUXCO
store 1 to I
do while I <= AUXCO
  if I < 10
    store str (I,1) to IC
  else
    store str (I,2) to IC
  endif
  replace S&IC with AUXS&IC
  store I + 1 to I
enddo

select a
store 'U' to B
skip

do while .not.eof() .and. B='U'
  if NOMAXE = AUXNOM
    if NIVOBS > NIVO
      store NIVOBS to NIVO
    endif

    store CARDINAL to AUXCO
    store 1 TO I
    do while I <= CARDINAL
      if I < 10
        store str (I,1) to IC
      else
        store str (I,2) to IC
      endif
      store S&IC to AUXS&IC
      store I + 1 to I
    enddo

    select b
    store CARDINAL + 1 to J
    store 1 TO I
    do while J <= CARDINAL + AUXCO
      if J < 10
        store str (J,1) to JC
      else

```

ANNEXES

```

        store str (J,2) to JC
    endif

    if I < 10
        store str (I,1) to IC
    else
        store str (I,2) to IC
    endif

    replace S&JC with AUXS&IC
    store J + 1 to J
    store I + 1 to I
enddo
replace CARDINAL with CARDINAL + AUXCO

select a
skip
else
store 'F' to B
endif
enddo
select B
replace CHAPITRE with AUXCHAP, TYPE with AUXTYP, NIVOBSAXE with NIVO
replace NOMAXE with AUXNOM
if CARDINAL > MAXS
store CARDINAL to MAXS
endif

select A
enddo

select a
use
select b
use
select a
use AXES
append from SCORESCO

index on str(NIVOBSAXE, 1) + TYPE + NOMAXE to INDEXAX
set index to indexax
go top
store 1 TO I
do while .not.eof()
replace numaxe with I
store I + 1 TO I
skip
enddo
sort on NUMAXE to AXES1
use
erase AXES.DBF
rename AXES1.DBF to AXES.DBF

** CREATION DE LSTSCORE.DBF LE FICHER DES SCORES DONNANT LA LISTE DES ES
** DE CHAQUE SCORE

select A
use SCORES

```

## ANNEXES

```
select B
use LSTSCORE
zap
select A
do while .not. eof()
  * nouvelle serie d'un meme score
  store 1 to NIVO
  store CHAPITRE to AUXCHAP
  store TYPE to AUXTYP
  store NOMAXE to AUXNOM
  store 1 to CO
  store NUM to AUXNUM

  select b
  append blank
  replace E1 with AUXNUM

  select a
  store 'V' to B
  skip

  do while .not.eof() .and. B='V'
    if NOMAXE = AUXNOM
      store NUM to AUXNUM
      store CO + 1 to CO
      if NIVOBS > NIVO
        store NIVOBS to NIVO
      endif

      select b
      if CO < 10
        store trim ( str (CO,1)) to D
      else
        store trim (str (CO, 2)) to D
      endif
      replace E&D with AUXNUM

      select a
      skip
    else
      store 'F' to B
    endif
  enddo

  select B
  replace CHAPITRE with AUXCHAP, TYPE with AUXTYP, NIVOBSAXE with NIVO
  replace NOMAXE with AUXNOM, CARDINAL with CO
  select A
enddo

** set echo on
do MAJSYMP
do MAJSYMP1
do BILAN
set talk on
quit
```

ANNEXES

ANNEXE 6

```
(*****  
(*****          GESTION EN PASCAL DU FICHER FSYNDROM          *****  
(*****
```

{ procédures principales:

```
RECHCODE  
LITABCOD  
LIRLSTCODE  
LIRFSYND  
RECHSYNDANSLST  
ECRIADRSYMPHOME  
CHANGADRESYMPHOME  
LIRNOSYMP  
LIRNOSYND  
INITFSYND  
DERBLOC  
APPENDFSYND  
ADRESLSTSYND  
AFFORMNATUR  
RECHSEC  
RECHDANSTFSYND  
RECHERCH  
LIRESYNDAJOUTE  
AJFSYND  
MODFSYND  
SUPSYMPT  
SUPSYND  
SUPFSYND  
AFFSYND
```

}

program GFSYND;

const

```
MAXSYNDCAND = 150;  
  
ST1      : string[30] = 'CREATION DU FICHER "FSYNDROM";  
ST2      : string[30] = ' AJOUTER AU FICHER "FSYNDROM";  
ST2A     : string[38] = 'AJOUT DE DONNEES AU FICHER "FSYNDROM";  
  
MAXELEMFSYND= 8000 ;  
MAXELEMtsy= 254;  
NBMAXELEM = 7935;  
NBMAXSY   = 254;
```

type

```
CHAINE = string [80];  
COD = string [3];  
  
TYPTCOD = array [1..MAXSYNDCAND] of COD;  
  
TYPTFSYND = array[1..MAXELEMFSYND] of integer;
```

ANNEXES

```

TYPTSY      = array[1..MAXELEMENTSY] of integer;
ESYPTOME    = record
    CAR_DB3   : char;
    FORMNAT:  string[90]; NUMERO:  string[4];
    ENTITSYM: string[45]; CODENTSYM: string[4];
    VALNAT:   string[25]; VALNUM:  string[3];
    NIVOSAXE: string[ 4]; ATTRIBUT: string[46];
    PERE:     string[4];  FILS:    string[4];
    FRERE:    string[4];  NIVObs:  char;
    NBDESynd: string[3];  ADLTSynd: string[4];
    CHAPITRE : string [2]; ESMERE:  string[82];
    AXE       : string[57]; NUMAXE  : string [3];
    TYPAXE    : string[2]
end;
( STRUCTURE DE DONNEES PASCAL CORRESPONDANTE A LA STRUCTURE DE
  DONNEES DU FICHER DBASEII SYMPTOME

```

```

ST70        = string[70];
ENSREP      = set of char;
var
CELLULE     : integer;
ESY         : ESYMPTOME;
FSPT        : file of ESYMPTOME; ( fic. interne PASCAL pour
                                  fic. DBASEII "'C:SYMPTOME.DBF'"
TFSYND      : TYPTFSYND;
STI         : string[10];
MESSAGECOH : boolean;
TFSYNDEXIST: boolean;
MAXTFSYND   : integer;( = ESTDERNIER(=dernier élément) du tableau
                                  TFSYND
A1, A2, B1, B2, C1, C2, SUP, I, J : integer;
REP, CC     : char;
F           : TEXT;
NUM         : byte;
T           : TYPTCOD;
C           : COD;
FAIT,BASCULE: boolean;

```

```

-----
{*****}
{**** PARTIE " OUTILS DE BASE " ****}
{*****}

```

```

procedure VIDECRAN;
begin
  write(clrhom);
end ( VIDECRAN );

procedure AVANCER;
var C : char;
begin
  writeln; writeln('          TAPEZ SUR LA TOUCHE <RC> POUR CONTINUER');
  readln(C);
end;
function CONTINUER : boolean;
var CC : char;

```

## ANNEXES

```

begin
  write(rvson,' VOULEZ VOUS CONTINUER (O/N) ?',rvsoff);
  write(' : ');
  read(CC);
  CONTINUER := (CC in ['O','o']);
  write(dellin);
end;

procedure TITRE ( S:ST70 );
var A, B : integer;
begin
  A:=len(S); B:=(70-A) div 2;
  write(' ':5);
  write(rvson); write(S:(A+B)); write(' ':B); write(rvsoff);
  writeln; writeln;
end;

procedure INITSY ( var T:TYPTSY );
var A : integer;
begin
  for A:=1 TO MAXELEMTSY do T[A]:=0;
end;

procedure LIRENTIER(var E:integer);
const
  MAR : string[10] = '          ';
  DE  : string[30] = 'ERREUR: Tapez un NOMBRE s.v.p.';
var
  CA      : string[4];
  ARRET, INVERSEVIDEO, BIS : boolean;
  X, Q    : integer;

begin
  E:=0; INVERSEVIDEO:=false; BIS:=false;
  repeat
    ARRET:=true; CA:=''; buflen:=4; Q:=0;
    read(CA);
    val(CA,E,Q);
    if (Q<>0) and (not (CA[1] = ' ')) then
      begin
        if not INVERSEVIDEO
          then
            begin
              writeln; write(MAR); write(rvson); write(DE); write(rvsoff);
              write(' : '); INVERSEVIDEO:=true;
            end
          else
            begin
              write(' '); write(rvson); write('??'); write(rvsoff); write(' : ');
              BIS:=true;
            end;
            ARRET:=false;
          end
        else if CA[1] = ' ' then E := 0;

      if INVERSEVIDEO and ARRET then
        ( PERMET D'ECRIRE LE RES. AU BON ENDROIT )

```

ANNEXES

```

begin
  write(dellin);
  if not BIS
  then
    begin
      if len(CA)=1
      then X:=7
      else if len(CA)=2
      then X:=6
      else if len(CA)=3
      then X:=5
      else X:=4;
      write('-> ',X,E);
    end
  else write('-> ',E);
end;
until ARRET;
writeln;
end; ( LIRENTIER )

procedure LIREQSTI ( var QSTI : char );
var CH : set of char;
begin
  CH:=['I','Q','S','T','i','q','s','t'];
  write('fréquence QSTI, s.v.p. (Q/S/T/I) : ':37); readln(QSTI);
  while not (QSTI in CH) do
    begin
      write('fréquence QSTI, s.v.p. (Q/S/T/I) : ':37);
      readln (QSTI)
    end
  end;

function CODENUM (NUM : integer): integer;
begin
  CODENUM := - (NUM + 256)
end;

function DECODENUM (CEL : integer): integer;
begin
  DECODENUM := - (CEL + 256)
end;

function CODENOMBRE (NB : integer): integer;
begin
  if NB = 0 then CODENOMBRE := - 255 else CODENOMBRE := - NB
end;

function DECODENOMBRE (CEL: integer): integer;
begin
  if CEL = - 255 then DECODENOMBRE := 0 else DECODENOMBRE := - CEL
end;

function CODESY (N : integer; QSTI : char): integer;
var Q : byte;
begin
  case QSTI of
    'Q' : Q := 0;
    'S' : Q := 1;
  end
end;

```

## ANNEXES

```

    'T' : Q := 2;
    'I' : Q := 3;
  end;
  CODESY := N * 4 + Q;
end;

procedure DECODESY (CEL : integer; var N : integer; var QSTI : char);
const T : string [4] = 'QSTI';
begin
  N := CEL DIV 4;
  QSTI := T [ CEL MOD 4 + 1 ];
end;

procedure CLOCHES;
begin
  write ( ^G^G^G )
end;

PROCEDURE LIREP ( X, Y:integer; E:ENSREP; var R:char );
begin
  gotoxy (X,Y); read (TRM,R);
  if R = chr (13) then R := ' ';
  while not (R in E) do
    begin CLOCHES; gotoxy (X,Y); writeln(' '); gotoxy(X,Y); read (TRM,R)
      if R = chr (13) then R := ' ';
    end
  end (* LIREP *);

PROCEDURE ECRIRE ( X,Y:integer; S : CHAINE);
begin
  gotoxy(X,Y); write (S);
end (* ECRIRE *);

procedure ECRIRINV ( X,Y : integer; S : CHAINE);
begin
  writeln (rvson); gotoxy (X,Y); write (S); writeln (rvsoff)
end ;

{=====
      {*****
      {**** SOUS-PROGRAMME " GESTION DU FICHER "FSYNDROME" " ****
procedure RECHCOD ( C:COD; T:TYPTCOD; SUP : byte; var TROUVE :boolean;
      var NUM : integer);
var I : integer;
begin
  I := 1;
  while (T[I] (> C) and (I <= SUP) do I := I + 1;
  TROUVE := I <= SUP;
  if TROUVE then NUM := I else NUM := 0
end ( RECHCOD );

procedure LITABCOD ( var SUP : integer);
VAR I : integer;
begin
  assign (F, 'LSTCODES.TXT'); reset (F);

```

ANNEXES

```

I := 1;
while not EOF(F) do
begin
  readln(F, TCII ); I:= I+1
end;
SUP := I-1;
end;

procedure LIRLSTCODES ( var FAIT, BASCULE : boolean );
var I : integer;
    R : char;
begin
  FAIT := true;
  writeln;
  write ('désirez-vous entrer les syndromes par c(ode ou par n(uméro ?:');
  read (TRM, R);writeln;
  BASCULE := R in ['c','C'];
end { LIRLSTCODES };

procedure LIRFSYND;
{ =lire fic. FSYNDRON, le placer dans le tableau TFSYND, le sauvegarder
dans fic. FSYN.MEM si besoin, mettre à jour la var. globale MAXTFSYND qui
est l'indice du dernier enregistrement contenu dans TFSYND }

var K, I : integer;
    FY, F2 : file;
begin
  assign ( FY,'FSYNDRON' ); reset ( FY, 2 );
  blockread ( FY, TFSYND, MAXELEMFSYND );
  if TFSYND[1] <> 0 then
  begin
    assign ( F2,'FSYN.MEM' ); rewrite ( F2,3 );
    blockwrite ( F2,TFSYND,MAXELEMFSYND );
    close ( F2 )
  end;
  close ( FY );

  TFSYNDEXIST := true;

  while (TFSYND[1] <> 0) and (I < MAXELEMFSYND) do I := I + 1;
  MAXTFSYND := I - 1

end { LIRFSYND };

procedure RECHSYNDANSLIST(INF,SUP,NOSYNDR:integer; var TROU:boolean;
var ADRESYNDR:integer);
var W : integer;
    CX : char;
begin
  TROU := false; ADRESYNDR := 0;
  if not (INF > SUP) then
    { rech. séquentielle du syndrome associé d'un symptôme donné }
    repeat
      DECODESY ( TFSYND[INF],W,CX ); INF := INF + 1;
    until (INF > SUP) or (W = NOSYNDR);

  if (W = NOSYNDR) then

```

## ANNEXES

```

begin
  TROU := true;
  ADRESYNDR := (INF - 1)
end;
end;

procedure ECRIDRDSYMPOTOME(NOSYMPT, E, NB : integer);
{ écriture de E ( =adresse de la liste de syndromes associés ) dans
  ADLSTSYND et de NB ( =nombre de syndromes ) dans NBDESYND de
  l'enregistrement numéro NOSYMPT de SYMPTOME.DBF }

var   CE : string[4];

begin
  seek ( FSPT , NOSYMPT+2 ); read ( FSPT,ESY );
  str ( E:4,CE ); ESY.ADLSTSYND := CE;
  str( NB:3,CE ); ESY.NBDESYND:=CE;
  seek ( FSPT,NOSYMPT+2 ); write( FSPT,ESY );
end;

procedure CHANGADRDSYMPOTOME (NOSYMPT, NBSAJOUTE, NOUVNBS : integer);
{ écrit dans l'enregistrement de SYMPTOME no. NOSYMPT le nouveau nombre de
  symptomes associés, puis change les adresses ADLSTSYND dans les
  enregistrements suivants }

var   CE      : string[4];
      NODERNSYMPT, I, A, P : integer;

begin
  NODERNSYMPT := length ( FSPT ) - 3;
  write('      modification dans le fichier SYMPTOME.DBF en cours...');

  seek ( FSPT,(NOSYMPT + 2) ); read ( FSPT,ESY );
  str ( NOUVNBS:3,CE ); ESY.NBDESYND := CE;
  seek ( FSPT, NOSYMPT + 2 ); write ( FSPT,ESY );
  NOSYMPT := NOSYMPT + 1;
  repeat
    seek ( FSPT, NOSYMPT + 2); read ( FSPT,ESY );
    CE := ESY.ADLSTSYND;
    I := 1;
    while CE[I] = ' ' do I := I + 1;
    val ( copy ( CE,I,len(CE) - (I - 1) ) ,A ,P );

    A := A + NBSAJOUTE;
    str ( A:4,CE ); ESY.ADLSTSYND := CE;
    seek ( FSPT, NOSYMPT + 2 ); write ( FSPT,ESY );
    NOSYMPT := NOSYMPT + 1;
  until NOSYMPT = NODERNSYMPT;
  write (dellin); writeln;
end { CHANGADRDSYMPOTOME };

procedure LIRNOSYMP ( var CELL: integer; var NO: integer);
var EC: integer;
begin
  writeln; write(' ':10);
  write(rvson);
  write('numéro du symptôme, s.v.p.',': ':15);
  LIRENTIER ( EC );

```

ANNEXES

```

write(rvsoff);
while (EC > NBMAXELEM) do
begin
write('nombre entre 1 et 7700, s.v.p.':40,' ':':11); LIRENTIER ( EC )
end;
CELL := CODENUM ( EC ); NO := EC
end;

procedure LIRNOSYND ( N:integer; var NOSYNDR: integer);
var EC : integer;
QSTI : char;
TROUVE : boolean;
begin
if not FAIT then LIRLSTCODES ( FAIT, BASCULE);
if BASCULE
then begin
TROUVE := false;
repeat
write ('code du syndrome ':29, N:3, ' associé, s.v.p. : ':19);
readln( C );
if len (C) > 0 then RECHCOD (C, T, SUP, TROUVE, NOSYNDR);
if ( not TROUVE ) and ( len ( C ) > 0 )
then writeln ( ' Erreur dans le code entré !' );
until TROUVE or (len (C) = 0);
if len (C) = 0 then NOSYNDR := 0
end
else begin
write('numéro du syndrome ':29,N:3, ' associé, s.v.p. : ':19);
LIRENTIER ( EC );
while (EC > 3999) do
begin
write('nombre entre 1 et 3999, s.v.p.':40,' ':':11);
LIRENTIER ( EC )
end;
NOSYNDR := EC
end
end { LIRNOSYND };

procedure LIRCELLSYNDR ( NOSYNDR:integer; var CELL: integer );
var QSTI : char;
begin
if (NOSYNDR > 0) then
begin
LIREQSTI ( QSTI );
CELL := CODESY ( NOSYNDR, QSTI );
end;
end { LIRCELLSYNDR };

procedure INITFSYND;

var FY : file;
A : integer;
begin
writeln('Veuillez patienter quelques instants s''il vous pla

```

## ANNEXES

```

t');
  for A := 1 TO MAXELEMFSYND do TFSYND[A] := 0;
  assign ( FY, 'FSYNDROM' ); rewrite( FY, 2 );
  blockwrite(FY, TFSYND, MAXELEMFSYND);
  close(FY);
  TFSYNDEXIST := true; MAXTFSYND := 0
end ( INITFSYND );

procedure DERBLOC ( var DERSYMP : INTEGER );
var K : integer;
begin
  { Recherche de DERSYMP le dernier symptôme }
  K := MAXTFSYND; DERSYMP := 0;
  if (K > 0) then
    begin
      while (TFSYND[K] < -8191) or (TFSYND[K] > -257) do K := K - 1;
      DERSYMP := DECODENUM ( TFSYND[K] )
    end;
end;

procedure APPENDFSYND ( S : ST70);

const  VU = 22;
var    FY           : file;
       CONT, STOP   : boolean;
       J, K, NB, ADRES, kB, NA, I : integer;
       TROU         : boolean;
       ADRESYNDR    : integer;
       NOSYMPT, NOSYNDR : integer;
       DERNSYMPTECRI : integer; { Numéro du DERNIER symptôme entré }

begin
  writeln('                initialisation en cours...');
  writeln;

  { si besoin, partie lire fic. FSYNDROM, sauvegarder dans fic. FSYN.MEM }
  if (not TFSYNDEXIST) then LIRFSYND;
  { Le tableau TFSYND corresp. à FSYNDROM est initialisé }

  DERBLOC (DERNSYMPTECRI);

  K := MAXTFSYND;
  { partie création du fichier ou ajout à ce fichier }
  NB := 1; NA := 0; CONT := true;

  while (CONT) do
    begin
      VIDECRAN; TITRE ( S ); STOP := false;
      I := 2;
      if (NB = 1) and (K <= (MAXELEMFSYND - 4)) then
        begin
          writeln(' ':10,rvson,'LE DERNIER NO. DE SYMPTOME ENTRE EST : ',
                DERNSYMPTECRI,' ',rvsoff); I := I + 1;
          LIRNOSYMP ( CELLULE, NOSYMPT ); I := I + 2;
          if (NOSYMPT < 1) then
            begin STOP := true; cont := false end;

          if (not STOP) and (NOSYMPT > DERNSYMPTECRI) { Contrôle de cohérence

```

## ANNEXES

```

then
  begin
    if ( NOSYMPT > (DERNSYMPTECRI + 1) ) then
      { Création des symptomes 'manquants' }
      repeat
        DERNSYMPTECRI := DERNSYMPTECRI + 1;
        K := K + 1; TFSYNDCKJ := CODENUM ( DERNSYMPTECRI );
        ADRES := K;
        TFSYND [ K + 1 ] := CODENOMBRE ( 0 );
        TFSYND [ K + 2 ] := A1;
        TFSYND [ K + 3 ] := A2;
        TFSYND [ K + 4 ] := B1;
        TFSYND [ K + 5 ] := B2;
        TFSYND [ K + 6 ] := C1;
        TFSYND [ K + 7 ] := C2;
        K := K + 1 + 6;
        ECRIADRSYMPHOME ( DERNSYMPTECRI,ADRES,0 );
      until DERNSYMPTECRI = (NOSYMPT - 1);

      DERNSYMPTECRI := NOSYMPT;
      K := K + 1; TFSYNDCKJ := CELLULE; ADRES := K;
      { ADRES : adresse du premier élé. de la liste }
      TFSYND [ K + 1 ] := CODENOMBRE ( 0 );
      TFSYND [ K + 2 ] := A1;
      TFSYND [ K + 3 ] := A2;
      TFSYND [ K + 4 ] := B1;
      TFSYND [ K + 5 ] := B2;
      TFSYND [ K + 6 ] := C1;
      TFSYND [ K + 7 ] := C2;
      K := K + 1 + 6;
      NA := K + 1; { NA : adresse du premier syndrome de la liste
                    pour un symptome donné }
    end
  else
    if (not STOP) then
      begin
        writeln;
        writeln(' Ce numéro est inférieur/égal à celui du symptome
          ' précédent : IMPOSSIBILITE');
        AVANCER; STOP := true;
      end { if (not STOP)...else ... };
    end { if (NB = 1) ... };

    if (NB > 1) and ( K <= (MAXELEMFSYND-2) ) then
      begin
        writeln;
        writeln(' ':10,rvson,'numéro du symptome',': ':23,NOSYMPT,rvsoff);
        writeln(' ':10,rvson,'nombre de syndromes associés entrés',
          ': ':6,(NB - 1),rvsoff);
        writeln(' ':10,'numéro du dernier syndrome entré',': ':9,NOSYNDR);
        I := I + 3;
      end;

    while (I <=VU) and (NB <255) and (K < (MAXELEMFSYND-1)) and (not STOP)
    begin
      LIRNOSYND ( NB,NOSYNDR ); I := I + 1;

      if (NOSYNDR > 0)

```

## ANNEXES

```

then
  begin
    RECHSYNDANSLIST ( NA,K,NOSYNDR,TROV,ADRESYNDR );

    if (not TROV)
      then
        begin
          LIRCELLSYNDR ( NOSYNDR, CELLULE ); I := I + 1;
          NB := NB + 1;
          J := K;
          while (TFSYND [J] > CELLULE) and (J >= ADRES + 8 ) do
            begin
              TFSYND [J + 1] := TFSYND [J]; J := J - 1
            end;
          TFSYND [J + 1] := CELLULE;
          K := K + 1 ;
        end
      else
        begin
          writeln(' Erreur, numéro de syndrome déjà entré');
          I := I + 1
        end
      end
    else
      begin
        STOP := true; NB := NB - 1;
        if (NB > 0) then TFSYND[ADRES + 1] := CODENOMBRE ( NB );
        ECRADRDSYMPHOME ( NOSYMPT,ADRES,NB )
      end
    end { while ... };

    if (STOP = true) then begin NB := 1; NA := 0 end;

    if (NB = 255) then
      begin
        writeln(' ':15,rvson,'NOMBRE DE SYNDROMES ENTRES > A 254',rvsoff);
        TFSYND[ADRES + 1] := CODENOMBRE ( NB - 1 );
        ECRADRDSYMPHOME ( NOSYMPT,ADRES,(NB - 1) ); NB := 1; NA := 0
      end;

    if (K >= (MAXELEMFSYND - 4)) then
      begin
        if (NB > 1) and (not STOP) then
          begin
            TFSYND[ADRES + 1] := NB - 1;
            ECRADRDSYMPHOME ( NOSYMPT,ADRES,(NB - 1) )
          end;
        write(rvson);
        writeln(' VOUS VENEZ D'ATTEINDRE LES LIMITES DU FICHER FSYNDROM.')
        writeln(' LE SYSTEME S'ARRETE : CF. DOCUMENTATION DU LOGICIEL');
        write(rvsoff);
        CONT := false;
        AVANCER;
      end;

    if (CONT) and (NB = 1) then CONT := CONTINUER;

end;

```

ANNEXES

```

MAXTFSYND := K;
  ( fin de la partie creation ou "ajout" du/ou fichier FSYNDROM )

  ( partie ecrire FSYNDROM )
assign( FY, 'FSYNDROM'); rewrite( FY, 2);
  blockwrite (FY, TFSYND, MAXELEMFSYND);
  close(FY);
end ( APPENDFSYND );

procedure ADRESLSTSYND (NOSYMPT:integer;var TROV:boolean;var ADRES:integer)
{ recherche de l'adresse de la liste associee au numero de symptome NOSYMP
  dans le fichier SYMPTOME.DBF }
var   CI      : string[4];
      ADRLSTDERNSYMPT, NODERNSYMPT, A, I, P : integer;

begin
  NODERNSYMPT := length ( FSPT ) - 3;

  if (NOSYMPT <= NODERNSYMPT) and (NOSYMPT > 0) then
    begin
      seek ( FSPT, NOSYMPT+2 ); { Acces direct au fichier DBASE III }
      read ( FSPT, ESY ); CI:=ESY.ADLSTSYND; { Lecture en record PASCAL
                                             du fichier DBASE III }

      I:=1;
      while (CI[I]=' ') do I:=I+1;
      val ( copy ( CI,I,(len (CI)-(I-1)) ),A,P);
      ADRES:=A; TROV:=true;
    end
  else TROV := false;
end ( ADRESLSTSYND );

procedure SAUTBLANC ( s : CHAINE; var I : byte );
begin
  I := 1;
  while S[I] = ' ' do I := I + 1
end;

procedure AFFORMNATUR ( NOSYMPT : integer);
{ affiche les renseignements de longueur fixe relatifs au symptome }
var CI : string [4];
    V, I, P : integer;
    B : byte;
begin
  write ( rvson, ' FORME NATURELLE      : ', rvsoff);
  SAUTBLANC ( ESY.FORMNAT, B );
  writeln ( copy ( ESY.FORMNAT, B, 55 ));
  write ( rvson, ' ENTITE SYMPTOMATIQUE : ', rvsoff);
  SAUTBLANC ( ESY.ENTITSYM,B );
  writeln ( copy ( ESY.ENTITSYM,B, 50 ));
  write ( rvson, ' VALEUR NATURELLE     : ', rvsoff);
  SAUTBLANC ( ESY.VALNAT, B );
  writeln ( copy ( ESY.VALNAT, B, 30 ));
  write ( rvson, ' VALEUR NUMERIQUE     : ', rvsoff);
  CI := ESY.VALNUM;
  I:=1;
  while (CI[I]=' ') do I:=I+1;
  val ( copy ( CI,I,(len (CI)-(I-1)) ), V, P);

```

## ANNEXES

```

writeln ( V );
write ( rvson, ' E.S. MERE : ', rvsoff);
SAUTBLANC ( ESY.ESMERE, B );
writeln ( copy ( ESY.ESMERE, B, 55 ));

end ( AFFORMNATUR );

procedure RECHSEC ( X : integer; var TROU : boolean; var ADRES : integer);
var I, NUMSYMP : integer;
    EXISTESUIV : boolean;

begin
    TROU := false;
    EXISTESUIV := true;
    I := 1;
    while (not TROU) and (EXISTESUIV) do
        begin
            NUMSYMP := DECODENUM ( TFSYND [I] );
            TROU := NUMSYMP = X;
            ADRES := I;
            I := I + 1 + 6 + DECODENOMBRE ( TFSYND [ADRES + 1] ) + 1;
            EXISTESUIV := TFSYND [I] <> 0
        end
    end ( RECHSEC );

procedure RECHDANSTFSYND ( NOSYMPT : integer; var TROUVE : boolean;
    var ADRES : integer );

var
    NBELEMDANS2FICEGAUX : boolean;
    P, NODERNSYMPT, ADRLSTDERNSYMPT, MAXTFSYNDBIS : integer;
    CI : string [4];
( adresse de ESTDERNIER dans TFSYND indiqué par le fichier "SYMPTOME.DBF )
begin
    NODERNSYMPT := length ( FSPT ) - 3;
    seek(FSPT, NODERNSYMPT + 2 ); read(FSPT,ESY); CI:=ESY.ADLSTSYND;
    I:=1;
    while (CI[I]= ' ') do I:=I+1;
    val ( copy ( CI,I,(len (CI)-(I-1)) ),ADRLSTDERNSYMPT,P);
    NBELEMDANS2FICEGAUX := true;

    RECHSEC ( NOSYMPT, TROUVE, ADRES );

    write ( dellin ); writeln;      ( Effaçage de 'RECHERCHE EN COURS ...' )

    if not TROUVE then              ( Cas résultat négatif )
        begin
            writeln('          DESOLE, CE SYMPTOME N'EXISTE PAS !'); writeln;
        end;

        ( Vérification de l'état de cohérence entre les 2 fic. SYMPTOME et
                                                FSYNDROM )
    if ( ADRLSTDERNSYMPT <> 0 )
        then
            MAXTFSYNDBIS := ADRLSTDERNSYMPT + 1 + 6 +
                DECODENOMBRE( TFSYND[ADRLSTDERNSYMPT+1] )
        else MAXTFSYNDBIS := 0;      ( MAXTFSYNDBIS = nbre théorique d'éléments dans

```

## ANNEXES

```

                                FSYNDROM calculé à partir de SYMPTOME
NBELEMDANS2FICEGAUX := MAXTFSYND = MAXTFSYNDBIS;
                                { MAXTFSYND = nbre d'éléments dans FSYNDROM

{ Affichage de la vérification si cela est nécessaire et n'a pas été
                                                                    fait
if not ( MESSAGECOH or NBELEMDANS2FICEGAUX ) then
begin
  for I := 1 TO 2 do writeln;
  write(' ':2, rvson);
  write('LES FICHIERS "SYMPTOME" ET "FSYNDROM" N'ONT PAS LE MEME',
        ' NOMBRE D'ELEMENTS', rvsoff);
  writeln; writeln;
  writeln('=) Manque de cohérence à corriger, s.v.p. (<=:57);
  AVANCER;
  MESSAGECOH := true; { Ceci établit que le message a été affiché
                                                                    une fois
end;
end { RECHDANSTFSYND };

procedure RECHERCH ( var NOSYMPT:integer; var TROUVE:boolean;
                    var ADRES :integer);
var   COHERENT : boolean;
      I       : integer;
      ADRLSTDERNIERSYMPTOME : integer;
begin
  ADRES := 0; TROUVE := true; COHERENT := true;

  writeln; writeln;
  write(' ':12,'numéro du symptôme concerné, s.v.p. : ');
  LIRENTIER ( NOSYMPT );

  if (NOSYMPT > 0)
  then
    begin
      write(' ':5,'RECHERCHE EN COURS ...');
      { Recherche dans le fichier SYMPTOME }
      ADRESLSTSYND ( NOSYMPT,TROUVE,ADRES );

      { Partie:lire fic. FSYNDROM, sauvegarder dans fic. FSYN.MEM si besoi
      if not TFSYNDEXIST then LIRFSYND;
      { Le tableau TFSYND corresp. à FSYNDROM est initialisé }

      { Réussite de la recherche dans le fic.SYMPTOME :vérif. de la
                                                                    cohérence
      if (TROUVE) then
        if ( DECODENUM ( TFSYND[ADRES] ) (<) NOSYMPT )
        then COHERENT:=false
        else
          { Cas 2 fic. SYMPTOME et FSYNDROM cohérent
          begin
            write(dellin); writeln { Effaçage de 'RECHERCHE EN COURS ...'
            end;

          { Cas echec de la recherche dans le fic. SYMPTOME ou incohérence :
                                                                    recherche dans le fic. FSYNDROM
      if (not TROUVE) or (not COHERENT) then
        RECHDANSTFSYND ( NOSYMPT,TROUVE,ADRES );

```

## ANNEXES

```

if not COHERENT then      { Affiche un message d'incohérence et l'efface
  begin
    write(rvson);
    write('ATTENTION, INCOHERENCE ENTRE LES FIC. SYMPTOME ET FSYNDROM'
          ' POUR LE NUMERO : ',NOSYMPT);
    write(rvsoff);
    for I := 1 to 20000 do for J := 1 to 9 do;
    write ( dellin ); writeln
  end
end
else TROUVE := false
end { RECHERCH };

procedure LIRESYNDAJOUTE ( NOSYMPT, NBSYNEXIS, ADRES:integer;
                          var TABUFF:TYPTSY; var NBSAJOUTE:integer);

const
  VU      = 22;

var
  CELLULE      : integer;
  I, H, NOSYNDR, ADRESYNDR : integer;
  STOP         : boolean;
  TROV         : boolean;

begin
  { Initialisation du tableau TABUFF : buffer des syndromes ajoutés }
  INITSY ( TABUFF );
  STOP := false; H := NBSYNEXIS; NBSAJOUTE := 1;

  while (not STOP) do
  begin
    VIDECRAN; TITRE ( ST2 ); I := 2;

    writeln(' ':10,rvson,'numéro du symptôme concerné',': ':14,NOSYMPT:4,' '
            rvsoff);
    writeln(' ':10,rvson,'nombre de syndromes associés existant : ',
            NBSYNEXIS:4,' ',rvsoff);
    writeln; I := I+3;

    STOP := H >= 254;
    while (not STOP) and (I <= VU) and (H < 255) do
    begin
      TROV := false;
      LIRNOSYND ( NBSAJOUTE,NOSYNDR ); I := I + 1;
      if NOSYNDR < 1 then begin STOP := true;
                          NBSAJOUTE := NBSAJOUTE - 1
                        end
      else
      begin
        if (NBSYNEXIS > 0) then
          RECHSYNDANSLIST ( ADRES + 8 , ADRES + 7 + NBSYNEXIS, NOSYNDR,TROV
                          ADRESYNDR );

        if not TROV
        then begin LIRCELLSYNDR ( NOSYNDR, CELLULE ); I := I + 1;
                  TABUFF[NBSAJOUTE] := CELLULE;
                  NBSAJOUTE := NBSAJOUTE + 1; H := H + 1
                end
        else begin
          writeln(' ':3,'Erreur, numéro de syndrome déjà existant');
        end
      end
    end
  end
end

```

ANNEXES

```

                I := I + 1
            end
        end
    end
end;

end;

if (H >= 254) then
    begin
        if (H > 254) then NBSAJOUTE := NBSAJOUTE - 1;
        for I:=1 TO 5 do writeln;
        writeln(' ':33,rvson,' ERREUR ! ',rvsoff); writeln;
        writeln(' ':5,rvson,'LE NOMBRE TOTAL DE SYNDROMES ASSOCIES EST >= 254
            rvsoff);
        writeln
        end
    end { LIRESYNDAJOUTE };

procedure AJFSYND;
var
    FY           : file;
    TBUF         : TYPTSY;
    CONT         : boolean;
    TROUVE       : boolean;
    NBSAJOUTE    : integer; { nombre d'enregistrements introduits }
    NOUVNBS, NBSDEJAECRI : integer;
    ADRES, ADRESUIVAN : integer;
    NOSYMPT     : integer;
    K, I, J     : integer;

begin
    CONT := true; ADRESUIVAN := 0; ADRES := 0;

    while CONT do
        begin
            VIDECRAN; TITRE(ST2); TROUVE := false;
            RECHERCH ( NOSYMPT, TROUVE, ADRES );
            if TROUVE then
                begin
                    NBSDEJAECRI := DECODENOMBRE ( TFSYND[ADRES + 1] );
                    LIRESYNDAJOUTE ( NOSYMPT, NBSDEJAECRI, ADRES, TBUF, NBSAJOUTE );
                    if (NBSAJOUTE > 0) then
                        begin
                            ADRESUIVAN := ADRES + 1 + 6 + NBSDEJAECRI + 1;

                            { decalage dans le tabl. TFSYND pour insertion }
                            for I := MAXTFSYND downto ADRESUIVAN do
                                TFSYND[I + NBSAJOUTE] := TFSYND[I];
                            MAXTFSYND := MAXTFSYND + NBSAJOUTE;

                            { insertion des composants de "FSYNDROM" dans TFSYND }
                            J:= ADRESUIVAN - 1;
                            for I := 1 to NBSAJOUTE do
                                begin
                                    K := J;
                                    while (TFSYND [K] > TBUF [I]) and (K >= ADRES + 8) do
                                        begin
                                            TFSYND [K+1] := TFSYND [K]; K := K - 1

```

## ANNEXES

```

        end;
        TFSYND [K+1] := TBUF [I]; J := J+1
    end;

    NOUVNBS := NBSDEJAECRI + NBSAJOUTE;
    TFSYND[ADRES + 1] := CODENOMBRE ( NOUVNBS );
    CHANGADRDSYMPOTOME ( NOSYMPT,NBSAJOUTE,NOUVNBS );
    writeln; writeln('                AJOUT EFFECTUE');
end;
end;

write(' ':10,rvson,'Ajouter dans un autre symptôme (O/N) ?',rvsoff,' : ');
read (TRM, CC);
CONT := CC in ['0','o'];
end;

assign (FY, 'FSYNDROM'); rewrite( FY, 2);
blockwrite(FY,TFSYND,MAXELEMFSYND);
close(FY);
end { AJFSYND };

procedure MODFSYND;
{ Modification dans la liste associée à un numéro de symptôme NOSYMPT acquis
  et recherché par la procédure RECHERCH }
const
    ST3 : string[30] = 'MODIFIER LE FICHER "FSYNDROM"';
var
    A, B, C : string [4];
    FY      : file;
    CELLULE : integer;
    NOTOK, CONT, STOP, TROUVE, TROV1,TROV2, TROV3 : boolean;
    CC, QSTI : char;
    NOSYMPT, NOSYNDR, NOUVSYNDR, AN, BN, CN, INF, SUP, NBS,
    P, ADRESSE, ADRES, ADRESYNDR1, ADRESYNDR2 :integer;
begin
    CONT := true;
    while CONT do
        begin
            VIDECRAN; TITRE ( ST3 );
            STOP := true;
            RECHERCH ( NOSYMPT,TROUVE,ADRES );
            { Cas réussite de la recherche dans SYMPTOME ou FSYNDROM }
            if (TROUVE) then begin NBS := DECODENOMBRE ( TFSYND[ADRES + 1] );
                                STOP := false
                            end;
            { Test si entrée des données par code ou par numéro }
            if (not STOP) and (not FAIT) then LIRLSTCODES ( FAIT, BASCULE);
            writeln;
            if not STOP
            then begin AFFORMNATUR ( NOSYMPT );
                    writeln ( ' ':12, rvson, 'nombre de syndromes associés : ',NBS, rvsoff
                            end;
            while not STOP do
                begin
                    TROV1 := false; TROV2 := true; TROV3 := false;
                    if BASCULE
                    then
                        begin

```

## ANNEXES

```

repeat
  write(' ':12,'code du syndrome associé à corriger, s.v.p. : ');
  readln( C );
  if len( C ) > 0 then RECHCOD ( C, T, SUP, TROU3, NOSYNDR );
  if ( not TROU3 ) and ( len ( C ) > 0 )
    then writeln( ' Erreur dans le code entré !' );
until ( TROU3 ) or ( len ( C ) = 0 );
if len ( C ) = 0 then NOSYNDR := 0
end
else
begin
  write(' ':12,'numéro du syndrome associé à corriger, s.v.p. : ');
  LIRENTIER ( NOSYNDR )
end;
if NOSYNDR > 0
then
begin
  { recherche du syndrome associé }
  RECHSYNDANSLIST ( ADRES + 8, ADRES + 7 + NBS, NOSYNDR, TROU1,
    ADRESYNDR1 );

  { trouvé }
  if (TROU1)
  then
  begin
    if BASCULE
    then
    begin
      repeat
        write(' ':12,'nouveau code, s.v.p. : ');
        readln( C );
        if len( C ) > 0 then RECHCOD ( C, T, SUP, TROU3, NOUVSYNDR );
        if not TROU3 then writeln( ' Erreur dans le code',
          'entré !' );
        until ( TROU3 ) or ( len ( C ) = 0 );
        if len ( C ) = 0 then NOUVSYNDR := 0
        end
      else
      begin
        write(' ':12,'nouveau numéro, s.v.p.',': ':26);
        LIRENTIER ( NOUVSYNDR )
        end;
      if (NOUVSYNDR > 0)
      then
      begin
        { Vérif. que NOUVSYNDR n'existe pas déjà }
        RECHSYNDANSLIST ( ADRES + 8, ADRES + 7 + NBS, NOUVSYNDR,
          TROU2, ADRESYNDR2 );

        if TROU2 then
          writeln(' ':3,'=> Ce numéro existe DEJA');
        LIRCELLSYNDR ( NOUVSYNDR, CELLULE );
        TFSYNDCIADRESYNDR1J := CELLULE
        end
      else STOP := true;
      end
    { non trouvé }
    else writeln(' ':3,'Erreur, numéro NON existant');
  end
end

```

ANNEXES

```
else STOP := true;

  { continuation ou non }
if (not STOP) then
  begin
    write('Autre correction dans ce même symptôme no: ',NOSYMPT,
      ' (O/N) ? : '); read(CC); write(dellin); writeln;
    STOP := not (CC in ['0','o'])
  end
end;

if (NBS = 0) and (TROUVE) then
  writeln('Nombre de syndromes associés = 0; Cf "AJOUT":47);

write (' ':12, rvson,'Modification des cha
```

## ANNEXES

```

nages (O/N) ?',rvsoff,' : ');
read(TRM, CC);writeln;
if CC in ['O','o']
then
begin
  writeln;
  writeln('cha
  writeln('cha
  writeln('cha
  writeln;
  writeln;
  writeln(rvson, 'MODIFICATIONS:', rvsoff);
  NOTOK := true;
  while NOTOK do
  begin
    write('cha
    readln (A);
    write('cha
    readln (B);
    write('cha
    readln (C);
    writeln;
    write ('OK ? (O/N) : ');
    read(TRM, CC);
    NOTOK := CC in ['N','n'] ;
    writeln;

  end;
  if len (A) <> 0 then
  begin
    val (A, AN, P);
    if AN <> 0
    then
    begin
      ADRESLSTSYND ( AN, TROUVE, ADRESSE);
      if TROUVE
      then begin TFSYND [ ADRES + 2 ] := CODENUM (AN);
                TFSYND [ ADRES + 3 ] := CODENOMBRE (ADRESSE);
              end
      else writeln (rvson,'ERREUR: SYMPTOME ''PERE'' INCONNU')
      end
      else begin TFSYND [ ADRES + 2 ] := CODENUM (0);
                TFSYND [ ADRES + 3 ] := CODENOMBRE (0);
              end
      end;
  if len (B) <> 0 then
  begin
    val (B, BN, P);
    if BN <> 0
    then
    begin
      ADRESLSTSYND ( BN, TROUVE, ADRESSE);
      if TROUVE
      then begin TFSYND [ ADRES + 4 ] := CODENUM (BN);
                TFSYND [ ADRES + 5 ] := CODENOMBRE (ADRESSE);
              end
      else writeln (rvson,'ERREUR: SYMPTOME ''FILS'' INCONNU')
      end
      else begin TFSYND [ ADRES + 4 ] := CODENUM (0);
    
```

## ANNEXES

```

                TFSYND [ ADRES + 5 ] := CODENOMBRE (0);
            end
        end;
    if len (C) <> 0 then
    begin
        val (C, CN, P);
        if CN <> 0
        then
            begin
                ADRESLSTSYND ( CN, TROUVE, ADRESSE);
                if TROUVE
                then begin TFSYND [ ADRES + 6 ] := CODENUM (CN);
                    TFSYND [ ADRES + 7 ] := CODENOMBRE (ADRESSE);
                        end
                    else writeln (rvson,'ERREUR: SYMPTOME ''FRERE'' INCONNU')
                end
            else begin TFSYND [ ADRES + 6 ] := CODENUM (0);
                TFSYND [ ADRES + 7 ] := CODENOMBRE (0);
                    end
            end
        end
    end;
    { continuation ou non }
write(' ':12,rvson,'Modification d''un autre symptôme (O/N) ?',rvsoff,
' ');
    read(TRM, CC);
    CONT := (CC in ['0','o']);
end;

    { Ecriture de TFSYND modifié dans le fic. FSYNDROM }
assign (FY, 'FSYNDROM'); rewrite(FY, 2);
blockwrite(FY,TFSYND,MAXELEMFSYND);
close(FY);
end { MODFSYND };

procedure SUPSYMPT;

const STR1:string[70]= 'SUPPRESSION D''UN SYMPTOME DANS LE FICHIER FSYNDROM'

var STI          : string[10];
    TROUVE       : boolean;
    CC           : char;
    I, K         : integer;
    ADRES        : integer;
    NBSYNDECRI  : integer; { nombre de syndromes associés à un symptôme
                             donné }
    NOSYMPT     : integer;

begin
    { Contrôle de sécurité }
    writeln; write(' ':12);
    write(rvson);write('MOT DE PASSE, S.V.P. : ');write(rvsoff);write(' ');
    readln(STI); writeln;

    if (STI='VINCENT') then
        begin
            write(' ':12); write(rvson);

```

ANNEXES

```

write('DANGER : CONFIRMEZ VOUS VOTRE CHOIX (O/N) ? : ');
write(rvsoff); write(' '); readln(CC);

    { Démarrage de la proc. }
if (CC in ['O','o']) then
begin
    VIDECRAN; TITRE(STR1);
    writeln; writeln;

    RECHERCH ( NOSYMPT,TROUVE,ADRES );

    { Cas réussite de la recherche dans les fic. SYMPTOME
                                                    ou FSYNDROM
if TROUVE then
begin
    NBSYNDECRI := DECODENOMBRE ( TFSYNDIADRES + 1J );
    if (NBSYNDECRI > 0)
    then
    begin
        K := MAXTFSYND - NBSYNDECRI;
        for I:= ADRES + 8 TO K do
            TFSYNDI[I] := TFSYNDI[I] + NBSYNDECRI;
        TFSYNDIADRES + 1J := CODENOMBRE ( 0 );
        for I := (K + 1) TO MAXTFSYND do TFSYNDI[I]:=0;
        MAXTFSYND:=MAXTFSYND-NBSYNDECRI;

        { Modification du fic. SYMPTOME }
        CHANGADRDSYMPTOME ( NOSYMPT,-NBSYNDECRI,0);

        { Affichage écran }
        writeln('                                SUPPRESSION EFFECTUEE');
        writeln;
        writeln(' VEUILLEZ SUPPRIMER CE SYMPTOME DANS LE FICHIER
                ' SYMPTOME');
        writeln;
        writeln(' VEUILLEZ SUPPRIMER LES REFERENCES A CE SYMPTOME
                ' DANS LE FICHIER FSYMPTOM');
        AVANCER
    end
    else
    begin
        writeln(' ':15,'SYMPTOME SANS "AUCUN" SYNDROME ASSOCIE !');
        AVANCER
    end
end
end
end
end { SUPSYMPT };

procedure SUPSYNDR;
const
    STR2:string[70]='SUPPRESSION DE SYNDROMES DANS LE FICHIER "FSYNDROM"
var
    CONT, STOP      : boolean;
    CC              : char;
    NBSUPPR, J      : integer;
    TROUVE          : boolean;
    TROU1, TROU2    : boolean;

```

## ANNEXES

```

NBSYNDECRI   : integer; { nombre de syndromes associés à un symptôme
                             donné
NOSYNDR      : integer;
NOSYMPT      : integer;
ADRES, ADRESYNDR : integer;

```

```

begin
  CONT:=true;

  while CONT do
    begin
      VIDECRAN; TITRE(STR2); STOP:=true; NBSUPPR := 0;

      RECHERCH ( NOSYMPT, TROUVE, ADRES );

      if TROUVE
      then
        begin
          { Cas réussite de la recherche rapide ou lente }
          NBSYNDECRI := DECODENOMBRE ( TFSYNDCADRES + 1);
          STOP := false;

          if (NBSYNDECRI > 0) then
            writeln ( ' ':12, rvson, 'nombre de syndromes associés : ',
                      NBSYNDECRI, rvsoff );
            while (not STOP) and (NBSYNDECRI > 0) do
              begin
                TROU1 := false; TROU2 := false;

                { Test si entrée des données par code ou par numéro }
                if (not FAIT) then LIRLSTCODES ( FAIT, BASCULE);
                if BASCULE
                then
                  begin
                    repeat
                      write(' ':12, 'code du syndrome associé à supprimer,',
                              ' s.v.p. : '); readln( C );
                      if len ( C ) > 0 then RECHCOD ( C, T, SUP, TROU2, NOSYNDR)
                      if ( not TROU2 ) and ( len ( C ) > 0 )
                      then writeln(' Erreur dans le code entré !');
                      until ( TROU2 ) or ( len ( C ) = 0 );
                      if len ( C ) = 0 then NOSYNDR := 0
                    end
                  else
                    begin
                      write(' ':12, 'numéro du syndrome associé à supprimer,',
                              ' s.v.p. : ');
                      LIRENTIER ( NOSYNDR )
                    end;

                if (NOSYNDR > 0)
                then
                  begin
                    { opération de recherche du syndrome associé }
                    RECHSYNDANSLIST ( ADRES + 8, ADRES + 7 + NBSYNDECRI,
                                      NOSYNDR, TROU1, ADRESYNDR );

                    { trouvé }
                    if (TROU1)

```

## ANNEXES

```

then
  begin
    for J := ADRESYNDR TO (MAXTFSYND-1) do
      begin
        TFSYNDECJ := TFSYNDECJ + 1;
      end;
    TFSYNDCMAXTFSYND := 0; MAXTFSYND := MAXTFSYND - 1;
    TFSYNDCADRES + 1 := CODENOMBRE ( DECODENOMBRE (
      TFSYNDCADRES + 1 ) - 1 );
    NBSYNDECRI := NBSYNDECRI - 1; NBSUPPR := NBSUPPR + 1;
    writeln(' ':12,'SYNDROME ASSOCIE NO:',NOSYNDR,
      ' SUPPRIME');
  end
  { non trouvé }
else writeln(' Erreur, numéro non existant');

write(' ':12,'une autre suppression pour ce même symptôme '
  '(O/N) ? : ');
read(TRM, CC); write(dellin); writeln;
STOP := not (CC in ['0','o'])
end
else STOP := true
end; { endwhile }
if (NBSYNDECRI = 0) and TROUVE then
  writeln('nombre de syndromes associés = 0':41);
if (NBSUPPR > 0) then
  CHANGADRDSYMPROME ( NOSYMP, -NBSUPPR, NBSYNDECRI );
write(' ':12,rvson,'Suppression dans un autre symptôme (O/N) ?',
  rvsoff,' : ');
read(TRM, CC);
CONT := (CC in ['0','o'])
end
else CONT := false
end
end { SUPSYNDR };

procedure SUPFSYND;
const
  ST4 :string[36]= 'SUPPRIMER DANS LE FICHER FSYNDROM';
  STR1:string[70]= 'SUPPRESSION D'UN SYMPTOME DANS LE FICHER FSYNDROM
  STR2:string[70]='SUPPRESSION DE SYNDROMES DANS LE FICHER FSYNDROM';
var
  FY      : file;
  CC      : char;
  I       : integer;
begin
  { Menu }
  VIDECRAN; TITRE ( ST4 ); for I := 1 TO 5 do writeln;

  writeln('A:   ':15,STR1); writeln; writeln;
  writeln('B:   ':15,STR2); writeln; writeln;
  gotoxy (0, 14); write('VOTRE CHOIX S.V.P. (A/B) ? : ');
  LIREP ( 27, 14, ['A','a','B','b',' '], CC);
  case CC of
    'A','a' : SUPSYMPT ( suppression dans le cas A );
    'B','b' : SUPSYNDR ( suppression dans le cas B );
    ' '      : ( RIEN );
  end;
end;

```

## ANNEXES

```

assign (FY, 'FSYNDROM'); rewrite(FY, 2);
blockwrite(FY, TFSYND, MAXELEMFSYND);
close(FY);

end ( SUPFSYND );

procedure AFFSYND;

var STOP, TROUVE, COHERENT: boolean;
    C : string [4];
    J, P, NODERN, NOSYMP, ADRES, I, NUMSYND, NBSYND : integer;
    QSTI : char;

begin
  if not TFSYNDEXIST then LIRFSYND;
  STOP := false;
  while not STOP do
    begin
      VIDECRAN;
      ECRIRINV (10,0, 'VISUALISATION DU FICHER "FSYNDROME"');
      ECRIRINV (12,5, 'sympt8me concerné ?:');
      gotoxy ( 12 + 23, 5 );
      LIRENTIER ( NOSYMP );
      STOP := NOSYMP <= 0;
      if NOSYMP > 0
      then
        begin
          DERBLOC (NODERN);
          STOP := NOSYMP > NODERN;
          if not STOP
          then
            begin
              ADRESLSTSYND ( NOSYMP, TROUVE, ADRES);
              if TROUVE
              then
                begin COHERENT := DECODENUM ( TFSYND [ADRES] ) = NOSYMP;
                  if COHERENT
                  then begin AFFORMNATUR ( NOSYMP );
                    writeln('cha
                    writeln('cha
                    writeln('cha

                    NBSYND := DECODENOMBRE ( TFSYND [ADRES + 1] );
                    write (rvson, 'NOMBRE DE SYNDROMES : ', rvsoff);
                    writeln (NBSYND);
                    writeln; J := 10;
                    for I := ADRES + 8 to ADRES + 7 + NBSYND do
                      begin J := J + 1;
                        DECODESY ( TFSYND [I], NUMSYND, QSTI);
                        writeln ('syndrome no. ', NUMSYND : 4,
                          ' code : ', T [NUMSYND]: 4,
                          ' fréquence : ', QSTI);
                        if J mod 20 = 0 then
                          begin writeln (rvson, 'Appuyez sur <RC> pour continuer ', rvsoff); readln end;
                            end;
                          writeln; writeln;
                          write ('tapez sur <RC> pour continuer. ');

```



## ANNEXES

```

writeln('*','5 : VISUALISATION':28,'6 : FIN DE CESSION':38,'*':12)
writeln('*','*':78);
write('* ');
write(rvson); write('VOTRE CHOIX S.V.P. (1/2/3/4/5/6) ?'); write(rvsoff)
writeln(' : ','*':39);
writeln('*','*':78);
writeln('*','*':78);
writeln('*****',
'*****');
LIREP ( 40,17,[ '1'..'6', ' ' ],REP );

( PARTIE GESTION DU FICHIER FSYNDROM PROPEMENT DITE

if (REP in [ '1'..'5' ]) then
begin
  assign ( FSPT, 'C:SYMPTOME.DBF' ); reset ( FSPT );
  case REP of
    '1' : begin
      VIDECRAN; TITRE ( ST1 );
      gotoxy ( 0,8 );
      writeln (rvson,'ATTENTION, VOUS ALLEZ DETRUIRE LE FICHIER':44
      ' FSYNDROM S'IL A DEJA ETE CREE ',rvsoff);
      writeln;
      write ( ' ':10,rvson,'MOT DE PASSE, S.V.P.',rvsoff,' : ' );
      readln(STI); writeln;
      if STI='VINCENT' then
      begin
        write('ETES VOUS D'ACCORD ? (O/N) : ':49); readln(CC);
        if (CC in [ '0', 'o' ]) then
          begin
            INITFSYND; APPENDFSYND ( ST1 )
          end
        end;
      end;
    '2' : begin ( menu de la partie AJOUT )
      VIDECRAN; TITRE ( ST2A ); for I := 1 TO 5 do writeln;
      writeln ( ' A - AJOUT D'UN BLOC-SYMPTOME' );
      writeln; writeln;
      writeln ( ' B - INSERTION DE SYNDROMES POUR UN SYMPTOME EXISTANT' );
      gotoxy ( 0,14 ); write('VOTRE CHOIX S.V.P. (A/B) ?');
      LIREP ( 27, 14, [ 'A', 'a', 'B', 'b', ' ' ], CC );
      case CC of
        'A', 'a' : APPENDFSYND ( ST2 );
        'B', 'b' : AJFSYND;
      end;
    end;
    '3' : MODFSYND;
    '4' : SUPFSYND;
    '5' : AFFSYND;
  end;
  close(FSPT);
end;

( FIN DE LA PARTIE GESTION "PURE" DE "FSYNDROM" )
until REP in [ '6', ' ' ];
VIDECRAN;
gotoxy(27,11); write(rvson); write(' AU REVOIR '); write(rvsoff);
gotoxy(0,23);
end.

```

ANNEXES

ANNEXE 7

PROGRAMME DE GENERATION DU FICHER FSYMPTOM.

```
program GENFSYMP;

const  MAXELEMFSYND = 8000;
       MAXELEMFSYMP = 2600;
       NBSYND       = 102;

type
  ENSREP      = set of char;
  CHAINE      = string [80];
  ENSEP       = set of char;
  TYPTFSYND   = array [1..MAXELEMFSYND] of integer;
  TYPTFSYMP   = array [1..MAXELEMFSYMP] of integer;
  ESYNDROME   = record
    LIBRE      : char;
    NOMSYND    : string[120];
    NUMERO     : string[4];
    CODESYND   : string[3];
    NBDESYMP   : string[4];
    RENSEIGT   : string[33];
    PREVALEN   : string[1];
    ADLSTSYMP  : string[4];
    ADLSTES    : string [4]
  end;

var    TFSYND : TYPTFSYND ;
       TFSYMP : TYPTFSYMP ;
       FSD    : file of ESYNDROME;
       BUFFSD : ESYNDROME;
       ENCORUNSYMP, APPARTIENT : boolean;
       FSYND, FSYMP : file;
       N, CEL, X, Y, NBS, I, J, K, L, NUMSYMP, NUM : integer;
       QSTI, R : char;

function CODENUM (NUM : integer): integer;
begin
  CODENUM := - (NUM + 256)
end;

function DECODENUM (CEL : integer): integer;
begin
  DECODENUM := - (CEL + 256)
end;

function CODENOMBRE (NB : integer): integer;
begin
  if NB = 0 then CODENOMBRE := - 255 else CODENOMBRE := - NB
end;

function DECODENOMBRE (CEL: integer): integer;
begin
  if CEL = - 255 then DECODENOMBRE := 0 else DECODENOMBRE := - CEL
end;
```

## ANNEXES

```

function CODESY (N : integer; QSTI : char): integer;
var Q : byte;
begin
  case QSTI of
    'Q' : Q := 0;
    'S' : Q := 1;
    'T' : Q := 2;
    'I' : Q := 3;
  end;
  CODESY := N * 4 + Q;
end;

procedure DECODESY (CEL : integer; var N : integer; var QSTI : char);
const T : string [4] = 'QSTI';
begin
  N := CEL DIV 4;
  QSTI := T [CEL MOD 4 + 1];
end;

procedure VIDECRAN;
begin
  writeln (CLRHM)
end;

procedure ECRIRINV (X,Y : integer; S : CHAINE);
begin
  writeln (rvson ); gotoxy (X,Y); write (S, rvsoff );
  gotoxy (X + len (S) + 1,Y);
end;

procedure LIREP (X,Y : integer; E : ENSREP; var R : char);
begin
  gotoxy (X,Y); read (R);
  while not (R in E) do
    begin gotoxy (X,Y); writeln (' '); gotoxy (X,Y); read (R) end
end (* LIREP *);

begin
  { PROGRAMME PRINCIPAL }
  X := 0; Y := 3;
  VIDECRAN;
  ECRIRINV (X,Y,
  'ATTENTION, VOUS ALLEZ DETRUIRE LE FICHER FSYMPTOM EXISTANT');
  ECRIRINV (X,Y + 1, 'MAINTENEZ-VOUS VOTRE DEMANDE ? (O/N) :');
  LIREP (X + 38, Y + 1, ['o', 'O', 'n', 'N', ' '], R);
  if R in ['o', 'O'] then
    begin
      assign (FSYND, 'FSYNDROM'); reset (FSYND, 2);
      assign (FSD, 'SYNDROME.DBF'); reset (FSD);
      assign (FSYMP, 'FSYMPTOM'); rewrite (FSYMP, 2);
      blockread (FSYND, TFSYND, MAXELEMFSYND);
      J := 1;
      for I := 1 to NBSYND do
        begin
          TFSYMP [J] := CODENUM (I);
          TFSYMP [J + 1] := CODENOMBRE (I);
          ENCORUNSYMP := true;
          K := 1;
          while ENCORUNSYMP do

```

## ANNEXES

```

begin
  APPARTIENT := false;
  NBS := DECODENOMBRE (TFSYND [K + 1]);
  L := 1;
  while (L <= NBS) and not APPARTIENT do
  begin
    DECODESY (TFSYND [K + 7 + L], NUM, QSTI);
    APPARTIENT := NUM = I;
    L := L + 1;
  end;
  if APPARTIENT then
  begin
TFSYMP [J + 1] := CODENOMBRE ( DECODENOMBRE (TFSYMP [J + 1]) + 1);
TFSYMP [J + 1 + DECODENOMBRE (TFSYMP [J + 1])]
:= CODESY ( DECODENUM ( TFSYND [K]), QSTI);
    end;
    K := K + 1 + 6 + DECODENOMBRE ( TFSYND [K + 1]) + 1;
    ENCORUNSYMP := TFSYND [K] <> 0;
  end;
  seek (FSD, I + 2); read (FSD, BUFFSD);
  N := DECODENOMBRE (TFSYMP [J+1]);
  str(N, BUFFSD.NBDESYMP);
  str(J, BUFFSD.ADLSTSYMP);
  seek (FSD, I + 2); write (FSD, BUFFSD);
  J := J + 1 + DECODENOMBRE ( TFSYMP [J + 1]) + 1;
end;
TFSYMP [J] := 0;
VIDECRAN;
ECRIRINV (36, 12, 'TERMINE');
gotoxy (0, 23)
end;
blockwrite (FSYMP, TFSYMP, MAXELEMFSYMP);
close (FSYND); close (FSYMP);
end (* GENFSYMP *).

```

## ANNEXES

## ANNEXE 8

GÉNÉRATION DU FICHIER FES DES LISTES DE COUPLES (es,v),  
LIÉ à SYNDROME.DBF

```

program GENFES;
( génération du fichier des listes de couples (es,v), lié à SYNDROME.DBF )

const
  MAXELEMFSYND = 8000;
  MAXELEMFSYMP = 2600;
  MAXSYND      = 102;
  MAXELEMFES  = 14000;
  MAXES       = 200;

type
  ST6 = string [6];
  ENSREP = set of char;
  CHAINE = string [80];
  ENSEP = set of char;
  TYPTFSYND= array [1..MAXELEMFSYND] of integer;
  TYPTFSYMP= array [1..MAXELEMFSYMP] of integer;
  TYPTFES = array [1..MAXELEMFES] of integer;
  ESYMPTOME = record
    CAR_DB3 : char;
    FORMNAT: string[90]; NUMERO: string[4];
    ENTITSYM: string[45]; CODENTSYM: string[4];
    VALNAT: string[25]; VALNUM: string[3];
    NIVOBSAXE: string[ 4]; ATTRIBUT: string[46];
    PERE: string[4]; FILS: string[4];
    FRERE: string[4]; NIVOBS: char;
    NBDESYND: string[3]; ADLSTSYND: string[4];
    CHAPITRE : string [2]; ESMERE : string[82];
    AXE : string[57]; NUMAXE : string[3];
    TYPAXE : string [2]
  end;
  ESYNDROME = record
    LIBRE : char;
    NOMSYND : string[120];
    NUMERO : string[4];
    CODESYND : string[3];
    NBDESYMP : string[4];
    RENSEIGT : string[33];
    PREVALEN : string[1];
    ADLSTSYMP: string[4];
    ADLSTES : string [4]
  end;

var
  NOES, NBES, P, NOSYMP, NBSYMP, ADR, ADRES, I, J, K : integer;
  FES, FSYMP : file;
  FSP : file of ESYMPTOME;
  FSD : file of ESYNDROME;
  ESP : ESYMPTOME;
  ESD : ESYNDROME;
  TFSYMP : TYPTFSYMP;
  TFES : TYPTFES;
  QSTI : char;

```

## ANNEXES

```

TROUVE : boolean;

function CODENUM (NUM : integer): integer;
begin
  CODENUM := - (NUM + 256)
end;

function DECODENUM (CEL : integer): integer;
begin
  DECODENUM := - (CEL + 256)
end;

function CODENOMBRE (NB : integer): integer;
begin
  if NB = 0 then CODENOMBRE := - 255 else CODENOMBRE := - NB
end;

function DECODENOMBRE (CEL: integer): integer;
begin
  if CEL = - 255 then DECODENOMBRE := 0 else DECODENOMBRE := - CEL
end;

function CODESY (N : integer; QSTI : char): integer;
var Q : byte;
begin
  case QSTI of
    'Q' : Q := 0;
    'S' : Q := 1;
    'T' : Q := 2;
    'I' : Q := 3;
  end;
  CODESY := N * 4 + Q;
end;

procedure DECODESY (CEL : integer; var N : integer; var QSTI : char);
const T : string [4] = 'QSTI';
begin
  N := CEL DIV 4;
  QSTI := T [CEL MOD 4 + 1];
end;

begin
  ( main program )

  assign (FSYMP, 'FSYMP.TOM'); reset (FSYMP, 2);
  assign (FES, 'FES'); rewrite (FES, 2);
  assign (FSD, 'SYNDROME.DBF'); reset (FSD);
  assign (FSP, 'SYMPTOME.DBF'); reset (FSP);
  blockread (FSYMP, TFSYMP, MAXELEMFSYMP);

  ADR := 1; ADRES := 1;
  for I := 1 to MAXSYND do
  begin
    seek (FSD, I + 2); read (FSD, ESD);
    str ( ADRES, ESD.ADLSTES);
    seek (FSD, I + 2); write ( FSD, ESD);

    TFES [ADRES] := TFSYMP [ADR]; ( numéro du syndrome )
    TFES [ADRES + 1] := TFSYMP [ADR + 1]; ( nb. de symptômes = couples (es,

```

## ANNEXES

```

NBSYMP := DECODENOMBRE ( TFSYMP [ADR + 1] );
for J := 1 to NBSYMP do
begin
  DECODESY ( TFSYMP [ADR + 1 + J], NOSYMP, QSTI );
  seek (FSP, NOSYMP + 2); read (FSP, ESP);

  K := 1;
  while ESP.NUMAXE [K] = ' ' do K := K + 1;
  val ( copy (ESP.NUMAXE, K, len (ESP.NUMAXE) - K + 1),
        TFES [ ADRES + 1 + 5 * J - 4 ], P);

  K := 1;
  while ESP.VALNUM[K] = ' ' do K := K + 1;
  val ( copy (ESP.VALNUM, K, len (ESP.VALNUM) - K + 1 ),
        TFES [ ADRES + 1 + 5 * J - 3 ], P);

  case ESP.TYPAXE [1] of
    'N' : TFES [ ADRES + 1 + 5 * J - 2 ] := 3;
    'T' : TFES [ ADRES + 1 + 5 * J - 2 ] := 1;
  end;

  TFES [ ADRES + 1 + 5 * J - 1 ] := ord ( ESP.NIVOBAXE [1] ) - 48;

  case ESP.TYPAXE [2] of
    'E' : TFES [ ADRES + 1 + 5 * J ] := 1;
    'S' : TFES [ ADRES + 1 + 5 * J ] := 2;
  end;
end;

ADRES := ADRES + 1 + 5 * NBSYMP + 1;
ADR := ADR + 1 + NBSYMP + 1;
end;
blockwrite (FES, TFES, MAXELEMES);
close ( FES )

end. ( GENFES )

```

## ANNEXES

## ANNEXE 9

VALEURS PROPRES ET POURCENTAGES SIMPLES ET CUMULES D'INERTIE,  
NIVEAU D'OBSERVATION n° 1 (clinique immédiat)

NOMBRE DE VARIABLES : 67

Nombre total de syndromes : 102

Nombre de syndromes non représentés à ce niveau : 0

VALEURS PROPRES :	% INERTIE :	% INERTIE CUMULEE :
D [ 1 ] = 13.05	39.80	39.80
D [ 2 ] = 3.73	11.37	51.18
D [ 3 ] = 2.91	8.86	60.04
D [ 4 ] = 1.81	5.53	65.57
D [ 5 ] = 1.37	4.19	69.75
D [ 6 ] = 0.96	2.94	72.69
D [ 7 ] = 0.88	2.69	75.39
D [ 8 ] = 0.80	2.43	77.82
D [ 9 ] = 0.73	2.22	80.04
D [10] = 0.60	1.82	81.86
D [11] = 0.59	1.80	83.66
D [12] = 0.53	1.61	85.27
D [13] = 0.50	1.53	86.80
D [14] = 0.40	1.23	88.03
D [15] = 0.36	1.08	89.11
D [16] = 0.33	1.01	90.12
D [17] = 0.31	0.93	91.05
D [18] = 0.29	0.88	91.93
D [19] = 0.26	0.80	92.73
D [20] = 0.24	0.73	93.46
D [21] = 0.21	0.65	94.11
D [22] = 0.20	0.60	94.70
D [23] = 0.19	0.56	95.27
D [24] = 0.16	0.50	95.77
D [25] = 0.13	0.41	96.17
D [26] = 0.12	0.37	96.54
D [27] = 0.11	0.35	96.89
D [28] = 0.10	0.31	97.20
D [29] = 0.09	0.28	97.48
D [30] = 0.09	0.26	97.74
D [31] = 0.08	0.25	97.99
D [32] = 0.07	0.21	98.20
D [33] = 0.06	0.19	98.39
D [34] = 0.05	0.15	98.54
D [35] = 0.05	0.15	98.69
D [36] = 0.05	0.14	98.84
D [37] = 0.04	0.12	98.96
D [38] = 0.04	0.11	99.07
D [39] = 0.03	0.01	99.17
D [40] = 0.03	0.09	99.26
D [41] = 0.03	0.08	99.35
D [42] = 0.03	0.08	99.43
D [43] = 0.02	0.07	99.50
D [44] = 0.02	0.07	99.56
D [45] = 0.02	0.06	99.62
D [46] = 0.02	0.05	99.67

ANNEXES

D [47]	=	0.01	0.05	99.72
D [48]	=	0.01	0.04	99.76
D [49]	=	0.01	0.04	99.80
D [50]	=	0.01	0.03	99.83
D [51]	=	0.00	0.03	99.86
D [52]	=	0.00	0.02	99.88
D [53]	=	0.00	0.02	99.90
D [54]	=	0.00	0.02	99.92
D [55]	=	0.00	0.02	99.94
D [56]	=	0.00	0.01	99.95
D [57]	=	0.00	0.01	99.97
D [58]	=	0.00	0.01	99.98
D [59]	=	0.00	0.00	99.99
D [60]	=	0.00	0.00	99.99
D [61]	=	0.00	0.00	100.00
D [62]	=	0.00	0.00	100.00
D [63]	=	0.00	0.00	100.00
D [64]	=	0.00	0.00	100.00
D [65]	=	0.00	0.00	100.00
D [66]	=	0.00	0.00	100.00
D [67]	=	0.00	0.00	100.00

## ANNEXES

## ANNEXE 10

## LISTE DES 12 PREMIERS VECTEURS PROPRES DU NIVEAU 1

AXE	V1	V2	V3	V4	V5	V6	V7	V8
1	0.00	0.00	-0.02	0.03	-0.04	-0.05	-0.02	-0.01
2	0.04	0.12	-0.09	0.06	-0.04	0.14	-0.14	0.02
3	0.00	-0.00	0.00	-0.00	-0.00	0.01	-0.04	0.02
4	0.00	0.02	-0.02	0.11	-0.13	0.02	0.04	0.09
5	0.00	-0.00	-0.00	0.00	0.04	0.00	0.01	0.02
6	0.00	0.02	-0.07	0.09	-0.21	-0.55	-0.33	-0.38
7	-0.00	0.00	0.00	-0.02	0.01	0.02	0.02	0.01
8	0.00	0.02	-0.02	-0.01	-0.02	-0.03	0.02	0.05
9	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.02	0.00
10	0.00	-0.00	0.00	0.00	-0.00	0.02	-0.04	0.01
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00
13	-0.00	0.00	-0.00	-0.01	0.00	-0.00	0.00	0.00
14	-0.00	0.00	-0.00	-0.01	-0.00	-0.00	0.00	0.00
15	-0.00	0.02	-0.03	-0.02	0.03	-0.07	0.03	0.06
16	-0.00	-0.01	-0.00	0.01	0.00	0.02	0.00	0.02
17	-0.00	0.00	-0.00	0.01	0.02	0.00	0.00	-0.01
18	-0.00	-0.00	0.00	-0.00	0.00	0.03	0.02	-0.00
19	-0.01	-0.02	0.05	-0.05	0.01	0.01	-0.03	0.00
20	-0.00	-0.00	0.02	-0.05	0.02	0.06	0.05	-0.04
21	-0.00	-0.00	0.00	-0.01	0.02	0.04	-0.00	0.06
22	-0.00	0.00	-0.00	-0.01	0.00	-0.00	0.00	0.00
23	-0.00	0.00	-0.00	-0.01	0.00	-0.00	0.00	0.00
24	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00
25	0.00	0.01	-0.00	-0.02	-0.02	0.00	-0.02	-0.01
26	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00
27	0.00	-0.00	-0.00	-0.00	0.02	-0.00	0.02	0.00
28	-0.00	0.00	-0.00	0.02	-0.00	0.00	-0.00	0.00
29	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00
30	-0.00	0.00	0.00	-0.00	-0.06	-0.03	0.03	0.03
31	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.02	0.00
32	0.01	0.05	-0.06	0.16	-0.05	0.13	0.21	0.10
33	0.04	-0.05	-0.03	0.29	-0.14	0.05	-0.56	0.15
34	0.03	0.03	-0.17	0.44	-0.29	-0.24	0.18	-0.00
35	0.00	0.00	-0.04	0.20	-0.03	0.07	0.10	-0.11
36	0.00	0.00	-0.01	0.01	0.03	-0.00	0.05	0.00
37	-0.00	-0.00	-0.03	0.11	-0.06	0.13	0.08	-0.07
38	0.00	0.02	-0.02	0.15	-0.08	0.08	0.07	0.02
39	0.00	0.05	-0.03	0.08	0.02	0.02	0.03	0.06
40	0.00	0.09	-0.15	0.46	0.22	0.15	0.02	-0.43
41	0.00	0.04	-0.03	-0.08	-0.03	-0.08	-0.11	-0.31
42	0.03	0.07	-0.04	0.11	-0.35	0.33	0.06	0.07
43	0.00	0.04	-0.06	0.06	0.13	-0.06	-0.07	-0.02
44	-0.02	0.12	0.01	-0.00	0.13	0.20	-0.05	-0.22
45	0.02	0.14	-0.01	0.18	0.41	0.25	0.01	-0.20
46	0.04	0.04	-0.08	0.34	0.00	-0.24	0.26	0.40
47	0.01	0.00	-0.03	0.04	-0.03	-0.07	-0.00	0.00
48	-0.00	0.03	-0.03	0.00	0.00	-0.02	0.03	0.03
49	0.01	0.01	-0.02	0.06	0.04	0.00	0.07	0.00
50	0.00	0.02	-0.02	0.00	0.04	0.02	0.03	0.04
51	0.02	0.15	-0.02	0.08	0.06	-0.06	-0.02	0.23
52	0.00	0.07	-0.02	-0.00	-0.06	0.05	0.01	-0.00

## ANNEXES

53	0.00	0.03	-0.00	-0.00	-0.03	0.05	-0.00	0.06
54	0.05	-0.04	0.00	0.06	-0.12	0.35	-0.43	0.12
55	0.00	-0.02	-0.00	0.11	-0.09	0.11	-0.01	0.06
56	0.03	0.04	-0.05	0.05	-0.15	0.18	-0.13	-0.07
57	0.01	0.03	-0.02	-0.05	-0.05	0.05	-0.03	0.05
58	0.02	-0.02	0.02	0.05	-0.00	0.03	-0.13	0.14
59	-0.02	0.44	0.70	0.03	-0.33	0.03	0.16	-0.20
60	0.99	-0.03	0.06	-0.05	0.04	-0.03	0.03	-0.03
61	0.00	-0.00	0.00	-0.01	0.00	0.00	-0.02	0.00
62	0.03	0.74	-0.49	-0.30	-0.07	-0.05	-0.05	0.09
63	-0.02	0.35	0.40	0.23	0.43	-0.16	-0.27	0.26
64	0.02	0.08	-0.07	-0.00	0.13	-0.04	0.01	-0.06
65	0.00	0.03	-0.00	0.05	-0.03	0.04	-0.01	0.04
66	0.02	0.06	-0.00	-0.03	-0.19	0.16	0.01	0.05
67	0.03	0.08	-0.07	0.07	-0.12	0.01	-0.06	-0.11

Vecteurs suivants :

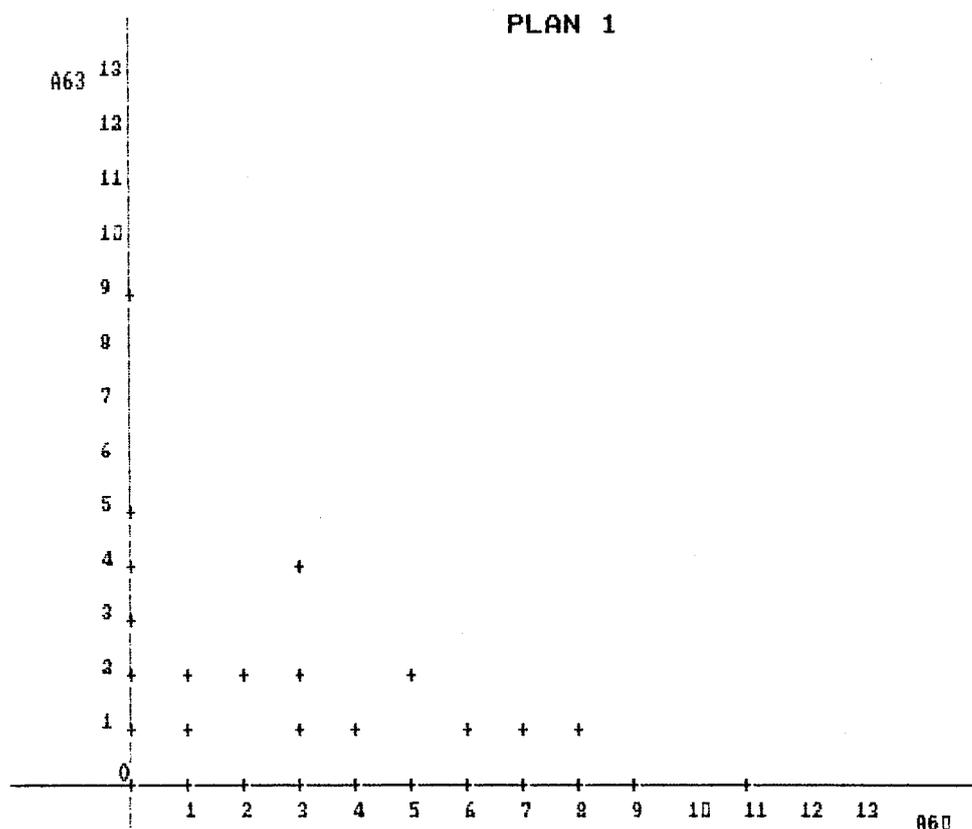
AXE	V9	V10	V11	V12
1	-0.04	0.04	0.04	-0.14
2	0.19	-0.45	-0.27	0.13
3	0.00	-0.00	0.00	0.03
4	-0.10	-0.17	0.08	-0.21
5	-0.01	-0.02	0.02	0.05
6	0.13	0.18	-0.33	-0.12
7	-0.00	-0.03	0.00	0.01
8	-0.03	0.03	-0.03	0.07
9	0.03	0.02	0.00	0.00
10	-0.00	-0.02	-0.01	0.00
11	0.00	0.00	0.00	0.00
12	0.00	0.00	0.01	0.02
13	0.00	0.00	0.00	0.02
14	-0.00	0.00	0.01	0.02
15	0.07	0.01	0.00	0.07
16	-0.04	0.00	0.05	0.00
17	-0.00	-0.01	-0.01	0.02
18	0.00	0.09	-0.04	0.05
19	-0.09	-0.02	0.04	-0.05
20	-0.03	0.06	-0.00	-0.06
21	-0.06	0.03	-0.00	0.04
22	0.00	0.00	0.00	0.02
23	0.00	0.00	0.00	0.02
24	0.00	-0.00	0.00	-0.00
25	-0.00	-0.00	-0.02	-0.00
26	0.00	-0.00	-0.00	0.00
27	0.05	-0.01	-0.00	-0.02
28	0.00	0.03	-0.04	0.05
29	0.00	0.00	0.00	0.00
30	-0.00	-0.02	0.00	-0.07
31	0.03	0.02	0.00	0.00
32	-0.09	-0.16	0.00	-0.22
33	0.07	-0.00	0.37	-0.13
34	-0.09	-0.05	0.26	0.05
35	-0.03	-0.01	0.03	-0.00
36	0.06	-0.00	0.02	-0.06
37	-0.04	-0.27	-0.05	0.05
38	-0.07	0.01	-0.13	-0.15

## ANNEXES

39	0.02	0.02	-0.08	-0.00
40	-0.18	0.17	-0.18	0.35
41	0.01	-0.32	0.14	-0.17
42	-0.18	0.37	-0.26	-0.31
43	0.12	-0.12	0.08	0.01
44	0.13	0.14	0.19	-0.13
45	0.36	0.09	0.11	-0.24
46	0.28	0.04	0.09	0.22
47	0.06	0.03	0.02	-0.03
48	0.09	-0.01	-0.02	0.02
49	-0.03	-0.12	-0.01	-0.05
50	0.18	0.04	0.00	-0.04
51	0.44	-0.02	-0.33	-0.28
52	0.21	0.08	-0.14	-0.09
53	-0.02	0.02	-0.16	-0.04
54	0.17	0.17	0.11	0.21
55	-0.10	-0.00	0.03	-0.01
56	0.01	-0.39	-0.15	0.18
57	-0.02	-0.02	-0.06	0.06
58	0.04	0.17	-0.00	0.03
59	0.20	0.02	0.18	0.12
60	-0.03	0.02	-0.00	0.00
61	0.00	0.00	0.00	0.02
62	-0.13	0.11	0.14	0.13
63	-0.38	-0.07	-0.19	-0.08
64	-0.11	-0.15	0.22	-0.33
65	-0.08	-0.04	-0.00	-0.06
66	0.04	-0.14	-0.22	0.07
67	-0.14	-0.12	0.00	-0.24

## ANNEXE 11

## PLAN 1



PLAN DES AXES A60 (abscisses) ET A63 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

( 0, 1): 11 49 53 58 66 70 5 26 52 60 61 71 87  
9 85

( 0, 2): 14 31 4 4 50

( 0, 3): 11

( 0, 4): 71 76

( 0, 5): 38

( 0, 9): 58

( 1, 0): 25 26 36 37 60 61 79 87 89 91 99 7 17  
18 21 22 27 29 33 36 37 39 41 42 68 74  
76 77 78 80 82 83 88 89 90 92 93 95 96  
99 100 102 22 39 62 70 82 92 101 102

( 1, 1): 18 29 11 28 75 47 80

( 1, 2): 25 52 78

( 1, 3): 14 32 53

( 1, 4): 38

( 2, 0): 1 2 28 35 86 79 81 101 43 77

( 2, 1): 6 9

( 2, 2): 29

( 2, 3): 7

( 3, 0): 19 20 45 91 19 60 79 95 96

( 3, 1): 44 53

( 3, 2): 71

( 3, 4): 21

( 4, 0): 2 20 28 89 94 100

( 4, 1): 72

ANNEXES

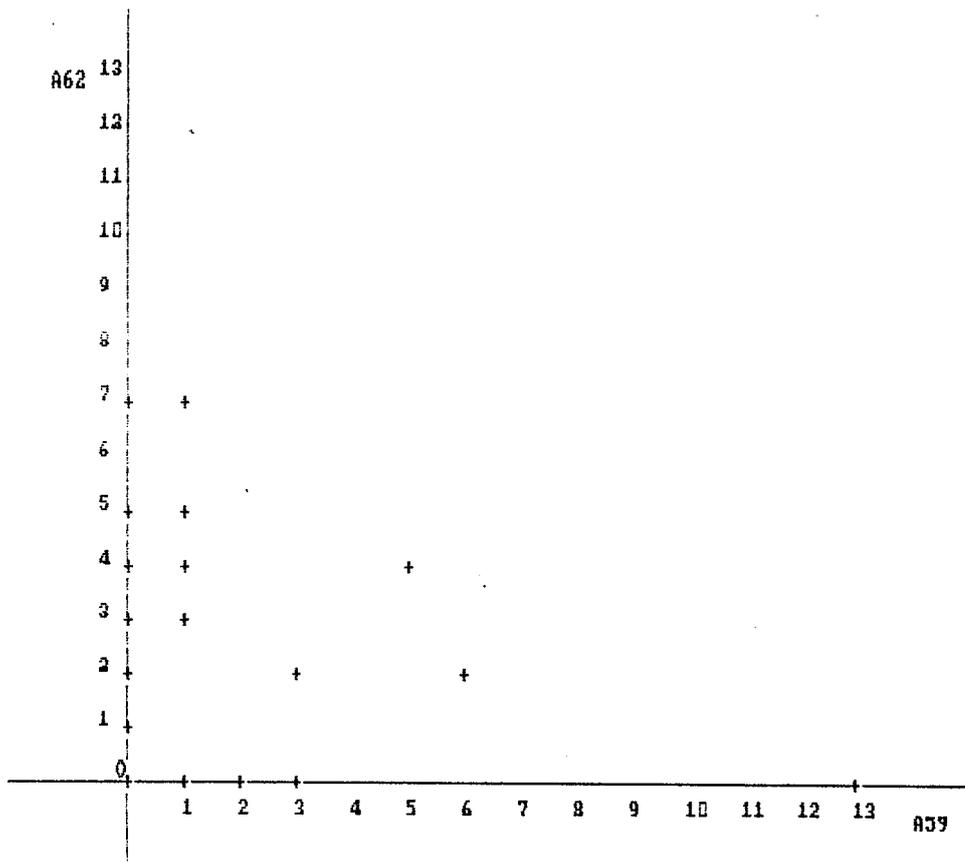
( 5, 0) : 10 26 86 87 93  
( 5, 2) : 27  
( 6, 0) : 33 69 74 83  
( 6, 1) : 1  
( 7, 0) : 25 36  
( 7, 1) : 45  
( 8, 0) : 57  
( 8, 1) : 30  
( 9, 0) : 32  
(11, 0) : 37

nombre de syndromes à l'origine :45

liste :

3	5	6	7	8	12	13	14	15	16	17	18	23	24	31
34	35	40	41	42	46	48	49	51	54	55	56	59	61	63
64	65	66	67	68	73	75	81	84	88	90	91	97	98	99

## PLAN 2



PLAN DES AXES A59 (abscisses) ET A62 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

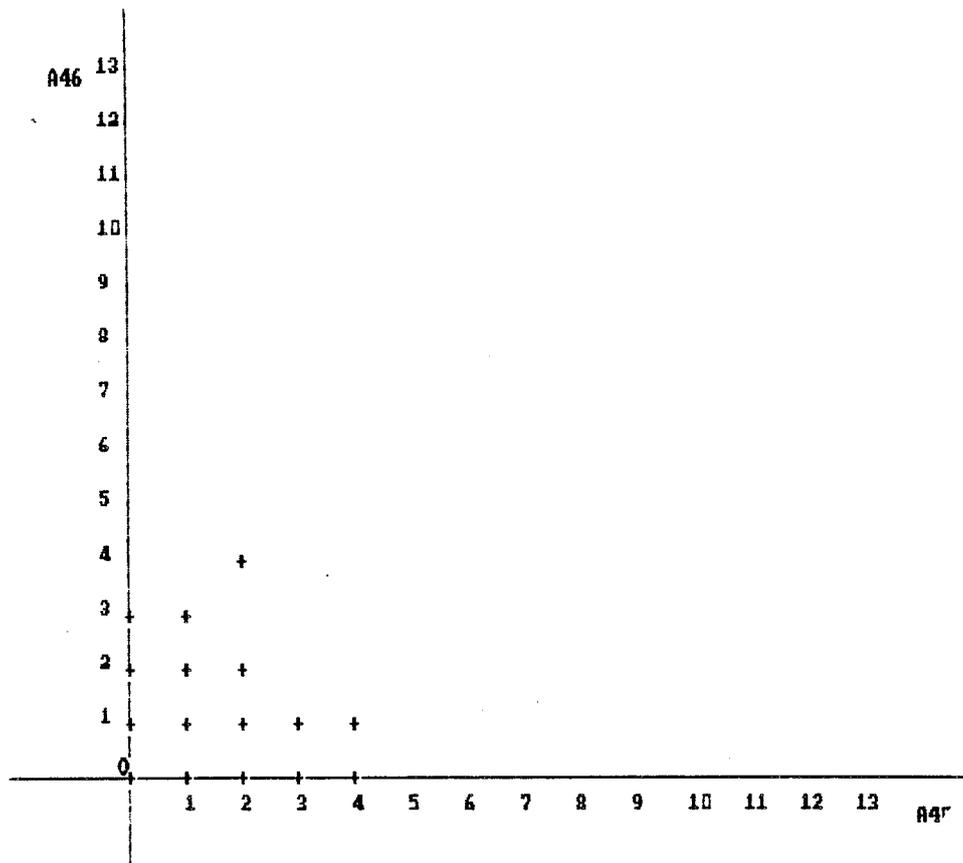
( 0, 1): 33  
 ( 0, 2): 17 72 90  
 ( 0, 3): 29 39  
 ( 0, 4): 4 5 27 34 37 52 59 80  
 ( 0, 5): 71 93  
 ( 0, 7): 11  
 ( 1, 0): 28  
 ( 1, 3): 16  
 ( 1, 4): 101  
 ( 1, 5): 14  
 ( 1, 7): 2  
 ( 2, 0): 46 74  
 ( 3, 0): 18  
 ( 3, 2): 25  
 ( 5, 4): 53  
 ( 6, 2): 49  
 (13, 0): 58

nombre de syndromes à l'origine :73

liste :

1	3	6	7	8	9	10	12	13	15	19	20	21	22	23
24	26	30	31	32	35	36	38	40	41	42	43	44	45	47
48	50	51	54	55	56	57	60	61	62	63	64	65	66	67
68	69	70	73	75	76	77	78	79	81	82	83	84	85	86
87	88	89	91	92	94	95	96	97	98	99	100	102		

## PLAN 3



PLAN DES AXES A40 (abscisses) ET A46 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

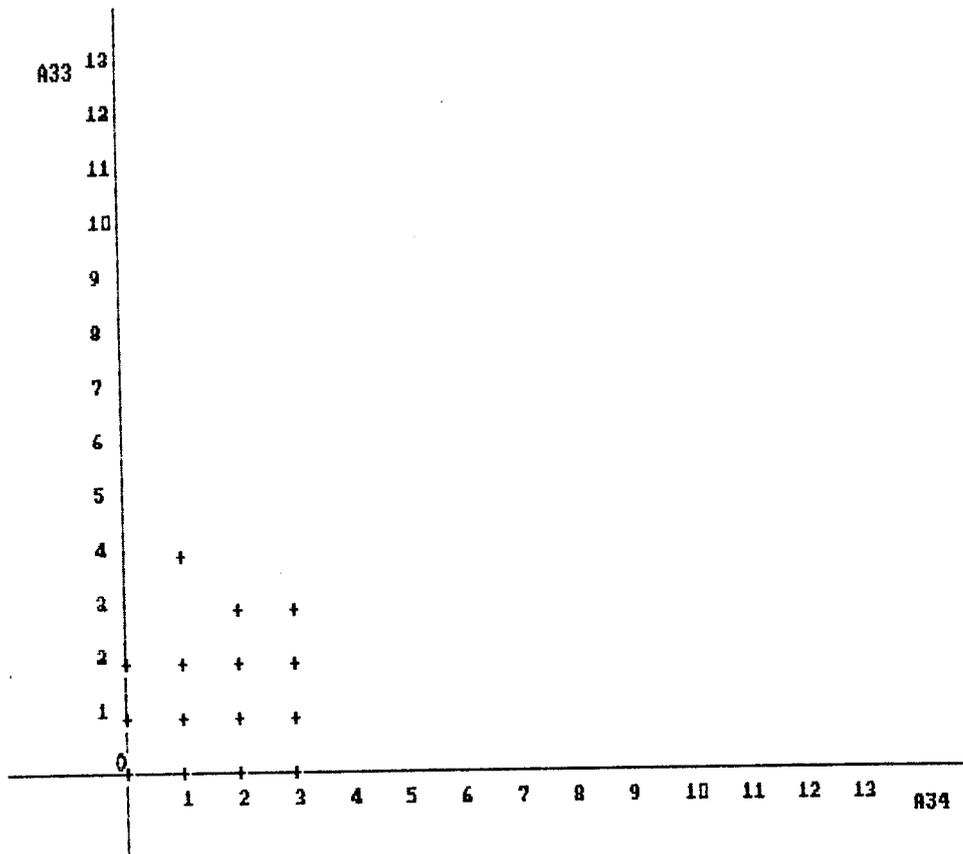
( 0, 1): 3 6 12 28 34 36 38 58 69 81 94  
 ( 0, 2): 37 75 90  
 ( 0, 3): 57  
 ( 1, 0): 9 19 22 25 30 44 45 53 71 77 80 85  
 ( 1, 1): 27 63 78 82 89 98  
 ( 1, 2): 5 91  
 ( 1, 3): 4 26 32  
 ( 2, 0): 7 50 70 100  
 ( 2, 1): 11 72 76 79 84 87 95 96 101  
 ( 2, 2): 83  
 ( 2, 4): 99  
 ( 3, 0): 31 97  
 ( 3, 1): 88  
 ( 4, 0): 29  
 ( 4, 1): 52

nombre de syndromes à l'origine :44

liste :

1	2	8	10	13	14	15	16	17	18	20	21	23	24	33
35	39	40	41	42	43	46	47	48	49	51	54	55	56	59
60	61	62	64	65	66	67	68	73	74	86	92	93	102	

## PLAN 4



PLAN DES AXES A34 (abscisses) ET A33 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

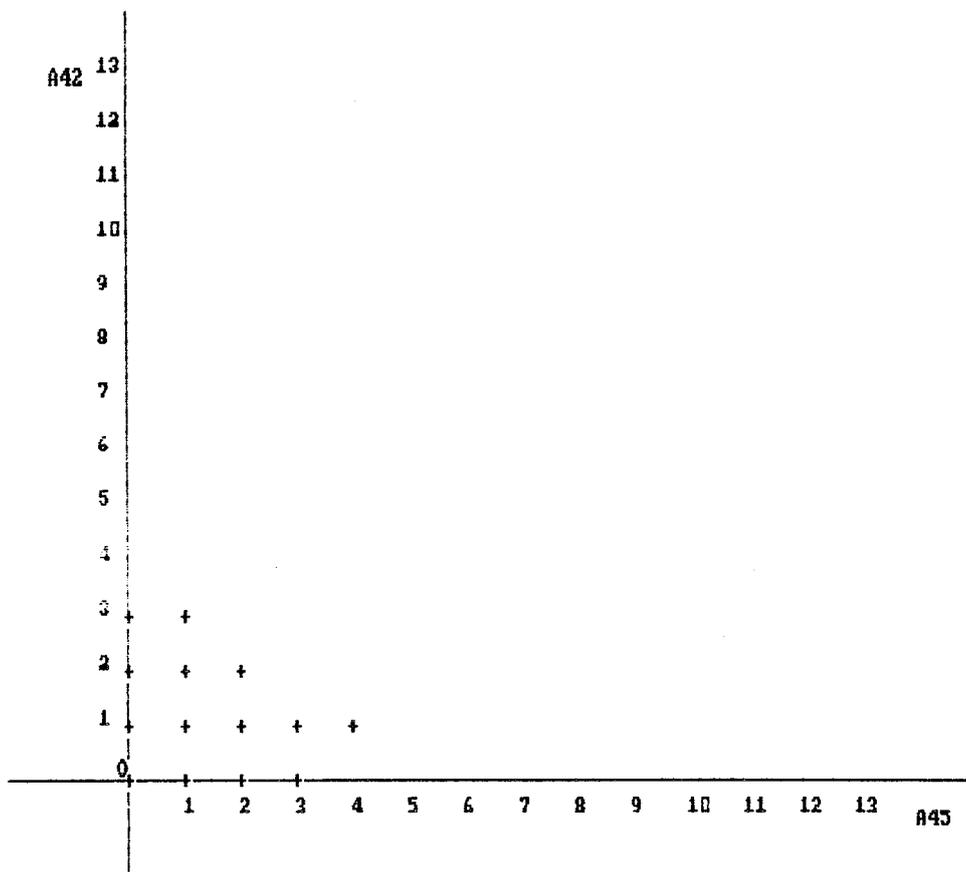
( 0, 1) : 1 12 18 47 60 61 71  
 ( 0, 2) : 19 28  
 ( 1, 0) : 5 6 7 9 16 31 33 37 40 42 44 49 59  
           62 63 68 69 81 84 85 88 89 94 96 102  
 ( 1, 1) : 11 30 32 35 82  
 ( 1, 2) : 86  
 ( 1, 4) : 45 78  
 ( 2, 0) : 27 41 52 53 73 92 101  
 ( 2, 1) : 93 100  
 ( 2, 2) : 57 91 97  
 ( 2, 3) : 77  
 ( 3, 0) : 26 79 90 95  
 ( 3, 1) : 4  
 ( 3, 2) : 99  
 ( 3, 3) : 87

nombre de syndromes à l'origine : 40

liste :

2	3	8	10	13	14	15	17	20	21	22	23	24	25	29
34	36	38	39	43	46	48	50	51	54	55	56	58	64	65
66	67	70	72	74	75	76	80	83	98					

## PLAN 5



PLAN DES AXES A45 (abscisses) ET A42 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

( 0, 1) : 10 15 30 35 37 39 45 46 49 54 55 57 63  
64 67 80 84 91

( 0, 2) : 3 8 86 90 95

( 0, 3) : 2 53

( 1, 0) : 1 12 22 27 31 34 44 50 56 58 65 68 70  
81 85 99 100

( 1, 1) : 7 13 17 19 25 28 33 38 51 77 82 83 93  
94 96 98

( 1, 2) : 26 72 76 78 87 89 92

( 1, 3) : 18 101

( 2, 0) : 11 14 66 88

( 2, 1) : 29

( 2, 2) : 79

( 3, 0) : 32 36 52

( 3, 1) : 5 9

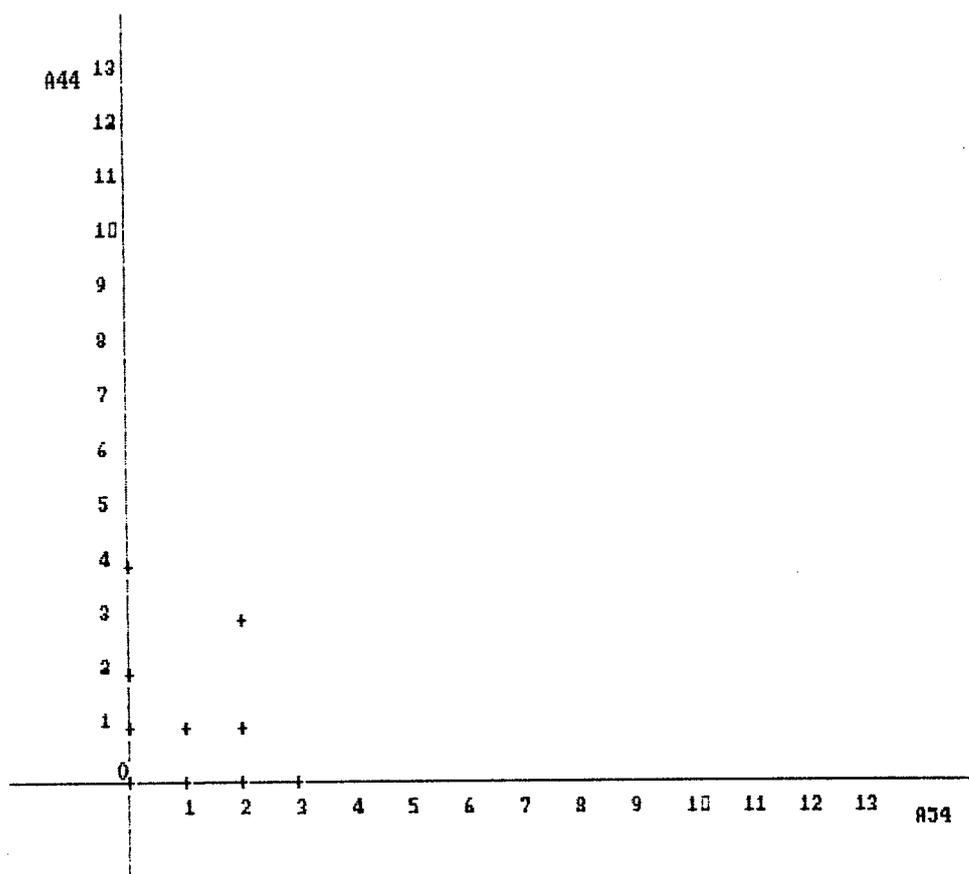
( 4, 1) : 71

nombre de syndromes à l'origine : 23

liste :

4 6 16 30 31 33 34 40 41 42 43 47 48 59 60  
61 62 69 73 74 75 97 102

## PLAN 6



PLAN DES AXES A54 (abscisses) ET A44 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

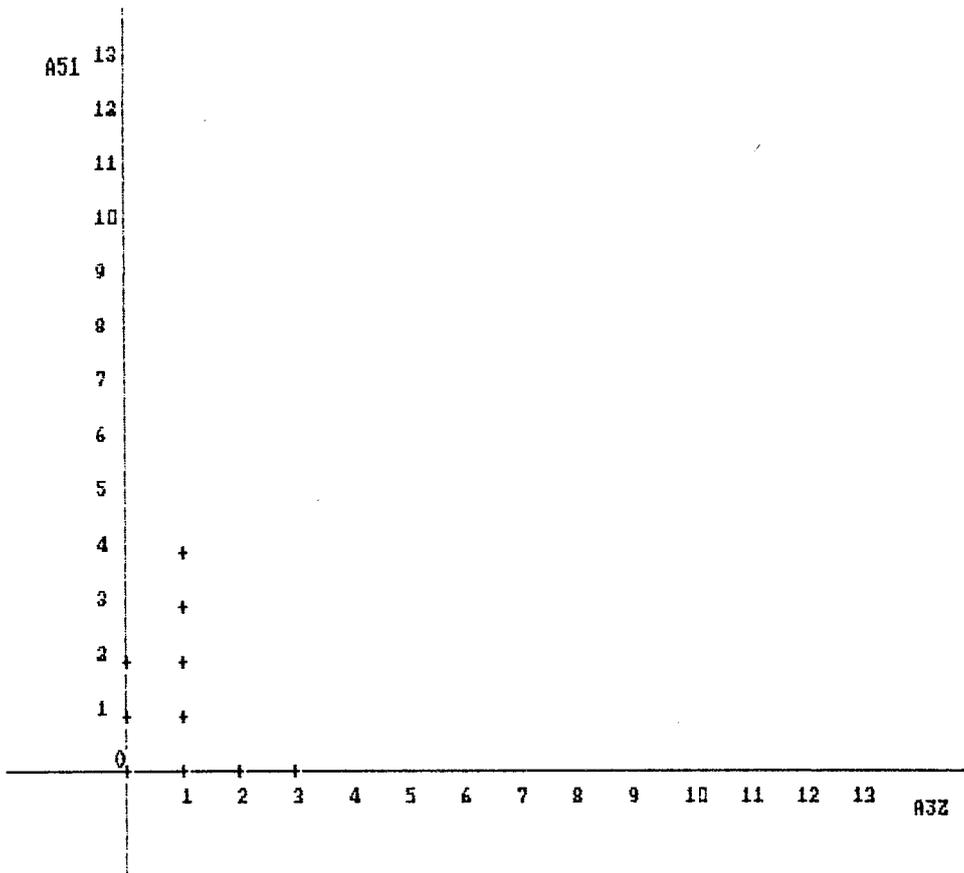
( 0, 1) : 11 49 53 58 66 70  
 ( 0, 2) : 14 31  
 ( 0, 4) : 71  
 ( 1, 0) : 25 26 36 37 60 61 79 87 89 91 99  
 ( 1, 1) : 18 29  
 ( 2, 0) : 1 2 28 35 86  
 ( 2, 1) : 6 9  
 ( 2, 3) : 7  
 ( 3, 0) : 19 20 45

nombre de syndromes à l'origine : 69

liste :

3	4	5	8	10	12	13	15	16	17	21	22	23	24	27
30	32	33	34	38	39	40	41	42	43	44	46	47	48	50
51	52	54	55	56	57	59	62	63	64	65	67	68	69	72
73	74	75	76	77	78	80	81	82	83	84	85	88	90	92
93	94	95	96	97	98	100	101	102						

PLAN 7



PLAN DES AXES A32 (abscisses) ET A51 (ordonnées)

POINTS DU PLAN OCCUPES et NUMEROS DE SYNDROMES

- ( 0, 1): 11 49 53 58 66 70 5 26 52 60 61 71 87
- ( 0, 2): 14 31 4
- ( 0, 4): 71
- ( 1, 0): 25 26 36 37 60 61 79 87 89 91 99 7 17  
18 21 22 27 29 33 36 37 39 41 42 68 74  
76 77 78 80 82 83 88 89 90 92 93 95 96  
99 100 102
- ( 1, 1): 18 29 11 28 75
- ( 1, 2): 25
- ( 1, 3): 14 32 53
- ( 1, 4): 38
- ( 2, 0): 1 2 28 35 86 79 81 101
- ( 2, 1): 6 9
- ( 2, 3): 7
- ( 3, 0): 19 20 45 91

nombre de syndromes à l'origine : 51

liste :

- 1 2 3 6 8 9 10 12 13 15 16 19 20 23 24
- 30 31 34 35 40 43 44 45 46 47 48 49 50 51 54
- 55 56 57 58 59 62 63 64 65 66 67 69 70 72 73
- 84 85 86 94 97 98

ANNEXES

ANNEXE 12

LISTE DES SYNDROMES

- 1 SYNDROME DE MARFAN
- 2 SYNDROME DU PTERYGIUM POPLITE
- 3 FUSION FILIFORME DES PAUPIERES
- 4 DYSOSTOSE MANDIBULO-FACIALE
- 5 SYNDROME DE WAARDENBURG-KLEIN
- 6 SYNDROME DE STICKLER
- 7 DYSOSTOSE CLEIDO-CRANIENNE
- 8 SYNDROME DE VAN DER WOUDE
- 9 NAEVOMATOSE BASO-CELLULAIRE
- 10 FENTE PALATINE, CAMPTODACTYLIE ET PIED BOT
- 11 SYNDROME D'APERT
- 12 NEVRITE RECIDIVANTE DU PLEXUS BRACHIAL ET FENTE PALATINE
- 13 HYPERTELORISME - HYPOSPADE
- 14 ECTRODACTYLIE
- 15 SYNDROME DE RAPP-HODGKIND
- 16 SYNDROME DE FONTAINE
- 17 DYSPLASIE OCULO-DENTO-OSSEUSE
- 18 SYNDROME DE ROBINOW
- 19 SYNDROME DE KNIEST
- 20 DYSPLASIE SPONDYLO-EPIPHYSIAIRE CONGENITALE
- 21 SYNDROME DE SAY
- 22 SYNDROME DE PALANT
- 23 SYNDROME DE LOWRY-MILLER
- 24 UVULE BIFIDE, SURDITE, SYNDROME NEPHROTIQUE ET ANOMALIES  
CONGENITALES DES VOIES URINAIRES ET DES DOIGTS
- 25 SYNDROME PSEUDO-THALIDOMIDE
- 26 SYNDROME DE MARDEN-WALKER
- 27 SYNDROME DE SMITH-LEMLI-OPITZ
- 28 CHONDRODYSPLASIE PONCTUEE TYPE RHIZOMELIQUE
- 29 SYNDROME DE MOHR
- 30 SYNDROME DE LARSEN
- 31 SYNDROME D'ELSAHY-WATERS
- 32 SYNDROME DE DUBOWITZ
- 33 SYNDROME DE MECKEL
- 34 ABLEPHARIE ET FENTE PALATINE
- 35 DYSPLASIE CAMPOMELIQUE
- 36 SYNDROME DE CHRISTIAN-ANDREWS-CONNELLY-MULLER
- 37 SYNDROME DE PTERYGION MULTIPLE
- 38 SYNDROME DE COFFIN-SIRIS
- 39 SYNDROME ORO-CRANIO-DIGITAL
- 40 NANISME MICROGNATHE
- 41 SYNDROME CEREBRO-COSTO-MANDIBULAIRE
- 42 NANISME DYSSEGMENTAIRE
- 43 SYNDROME DE AASE SMITH
- 44 SYNDROME DE LA CHAPELLE
- 45 NANISME DIASTROPHIQUE
- 46 SYNDROME DE BIRCH-JENSEN
- 47 SYNDROME DE RUDIGER
- 48 FENTE PALATINE ET TERATOME CONGENITAL DE LA REGION BUCCALE
- 49 SYNDROME DE HANHART
- 50 ACHRODYSOSTOSE
- 51 ENCEPHALOMENINGOCELE
- 52 DYSPLASIE FRONTONASALE

ANNEXES

53	SYNDROME DE CORNELIA DE LANGE
54	JUMEAUX THORACOPAGES ET FENTE LABIALE ET/OU PALATINE
55	MALADIES CARDIAQUES CONGENITALES ET FENTE LABIALE ET/OU PALATINE
56	FENTE LABIALE ET/OU PALATINE ET ANENCEPHALIE
57	SYNDROME DE GOLDENHAR
58	SYNDROME DE CHARLIE M
59	SYNDROME DE PIERRE ROBIN
60	SYNDROME PTERYGO-ACROVERTEBRAL
61	SYNDROME DE KLIPPEL-FEIL
62	SYNDROME DE HO
63	SYNDROME FEMOROFACIAL DE DAENTL
64	SYNDROME DE MAJEWSKI
65	FENTE PALATINE ET DUPLICATION ORALE
66	FENTE PALATINE, ANKYLOSE STAPEDOVESTIBULAIRE ET OLIGODONTIE
67	FENTE PALATINE ET NEUROBLASTOME CONGENITAL
68	SYNDROME DE BECWITH-WIEDEMANN
69	FENTE PALATINE, MICROGNATHIE, PIED BOT VARUS EQUIN, COMMUNICATION INTERAURICULAIRE ET PERSISTANCE DE LA VEINE CAVE SUPERIEUR GAUCHE
70	SYNDROME DE PALLISTER
71	SYNDROME ORO-DIGITO-FACIAL TYPE 1
72	SYNDROME OTO-PALATO-DIGITAL
73	FENTE PALATINE ET PETITE TAILLE
74	INTOXICATION FOETALE PAR L'AMINOP TERINE
75	INTOXICATION FOETALE PAR L'ALCOOL
76	INTOXICATION FOETALE A L'HYDANTOINE
77	SYNDROME 1 Q +
78	TRISOMIE 3 P
79	MONOSOMIE 4 P
80	MONOSOMIE 7 Q 3
81	TRISOMIE 7 2
82	TRISOMIE 7 P 2
83	MONOSOMIE 7 P 2
84	MONOSOMIE 10 P
85	TRISOMIE 8 Q 2
86	TRISOMIE 8 P
87	TRISOMIE 10 P
88	TRISOMIE 14
89	TRISOMIE 14 Q 1
90	TRISOMIE 17 Q 2
91	TRISOMIE 19 Q
92	CHROMOSOMIER R (18)
93	TRIPLOIDIE
94	TRISOMIE 7 Q 3
95	TRISOMIE 18
96	TRISOMIE 13
97	TRISOMIE 11 Q 2
98	TRISOMIE 11 P
99	TRISOMIE 10 Q 2
100	TRISOMIE 9
101	MONOSOMIE 4 Q 3
102	TRISOMIE 9 Q

## ANNEXES

## ANNEXE 13

GENERATION DES FICHIERS SYNDROME x ENTITE/SCORE POUR LES ACP,  
AUX 4 NIVEAUX D'OBSERVATION, nommés FACP1, FACP2, FACP3, FACP4

```

program GFICHACP;
{ génération des fichiers SYNDROME x ENTITE/SCORE pour les ACP,
  aux 4 niveaux d'observation, nommés FACP1, FACP2, FACP3, FACP4 }

const  MAXELEMFSYND = 8000;
       MAXELEMFSYMP = 2600;
       MAXSYND      = 102;
       MAXELEMFESE  = 14000; { 2600 + 2600 + 2 * ( 2600 - 102 * 2 ) }
       MAXAXES     = 200;

type
  ST8 = string [8];
  ESYMPTOME = record
    CAR_DB3 : char;
    FORMNAT: string[90]; NUMERO: string[4];
    ENTITSYM: string[45]; CODENTSYM: string[4];
    VALNAT: string[25]; VALNUM: string[3];
    NIBOBSAXE: string[ 4]; ATTRIBUT: string[46];
    PERE: string[4]; FILS: string[4];
    FRERE: string[4]; NIVOBS: char;
    NBDESYND: string[3]; ADLSTSYND: string[4];
    CHAPITRE : string [2]; ESMERE : string[82];
    AXE : string[57]; NUMAXE : string[3];
    TYPAXE : string [2]
  end;
  ESYNDROME = record
    LIBRE : char;
    NOMSYND : string[120];
    NUMERO : string[4];
    CODESYND : string[3];
    NBDESYMP : string[4];
    RENSEIGT : string[33];
    PREVALEN : string[1];
    ADLSTSYMP: string[4];
    ADLSTES : string [4]
  end;

  TYPTFSYMP= array [1..MAXELEMFSYMP] of integer;
  TYPTFES = array [1..MAXELEMFESE] of integer;
  TYPsyND = array [1..MAXSYND] of integer;

var
  NIVO, NBSYND, NOES, NBES, P, NOSYMP, NBSYMP, ADR, ADRES, I, J : integer;
  FES, FSYMP : file;
  RENS : text;
  FSP : file of ESYMPTOME;
  FSD : file of ESYNDROME;
  ESP : ESYMPTOME;
  ESD : ESYNDROME;
  TFSYMP : TYPTFSYMP;
  TFES : TYPTFES;
  TSYND : TYPsyND;
  QSTI : char;
  TROUVE : boolean;

```

ANNEXES

```

    CNIVO, NOMFACP : ST8;

function DECODENUM (CEL : integer): integer;
begin
    DECODENUM := - (CEL + 256)
end;

function DECODENOMBRE (CEL: integer): integer;
begin
    if CEL = - 255 then DECODENOMBRE := 0 else DECODENOMBRE := - CEL
end;

procedure DECODESY (CEL : integer; var N : integer; var QSTI : char);
const T : string [4] = 'QSTI';
begin
    N := CEL DIV 4;
    QSTI := T [CEL MOD 4 + 1];
end;

procedure GENFICHACP ( NOMFACP : ST8;
                      NIVO, NBSYND : integer; SYND : TYPsyND);
{ SYND : array [1..MAXSYND] of byte , contient les numéros ordonnés
  croissants des syndromes candidats }
var
    NBNONREP, NBNONREP1, RANG, K, A1, A2, K1, K2 , N, T, I : integer;
    TABSENT : array [ 1..MAXSYND ] of byte;
    ILYEST : boolean;
    NBAX : array [ 0..4 ] of byte;
    PREMAX : array [ 0..4 ] of byte;
    FACP : text;
    FSYMP, FES : file;
    ADFSYP, ADFES : TYPsyND;
    LIGNE : array [1..MAXAXES] of integer;
    POIDS : real;

begin
    NBAX [ 0 ] := 174;
    NBAX [ 1 ] := 67;
    NBAX [ 2 ] := 22;
    NBAX [ 3 ] := 32;
    NBAX [ 4 ] := 53;

    PREMAX [ 0 ] := 0 ;
    PREMAX [ 1 ] := 0 ;
    PREMAX [ 2 ] := PREMAX [ 1 ] + NBAX [ 1 ] ;
    PREMAX [ 3 ] := PREMAX [ 2 ] + NBAX [ 2 ] ;
    PREMAX [ 4 ] := PREMAX [ 3 ] + NBAX [ 3 ] ;

    NBNONREP := 0;
    assign (FSYMP, 'FSYMP.TOM'); reset (FSYMP, 2);
    assign (FES , 'FES'); reset (FES, 2);
    assign (FACP , NOMFACP ); rewrite ( FACP );
    blockread (FSYMP, TFSYMP, MAXELEMFSYMP);
    blockread (FES, TFES, MAXELEMFS);

    writeln (FACP, CNIVO);
    writeln (FACP, NBSYND );
    writeln (FACP, NBAX [ NIVO ] );

```

## ANNEXES

```

{ calcul des adresses des syndromes dans TFSYMP ET TFES }
K1 := 1;
K2 := 1;
for I := 1 to NBSYND do
begin
  {WRITELN (' K1 = ',K1:4,'          SYND [',I:3,'] = ',SYND [I],
            ',          ',DECODENUM ( TFSYMP [K1] ));}
  while DECODENUM ( TFSYMP [K1] ) < SYND [I] do
  begin
    K1 := K1 + 1 + DECODENOMBRE ( TFSYMP [K1+1] ) + 1;
    K2 := K2 + 1 + DECODENOMBRE ( TFES [ K2 + 1 ] ) * 5 + 1
  end;
  ADFSYP [ SYND [I] ] := K1;
  ADFES [ SYND [I] ] := K2;
  { recherche si le syndrome I est représenté au niveau NIVO }
  ILYEST := false; K := 1;
  NBSYMP := DECODENOMBRE ( TFSYMP [ K1 + 1 ] );
  while not ILYEST and ( K <= NBSYMP ) do
  begin ILYEST := ( NIVO = 0 ) or ( TFES [ K2 + 1 + 5 * K - 1 ] = NIVO )
        K := K + 1
  end;
  if not ILYEST then begin NBNONREP := NBNONREP + 1;
                        TABSENT [ NBNONREP ] := I
                    end;
end;

writeln (FACP, NBNONREP );
for I := 1 to NBNONREP do writeln (FACP, TABSENT [ I ] );

NBNONREP1 := 0;
writeln (rvson,' SYNDROMES NON REPRESENTES AU NIVEAU ', CNIVO,' : ',rvsoff
for I := 1 to NBSYND do
begin
  {WRITELN ('          I : ', I);}
  for J := 1 to NBAX [ NIVO ] do LIGNE [J] := 0;
  A1 := ADFSYP [ SYND [I] ];
  A2 := ADFES [ SYND [I] ];
  NBSYMP := DECODENOMBRE ( TFSYMP [ A1 + 1 ] );
  N := 0; T := 0;
  {WRITELN ('NBSYMP = ',NBSYMP);}
  ILYEST := false;
  for K := 1 to NBSYMP do
  begin
    if ( NIVO = 0 ) or ( TFES [ A2 + 1 + 5 * K - 1 ] = NIVO )
    then begin
      if ( PREMAX [NIVO] >= TFES [ A2 + 1 + 5 * K - 4 ] )
      or ( PREMAX [NIVO] + NBAX [NIVO] < TFES [ A2 + 1 + 5 * K - 4 ] )
      then WRITELN ('ERREUR syndrome n° ', I,' axe : ',
                    TFES [ A2 + 1 + 5 * K - 4 ], ' Niv. obs. : ',
                    TFES [ A2 + 1 + 5 * K - 1 ])
    end
  else
  begin RANG := TFES [ A2 + 1 + 5 * K - 4 ] - PREMAX [ NIVO ];
        {WRITELN ('          RANG : ', RANG );}
        if TFES [ A2 + 1 + 5 * K ] = 1
        then LIGNE [ RANG ] := TFES [ A2 + 1 + 5 * K - 3 ]
        else LIGNE [ RANG ] := LIGNE [ RANG ]
            + abs (TFES [ A2 + 1 + 5 * K - 3 ] );
  end;
end;

```

## ANNEXES

```

      (WRITELN ( ' TFES [ A2 + 1 + 5 * K - 2 ] : ',
              TFES [ A2 + 1 + 5 * K - 2 ] );)
    case TFES [ A2 + 1 + 5 * K - 2 ] of
      1 : T := T + 1;
      2 : N := N + 1
    end
  end;
  ILYEST := true
end
end;

if ILYEST
then begin
  for J := 1 to NBAX [ NIVO ] do write (FACP, LIGNE [J], ' ');
  { calcul du poids }
  (WRITELN ( ' I : ', I:3, ' T = ', T:5, ' N = ', N);)
  POIDS := T / ( N + T );
  write ( FACP, POIDS, ' ');
  end
else begin NBNONREP1 := NBNONREP1 + 1;
  writeln ( I:25)
end;

end;
if NBNONREP <> NBNONREP1
then writeln ( rvson, ' ERREUR DE DECOMPTE DE NON-REPRESENTES ', rvsoff,
              ' NBNONREP = ', NBNONREP, ' NBNONREP1 = ', NBNONREP1);

if NBNONREP = 0 then writeln ( 'Tous représentés');
close ( FACP )
end ( GENFICHACP );

begin ( MAIN PROGRAM )
(*)
  writeln (clrhom);
  ( NOM DU FICHIER A CREER DANS RENSACP );
  assign ( RENS, 'RENSACP'); reset ( RENS );
  readln ( RENS, S);
  WRITELN ( S );
  readln ( RENS, NIVO);
  WRITELN ( 'NIVEAU : ', NIVO );
  readln ( RENS, NBSYND);
  WRITELN ( 'NBSYND : ', NBSYND );
  if NBSYND > MAXSYND then
  begin NBSYND := 102;
    for I := 1 to MAXSYND do TSYND [I] := I
  end
  else for I := 1 to NBSYND do read ( RENS, TSYND [I] );
*)
  NBSYND := 102;
  for I := 1 to MAXSYND do TSYND [I] := I;
  for NIVO := 0 to 4 do
  begin
    str ( NIVO:1, CNIVO );
    NOMFACP := concat ( 'FACP', CNIVO);
    GENFICHACP ( NOMFACP, NIVO, NBSYND, TSYND )
  end
end.

```

## ANNEXES

## ANNEXE 14

## PROGRAMME DE CALCUL DE MATRICES DE VARIANCES-COVARIANCES, DE VALEURS PROPRES ET DE VECTEURS PROPRES

```

program J;
( programme de calcul de vecteurs et valeurs propres par niveau,
  destiné en premier lieu à la pré-étude statique du domaine.
  Calcule la matrice V de variance-covariance à partir d'un fichier
  individus x caractères nommé FACP<niv>, créé par le programme GFICHACP,
  et la range dans le fichier MATCOV<niv>.
  ACP par l'algo. de JACOBI.
  Sauvegarde val. et vect. propres dans SAUVACP<niv>,
  création du listing de sortie dans LISTACP<niv>.
)

const
  MAXINDIV = 102;
  MAXVAR   = 67;

type
  TYPVECTEUR = array [1..MAXVAR] of real;
  TYPV       = array [1..MAXVAR] of TYPVECTEUR;
  {TYPTABCOORD= array [1..MAXINDIV] of array [1..2] of real;}
  CHAINE     = string [10];
  TYPTABSENT = array [ 1..MAXINDIV ] of integer;

var
  D           : TYPVECTEUR at $3000:$0000;
  EIVVEC, FIN : boolean;
  NBVEC, I, NBVAR, NBIND, NBNONREP, J, K, K1, K2: integer;
  DE, NU, PCENT : real;
  {COORD       : TYPTABCOORD;}
  A : TYPV at $4000:$0000;
  LIST, FV, SV : text;
  REP, R : char;
  C, C1, C2 : CHAINE;
  TABSENT : TYPTABSENT at $3700:$0000;
  NIVO : char;

procedure CALCULV ( NIVO : char; var NBIND, NBVAR, NBNONREP : integer;
                   var TABSENT : TYPTABSENT; var V : TYPV);
( calcule la matrice V de var. covar. à partir du fichier de nom (C)
  contenant X )
(type
  TYPX = array [1..MAXINDIV, 1..MAXVAR ] of real;
)
var NBREP, N, N2, I, J, K : integer;
(   SX,SY,SXY,SX2,SY2 : real;
  COLONNE : array [ 1..MAXVAR ] of integer;
  POIDS : real;
  S : array [1..MAXVAR ] of integer;
  F, FV : text;
)
(   S, S2 : array [ 1.. MAXVAR ] of real;
  C : string [10];
)
begin
( write (' NOM DU FICHIER DES INDIVIDUS ? : '); readln (C);)

```

## ANNEXES

```

C := concat ('FACP', NIVO);
writeln;
writeln (rvson, ' lecture du fichier ', C , rvsoff);
assign ( F, C ); reset ( F );
readln (F, NIVO ); writeln ( 'NIVEAU D'OBSERVATION : ', NIVO );
readln (F, NBIND); writeln ( 'NBIND : ', NBIND);
readln (F, NBVAR); writeln ( 'NBVAR : ', NBVAR);
readln (F, NBNONREP); writeln ( 'NBNONREP : ', NBNONREP);
for K := 1 to NBNONREP do readln (F, TABSENT [ K ]);
NBREP := NBIND - NBNONREP;
{ for K := 1 to NBREP Do
    for I := 1 TO NBVAR do read (F, X [ K, I ] );}

writeln
(' Calcul de la matrice de variance-covariance; veuillez patienter');
writeln (' quelques minutes. ');
N := NBREP;
N2 := N * N ;
for I := 1 to NBVAR do for J := I to NBVAR do V [I, J] := 0.0;
for I := 1 to NBVAR do S [I] := 0;

for K := 1 to N do
begin
  {WRITELN (' K : ',k);}
  for I := 1 to NBVAR do
  begin read (F, COLONNE [I] );
    {WRITELN (' I : ',I, ' COLONNE [I]');}
    S [I] := S [I] + COLONNE [I]
  end;
  read ( F, POIDS );
  WRITELN (POIDS:8:5);
  for I := 1 to NBVAR do
    for J := I to NBVAR do V [I, J] := V [I, J] + COLONNE [I] * COLONNE [J];
end;

for I := 1 to NBVAR do
  for J := I to NBVAR do V [I, J] := V [I, J] / N - S [I] * S [J] / N2 ;

for I := 1 to NBVAR - 1 do
  for J := I + 1 to NBVAR do V [J, I] := V [I, J];

{ sauvegarde de V }
C := concat ( 'MATCOV', NIVO );
writeln;
WRITELN (' SAUVEGARDE MATRICE VAR-COVAR DANS ', C );
assign ( FV , C ); rewrite (FV);
writeln ( FV , NBVAR);
for I := 1 to NBVAR do
  for J := 1 to NBVAR do write ( FV, ' ', V [ I, J ] );
close ( FV )

end (CALCULV) ;

procedure JACOBI (N: integer; var A: TYPV;
                  EIVC: BOOLEAN ; var D:TYPVECTEUR);
{ N : nombre d'individus
  V : matrice des corrélations, transformée en matrice de vecteurs propres
  normés

```

ANNEXES

```

EIVEC : booléen pour calcul des vecteurs propres
D : vecteur des valeurs propres
}
{ procedure non transcrite, appartenant à la bibliothèque B.A.N.G. }

end { JACOBI } ;

procedure LECTURMAT ( NIVO : char; var NBIND, NBVAR, NBNONREP : integer;
                    var TABSENT : TYPTABSENT; var A : TYPV);
{ lecture de la matrice de variance-covariance sauvegardée par CALCULV }
var C : string [10];
    F, FV : text;
    I, J : integer;

begin
    { lecture de V }
    C := concat ('FACP', NIVO);
    writeln ( rvson, ' lecture du fichier ', C , rvsoff);
    assign ( F, C ); reset ( F );
    readln ( F, NIVO ); writeln ( 'NIVEAU D'OBSERVATION : ', NIVO );
    readln ( F, NBIND); writeln ( 'NBIND : ', NBIND);
    readln ( F, NBVAR); writeln ( 'NBVAR : ', NBVAR);
    readln ( F, NBNONREP); writeln ( 'NBNONREP : ', NBNONREP);
    for K := 1 to NBNONREP do readln ( F, TABSENT [ K ] );
    close ( F );

    C := concat ('MATCOV', NIVO);
    writeln ( rvson, ' lecture du fichier ', C , rvsoff);
    assign ( FV , C ); reset ( FV );
    readln ( FV , NBVAR);

    for I := 1 to NBVAR do
        for J := 1 to NBVAR do read ( FV, A [ I, J ] );
    end;

    close ( FV )
end { LECTURMAT };

procedure ECRITURVP ( NIVO : char; var NBIND, NBVAR, NBNONREP : integer;
                    var TABSENT : TYPTABSENT; A : TYPV );
var SV : text;
    C : string [10];
    I, J, K : integer;

begin
    C := concat ('SAUVACP', NIVO);
    writeln ( rvson, ' Sauvegarde du fichier ', C , rvsoff);
    assign ( SV , C ); rewrite ( SV );
    writeln ( SV , NIVO);
    writeln ( SV , NBIND);
    writeln ( SV , NBVAR);
    writeln ( SV , NBNONREP);
    for K := 1 to NBNONREP do writeln ( SV , TABSENT [ K ] );

    for I := 1 to NBVAR do write ( SV, ' ', D [ I ] );
    for I := 1 to NBVAR do
        for J := 1 to NBVAR do write ( SV, ' ', A [ I, J ] );
    end;
    close ( SV )
end { ECRITURVP };

```

## ANNEXES

```

procedure LISTING ( NIVO : char) ;
var P, MAX, NBVEC, I, J, K : integer;
    DE, NU, PCENT : real;
    SV, LIST : text;
    C, CJ : string [10];
    NBAXE, PREMAXE : array [0..4] of byte;
    NIVEAU : byte;

begin
  ( lecture fichier des valeurs et vecteurs propres )
  C := concat ('SAUVACP', NIVO);
  writeln (' LISTING ');
  writeln ( rvson, ' Lecture du fichier ', C , rvsoff);
  assign ( SV , C ); reset (SV);
  readln ( SV , NIVO);
  readln ( SV , NBIND);
  readln ( SV , NBVAR);
  readln ( SV , NBNONREP);
  for K := 1 to NBNONREP do readln ( SV , TABSENT [ K J]);

  for I := 1 to NBVAR do read ( SV, D [I]);
  for I := 1 to NBVAR do
    for J := 1 to NBVAR do read ( SV, A [ I, J ] );
  close ( SV );

  NBAXE [0] := 174;
  NBAXE [1] := 67;
  NBAXE [2] := 22;
  NBAXE [3] := 32;

  PREMAXE [1] := 0 ;
  PREMAXE [2] := PREMAXE [1] + NBAXE [1] ;
  PREMAXE [3] := PREMAXE [2] + NBAXE [2] ;
  PREMAXE [4] := PREMAXE [3] + NBAXE [3] ;

  val ( NIVO, NIVEAU, P);

  C := concat ('LISTACP', NIVO );
  assign ( LIST, C); rewrite ( LIST );
  DE := 0;
  for I := 1 to NBVAR do DE := DE + D [I];
  NU := 0;
  writeln ( LIST);
  writeln ( LIST, '
                                     VALEURS ET VECTEURS PROPRES');
  writeln ( LIST);
  writeln ( LIST, ' NOMBRE DE VARIABLES : ', NBVAR);
  writeln ( LIST);
  writeln ( LIST, ' Niveau d'observation : ', NIVO);
  writeln ( LIST, ' Nombre de syndromes total : ', NBIND);
  writeln ( LIST, ' Nombre de syndromes non représentés à ce niveau : ',
          NBNONREP);
  writeln ( LIST);
  writeln (LIST, ' VALEURS PROPRES :          % INERTIE :          % INERTIE CUMULEE
  for I := 1 TO NBVAR do
  begin
    NU := NU + D [I];
    PCENT := (NU / DE) * 100;

```

## ANNEXES

```

      writeln (LIST, ' D [', I:2, ' ] = ', D [I]:6:2,
              ', (D [I]/DE)*100 :6:2 ,
              ', PCENT :6:2 )
    end;
  writeln ( LIST );

  if NBVAR < 16 then NBVEC := NBVAR else NBVEC := 16;
  writeln (LIST, NBVEC : 2, ' premiers vecteurs propres : ' );
  writeln ( LIST );
  write ( LIST, 'AXE3 ');
  for J := 1 to NBVEC do
  begin
    str ( J:2, CJ);
    if (len (CJ) = 2) and (CJ [1] = ' ') then CJ := CJ [2];
    write ( LIST, ' V', CJ, ' ');
    if len (CJ) = 1 then write ( LIST, ' ');
    write (LIST, ' 3 ');
  end;
  writeln ( LIST );

  for I := 1 to NBVAR do
  begin write (LIST, I + PREMAXE [ NIVEAU] :3, ' 3 ');
        for J := 1 to NBVEC do write (LIST, A [I, J]:5:2, ' 3 ');
        writeln ( LIST )
  end;

  writeln ( LIST );
  writeln ( LIST );
  writeln ( LIST, ' Vecteurs suivants (jusqu''à 16 en plus) : ');
  writeln ( LIST );
  if NBVAR > 16 then
  begin write ( LIST, 'AXE3 ');
        if NBVAR > 32 then MAX := 32 else MAX := NBVAR;
        for J := 17 to MAX do
        begin
          str ( J:2, CJ);
          if (len (CJ) = 2) and (CJ [1] = ' ') then CJ := CJ [2];
          write ( LIST, ' V', CJ, ' ');
          if len (CJ) = 1 then write ( LIST, ' ');
          write (LIST, ' 3 ');
        end;
        writeln ( LIST );

        for I := 1 to NBVAR do
        begin write (LIST, I + PREMAXE [ NIVEAU] :3, ' 3 ');
              for J := 17 to MAX do write (LIST, A [I, J]:5:2, ' 3 ');
              writeln ( LIST )
        end;
        writeln ( LIST );
        writeln ( LIST );
  end;
  close (LIST)
end { LISTING };

begin                                     { MAIN PROGRAM }
  write (clrhom);

  for NIVO := '1' to '4' do

```

## ANNEXES

```
begin
  { write ( ' Voulez-vous toute l''E(TUDE ou seulement le L(ISTING ? : ' );
    read ( TRM, R ); }
  R := 'E';
  if R in [ 'E', 'e' ] then
  begin
    writeln;
    {write ( 'MATRICE DE COVARIANCES PRETE ? (O/N) : ' ); read ( trm,R ); writeln
    R := 'N';
    if R in [ 'N', 'n' ] then CALCULV ( NIVO, NBIND, NBVAR, NBNONREP, TABSENT, A
      else LECTURMAT( NIVO, NBIND, NBVAR, NBNONREP, TABSENT, A

    { affichage matrice A de corrélation );
    JACOBI ( NBVAR, A, true, D );
    ECRITURVP ( NIVO, NBIND, NBVAR, NBNONREP, TABSENT, A );
  end;
  LISTING ( NIVO );
end
end.
```

## ANNEXE 15

## REPRESENTATION CARTESIENNE DE SYNDROMES SUR UN PLAN

```

program PROJEC ;
{ trace LES PROJECTIONS DES SYNDROMES SUR DEUX AXES QUELCONQUES }

{#I GRAPH_0L}
const
    MAXINDIV = 102;
    MAXVAR   = 67;
type
    TYPAXES = array [1..MAXINDIV] of record A1, A2: integer end;
    TYPTABCOORD= array [1..MAXINDIV] of array [1..2] of integer ;
    TYPOINT = array [-2..13,-2..13] of record NB : integer;
                                                SYND : array [1..80] of byte
                                                end;
    CHAINE = string [10];
var
    TABSENT : array [1..MAXINDIV] of integer at $3700:$0000;
    MAX1, MAX2, AXE1, AXE2, I, NBVAR, NBIND, NBNONREP,
    L, J, K, K1, K2 : integer;
    COLON : array [1..174] of integer;
    POIDS : real;
    COORD : TYPTABCOORD;
    PTS : TYPOINT at $5000:$0000;
    ORI : array [1..102] of integer;
    SV, PO : text;
    R , NIVO : char;
    C : CHAINE;
    C1, C2 : string [40];

procedure VISU (CO : TYPTABCOORD ; NBVAR : integer);
{ O1, O2 : abscisse et ordonnée du centre 0 }
var U1, U2, R1, R2, I, O1, O2, X, Y: integer;
    SP, S, S1 : string [40];

begin
    writeln;
    { R1 := 580; R2 := 380; }
    U1 := 37; U2 := 24;
    O1 := 79; O2 := 53;
    INEGRA (7, 0, 1, 3, 3 ); { initialisation de l'écran graphique }
    PLOT ( 0 , O2 ); { axe horizontal }
    DRAW ( 639, O2 );
    PLOT ( 600, O2 - 20 ); WRITEGR ( C1, 1);
    PLOT ( O1 , 0); { axe vertical }
    DRAW ( O1, 399 );
    PLOT ( O1 - 50 , 360 ); WRITEGR ( C2, 1);
    PLOT (O1-7, O2+2); S := '0'; WRITEGR ( S, 1);
    for I := 1 to 13 do
    begin str (I,S);
        PLOT (O1+ I*U1 , O2-15 );
        WRITEGR (S,1);
    end
end

```

## ANNEXES

```

        PLOT (01-18 ,02 + I*U2 );
        WRITEGR (S,1)
    end;

( PLOT (1, 360); S := 'Nombre de variables : '; WRITEGR (S, 1);
  str (NBVAR, S); WRITEGR ( S, 1);
  SP := '+';
  for I := 1 to NBVAR do
  begin X := U1 * CO [I, 1];
        Y := U2 * CO [I, 2];
        {str ( X, S); str (Y, S1);
        S :=
        PLOT (1, 360 - 9 * I); WRITEGR (S, 1); }
        PLOT ( X + 01 - 2, Y + 02 - 4);
        str ( I:2, S) ; WRITEGR (SP, 1); {WRITEGR ( S, 1 )}
  end;
  REDEFAN;
end ( VISU );

begin
    { MAIN PROGRAM }
    write (clrhom);

    for I := -2 to 13 do for j := -2 to 13 do PTS [ I, J ].NB := 0;
    NBNONREP := 0;

    repeat
        { demande des axes }
        write (' NUMERO DU PREMIER AXE ? : '); readln ( AXE1 );
        write (' NUMERO DU SECOND AXE ? : '); readln ( AXE2 );
        str (AXE1, C1); str ( AXE2, C2);
        C1 := concat ('A',C1);
        C2 := concat ('A',C2);

        { lecture fichier matrice des syndromes }
        C := 'FACPO';
        writeln ( rvson, ' Lecture du fichier ', C , rvsoff);
        assign ( SV , C ); reset (SV);
        readln ( SV, NIVO );
        readln ( SV , NBIND);
        readln ( SV , NBVAR);
        readln ( SV, NBNONREP );
        for I := 1 to NBIND do
            begin
                for J := 1 to NBVAR do read ( SV, COLON [J] );
                read ( SV, POIDS );
                {WRITELN ('SYND. ',I:3,' poids : ', POIDS);}
                COORD [I, 1 ] := COLON [ AXE1 ];
                COORD [I, 2 ] := COLON [ AXE2 ];
                if (COLON [AXE1]=0)and (COLON [AXE2]=0)
                then begin NBNONREP := NBNONREP + 1;
                           ORI [ NBNONREP ] := I
                end
                else
            begin PTS [ COLON [AXE1], COLON [AXE2] ].NB :=
                    PTS [ COLON [AXE1], COLON [AXE2] ].NB + 1;
            PTS [COLON [AXE1],COLON[AXE2] ]
                .SYND [ PTS [COLON[AXE1], COLON[AXE2]].NB ] := I;
            end;
        end;
    end;

```

## ANNEXES

```

      (WRITELN ( COLON [ AXE1 ]:5,' ', COLON [AXE2 ]))
    end;
    close ( SV );

VISU ( COORD, NBIND);
C := concat (C1,C2);
assign (P0, C ); rewrite (P0);
writeln (P0,'PLAN DES AXES A',AXE1,' (abscisses) ET A',AXE2,
        ' (ordonnées)');
writeln (P0);
writeln (P0,'POINTS DU PLAN OCCUPES et NUMEROS. DE SYNDROMES');

for I := -2 TO 13 do
  for J := -2 to 13 do
    if (I<>0) or (J<>0)
    then
      if PTS [ I, J ].NB > 0
      then
        begin L := 0;
          write ( P0,' (' ,I:2,',',J:2,') : ');
          for K := 1 to PTS [ I, J ].NB do
            begin write ( P0, ' ',PTS [ I, J ].SYND [K]);
              L := L + 1;
              if L = 13 then begin L := 0; writeln (P0);
                write (P0, ' ' )
              end;
            end;
          writeln ( P0 );
        end;
      end;
    writeln (P0);
    writeln (P0, ' nombre de syndromes à l'origine :',NBNONREP);
    writeln (P0, ' liste : ');
    J := 1;
    for I := 1 to NBNONREP do
      begin write (P0,' ',ORI [I]:3); J := J + 1;
        if J=16 then begin J := 1; writeln (P0) end
      end;
    writeln (P0);
    writeln (P0);
    writeln (P0);

close (P0);

write (' Une autre projection ? (O/N) : '); read (TRM, R);
writeln
until not ( R in ['0', 'o']);

end.

```

ANNEXES

ANNEXE 16

RECHERCHE DES AXES PERTINENTS ET DE LEURS VALEURS

```
program LDISC;
{ recherche des axes pertinents et de leurs valeurs }
const MAXVAR = 70;
      MAXSYND = 102;

type
  CELLULE = record NUMAX : byte;
                  VAL : real
            end;

  TYPVECT = array [1..MAXVAR] of CELLULE;
  TYPVECTEUR = array [ 1..MAXVAR] of real;
  TYPV = array [1..MAXVAR] of TYPVECTEUR;
  TYPABSENT = array [1..MAXSYND] of integer;

var
  D : TYPVECTEUR at $3000:$0000;
  A : TYPV at $4000:$0000;
  TABSENT : TYPABSENT at $3700:$0000;
  E, S : real;
  DD, NBIND, NBVAR, NBNONREP, NIVO, MAX, I, J, K, L : integer;
  SV, F : text;
  C : string [10] ;
  V : TYPVECT ;
  AUX : CELLULE ;
  NBAXE, PREMAXE : array [0..4] of byte;
  NIVEAU : char;

begin
  NBAXE [0] := 174;
  NBAXE [1] := 67;
  NBAXE [2] := 32;
  NBAXE [3] := 32;

  PREMAXE [0] := 0;
  PREMAXE [1] := 0;
  PREMAXE [2] := PREMAXE [1] + NBAXE [1];
  PREMAXE [3] := PREMAXE [2] + NBAXE [2];
  PREMAXE [4] := PREMAXE [3] + NBAXE [3];

  for NIVEAU := '1' to '4' do
  begin
    { lecture fichier des valeurs et vecteurs propres }
    C := concat ('SAUVACP', NIVEAU);
    writeln ( rvson, ' Lecture du fichier ', C , rvsoff);
    assign ( SV , C ); reset (SV);
    readln ( SV , NIVO);
    readln ( SV , NBIND);
    readln ( SV , NBVAR);
    readln ( SV , NBNONREP);
    for K := 1 to NBNONREP do readln ( SV , TABSENT [ K J]);
  end;
end;
```

## ANNEXES

```

for I := 1 to NBVAR do (WRITELN (' I = ',I)); read ( SV, D [I] );
for I := 1 to NBVAR do
  for J := 1 to NBVAR do read ( SV, A [ I, J ] );
close ( SV );

C := concat ('AD', NIVEAU);
assign ( F, C); rewrite ( F );

if NBVAR > 16 then MAX := 16 else MAX := NBVAR;

for I := 1 to MAX do
begin
  writeln (rvson, ' axe ', I:3, rvsoff);
  for J := 1 to NBVAR do begin V [J].NUMAX := PREMAXE [NIVO] + J;
                               V [J].VAL := A [ J, I]
                               end;
  ( TRI BULLE DE V SUR V.VAL )
  writeln (' tri ');
  K := NBVAR ;
  while K >= 2 do
  begin DD := 1 ;
        for L := 1 to K - 1 do
          if V [L].VAL < V [L+1].VAL then begin AUX := V [L+1];
                                             V [L+1] := V [L];
                                             V [L] := AUX;
                                             DD := L
                                             end;

          K := DD
        end;
  end;

( RECHERCHE DES MEILLEURS AXES PERTINENTS ET DES COEFFICIENTS
  AU SEUIL E )
  writeln (' recherche des meilleurs axes ');
  J := 0;
  E := 0.85;
  S := 0.0;
  while (S < E) and ( J < 10 ) do
  begin
    J := J + 1 ;
    S := S + V [J].VAL * V [J].VAL
  end;

  ( ECRITURE SUR LE FICHER )
  writeln (' écriture sur fichier ', C);
  writeln (F);
  writeln (F, ' VECTEUR ', I : 2);
  writeln (F, ' nombre d'axes ', J : 2);
  writeln (F);
  for L := 1 to J do writeln (F, V [L].NUMAX : 3, ' ', V [L].VAL:6:2);
end;

  writeln (F);
  writeln (F);
  writeln (F);

close (F)
end.
end.

```

ANNEXES

ANNEXE 17

ENSEMBLES DICHOTOMIQUES

E11 EFFECTIF : 45

3	5	6	7	8	12	13	14	15	16	17	18	23	24	31	34	35	40
41	42	46	48	49	51	54	55	56	59	61	63	64	65	66	67	68	73
75	81	84	88	90	91	97	98	99									

E12 EFFECTIF 57

1	2	4	9	10	11	19	20	21	22	25	26	27	28	29	30	32	33
36	37	38	39	43	44	45	47	50	52	53	57	58	60	62	69	70	71
72	74	76	77	78	79	80	82	83	85	86	87	89	92	93	94	95	96
100	101	102															

E21 EFFECTIF 73

1	3	6	7	8	9	10	12	13	15	19	20	21	22	23	24	26	30
31	32	35	36	38	40	41	42	43	44	45	47	48	50	51	54	55	56
57	60	61	62	63	64	65	66	67	68	69	70	73	75	76	77	78	79
81	82	83	84	85	86	87	88	89	91	92	94	95	96	97	98	99	100
102																	

E22 EFFECTIF : 29

2	4	5	11	14	16	17	18	25	27	28	29	33	34	37	39	46	49
52	53	58	59	71	72	74	80	90	93	101							

E31 EFFECTIF : 44

1	2	8	10	13	14	15	16	17	18	20	21	23	24	33	35	39	40
41	42	43	46	47	48	49	51	54	55	56	59	60	61	62	64	65	66
67	68	73	74	86	92	93	102										

E32 EFFECTIF : 58

3	4	5	6	7	9	11	12	19	22	25	26	27	28	29	30	31	32
34	36	37	38	44	45	50	52	53	57	58	63	69	70	71	72	75	76
77	78	79	80	81	82	83	84	85	87	88	89	90	91	94	95	96	97
98	99	100	101														

E41 EFFECTIF : 40

2	3	8	10	13	14	15	17	20	21	22	23	24	25	29	34	36	38
39	43	46	48	50	51	54	55	56	58	64	65	66	67	70	72	74	75
76	80	83	98														

ANNEXES

E42 EFFECTIF : 62

1	4	5	6	7	9	11	12	16	18	19	26	27	28	30	31	32	33
35	37	40	41	42	44	45	47	49	52	53	57	59	60	61	62	63	68
69	71	73	77	78	79	81	82	84	85	86	87	88	89	90	91	92	93
94	95	96	97	99	100	101	102										

E51 EFFECTIF : 23

4	6	16	20	21	23	24	40	41	42	43	47	48	59	60	61	62	69
73	74	75	97	102													

E52 EFFECTIF : 79

1	2	3	5	7	8	9	10	11	12	13	14	15	17	18	19	22	25
26	27	28	29	30	31	32	33	34	35	36	37	38	39	44	45	46	49
50	51	52	53	54	55	56	57	58	63	64	65	66	67	68	70	71	72
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93
94	95	96	98	99	100	101											

E61 EFFECTIF : 69

3	4	5	8	10	12	13	15	16	17	21	22	23	24	27	30	32	33
34	38	39	40	41	42	43	44	46	47	48	50	51	52	54	55	56	57
59	62	63	64	65	67	68	69	72	73	74	75	76	77	78	80	81	82
83	84	85	88	90	92	93	94	95	96	97	98	100	101	102			

E62 EFFECTIF : 33

1	2	6	7	9	11	14	18	19	20	25	26	28	29	31	35	36	37
45	49	53	58	60	61	66	70	71	79	86	87	89	91	99			

E71 EFFECTIF : 51

1	2	3	6	8	9	10	12	13	15	16	19	20	23	24	30	31	34
35	40	43	44	45	46	47	48	49	50	51	54	55	56	57	58	59	62
63	64	65	66	67	69	70	72	73	84	85	86	94	97	98			

E72 EFFECTIF : 51

4	5	7	11	14	17	18	21	22	25	26	27	28	29	32	33	36	37
38	39	41	42	52	53	60	61	68	71	74	75	76	77	78	79	80	81
82	83	87	88	89	90	91	92	93	95	96	99	100	101	102			

ANNEXES

ANNEXE 18

LISTE DES SYMPTOMES PORTES PAR LES AXES SELECTIONNES AU NIVEAU 1

(le numéro situé à gauche de l'expression naturelle est le code interne du symptôme)

AXE 60

SCORE VI.F FORME DES MEMBRES

432 INCURVATION DES MEMBRES INFERIEURS  
376 CLINODACTYLIE DU 5  
79 GROS ORTEIL LARGE  
435 GROS ORTEIL EN SPATULE  
32 PIED BOT CONVEXE  
33 PIED BOT VARUS  
34 PIED BOT VALGUS  
35 PIED BOT EQUIN  
36 PIED CREUX  
38 PIED PLAT  
39 PIED PLAT VALGUS  
309 PIED BOT  
562 PIED BOT TALUS  
551 CAMPTODACTYLIE DES ORTEILS  
434 POUCE EN SPATULE  
546 POUCE DIGITALISE  
622 POUCE EN ADDUCTION  
676 POUCE LARGE  
30 ARACHNODACTYLIE  
223 CAMPTODACTYLIE  
296 DOIGTS EN FUSEAU  
358 CLINODACTYLIE  
422 CLINODACTYLIE MULTIPLE DES DOIGTS  
37 MAIN BOTE EN FLEXION  
381 MAIN BOTE  
535 MAIN BOTE CUBITALE

AXE 63

SCORE VI.I BRACHYDACTYLIE

205 BRACHYMETACARPIE  
24 BRACHYDACTYLIE COMPLETE DES ORTEILS  
423 BRACHYDACTYLIE DES ORTEILS  
29 BRACHYDACTYLIE COMPLETE DES DOIGTS  
31 BRACHYDACTYLIE DES DOIGTS  
231 BRACHYDACTYLIE TYPE D  
314 BRACHYDACTYLIE TYPE B  
520 BRACHYDACTYLIE DES MAINS DE TYPE A

AXE 59

SCORE VI.D AMPUTATION DES MEMBRES

443 AMPUTATION RHIZOMELIQUE DES MEMBRES  
445 AMPUTATION DES MEMBRES

ANNEXES

475 PHOCOMELIE  
519 AMPUTATION ASYMETRIQUE DES MEMBRES  
49 AMPUTATION DES PHALANGES  
268 ECTRODACTYLIE  
515 AMPUTATION TOTALE PHALANGIENNE DES DOIGTS  
516 AMPUTATION CENTRALE DES PHALANGES DES DOIGTS  
287 AMPUTATION MESOMELIQUE PARTIELLE DU MEMBRE SUPERIEUR  
487 APLASIE DE L'AVANT BRAS  
639 AMPUTATION CONGENITALE DES MEMBRES SUPERIEURS  
441 AMPUTATION MESOMELIQUE PARTIELLE DU MEMBRE INFERIEUR  
517 AMPUTATION TOTALE DES PHALANGES DES ORTEILS  
518 AMPUTATION CENTRALE DES PHALANGES DES ORTEILS  
545 AMPUTATION PARTIELLE DES ORTEILS  
442 AMPUTATION METACARPIENNE

AXE 62  
SCORE VI.H SYNDACTYLIE

26 SYNDACTYLIE DES DOIGTS  
277 SYNDACTYLIE PARTIELLE DES DOIGTS  
27 SYNDACTYLIE DES ORTEILS  
28 SYNDACTYLIE PARTIELLE DES ORTEILS

AXE 40  
SCORE III.J FORME DU NEZ

119 ENSELLURE NASALE PLATE  
120 ENSELLURE NASALE SAILLANTE  
341 ARETE DU NEZ LARGE  
635 ARETE NASALE ETROITE  
152 RACINE DU NEZ LARGE  
153 RACINE DU NEZ ETROITE  
183 NEZ RETROUSSE  
318 NEZ CROCHU  
331 EXTREMITE DU NEZ LARGE  
454 NEZ BIFIDE  
558 NEZ POINTU

AXE 46  
SCORE III.R PAUPIERES

103 BLEPHAROPHYMOSIS  
104 BLEPHAROPHYMOSIS INCOMPLET  
362 FENTES PALPEBRALES ETROITES  
115 OBLIQUITE MONGOLODE DES FENTES PALPEBRALES  
116 OBLIQUITE ANTIMONGOLODES DES FENTES PALPEBRALES  
150 EPICANTHUS  
609 ABLEPHARIE  
123 COLOBOME DE LA PAUPIERE INFERIEURE  
124 COLOBOME DE LA PAUPIERE SUPERIEURE

ANNEXES

AXE 34  
SCORE III.D REGION MANDIBULAIRE

428 FENTE DU MAXILLAIRE INFERIEUR  
229 OREILLES BASSES  
320 ROTATION POSTERIEURE DE L'OREILLE  
114 MICROGNATHIE  
129 APPENDICE PRETRAGIEN  
130 FOSSETTE PRETRAGIENNE  
112 PROGNATHIE  
113 RETROGNATHIE

AXE 33  
SCORE III.C FACE

408 HEMIATROPHIE FACIALE  
677 TRAITTS GROSSIERS  
67 FACE ALLONGEE  
68 RACE RONDE  
69 FACE TRIANGULAIRE  
70 FACE CARREE  
228 ASYMETRIE FACIALE  
590 FACE APLATIE  
111 APLASIE DE LA REGION MALAIRE  
361 JOUES RONDES

AXE 45  
SCORE III.P POSITION DES YEUX

157 HYPERTELORISME  
158 HYPOTELORISME  
63 ENOPHTALMIE  
64 EXOPHTALMIE  
148 TELECANTHUS  
149 DYSTOPIE DES CANTHUS

AXE 42  
SCORE III.L ORIFICE BUCCAL

285 GRANDE BOUCHE  
369 BOUCHE PETITE  
363 LEVRES EPAISSES  
379 LEVRES FINES  
374 LEVRE INFERIEURE EPAISSE  
463 LEVRE INFERIEURE FINE  
391 LEVRE SUPERIEURE FINE  
404 LEVRE SUPERIEURE EPAISSE  
14 FOSSETTE DE LA LEVRE INFERIEUR  
160 PROTRUSION DE LA LEVRE INFERIEURE  
286 COMMISSURE LABIALE TOMBANTE  
1 FENTE LABIALE

ANNEXES

AXE 54  
SCORE V.A COLONNE VERTEBRALE

17 SCOLIOSE  
18 LORDOSE  
19 CYPHOSE

AXE 44  
SCORE III.N DENTS

185 DYSPLASIE DENTAIRE  
170 ANOMALIES DENTAIRES  
66 ANOMALIE DE L'ARTICULE DENTAIRE  
184 DENT SURNUMERAIRE  
259 HYPODONTIE

AXE 32  
SCORE III.A VOUTE CRANIENNE

339 GLABELLE SAILLANTE  
257 MICROCEPHALIE  
283 MACROCEPHALIE  
674 PERIMETRE CRANIEN NORMAL  
180 GRANDE FONTANELLE ANTERIEURE  
406 FONTANELLE POSTERIEURE GRANDE

AXE 51  
SCORE IIII.F PHANERES

156 SYNOPHRIS  
264 SOURCILS RARES  
625 SOURCILS ABONDANTS  
131 CILS RARES  
230 ALOPECIE  
263 CHEVEUX RARES  
467 HIRSUTISME  
466 IMPLANTATION CAPILLAIRE POSTERIEURE BASSE  
458 IMPLANTATION CAPILLAIRE FRONTALE EN V

## ANNEXES

## ANNEXE 19

## VALEURS ET VECTEURS PROPRES DE NIVEAU 2 - EXEMPLE TRAITE

NOMBRE DE VARIABLES : 22

Nombre de syndromes total : 7

Liste des syndromes : 32 77 78 82 95 96 100

VALEURS PROPRES :	% INERTIE :	% INERTIE CUMULEE :
D [ 1 ] = 3.72	60.01	60.01
D [ 2 ] = 1.21	19.51	79.52
D [ 3 ] = 0.78	12.62	92.14
D [ 4 ] = 0.27	4.27	96.41
D [ 5 ] = 0.13	2.11	98.52
D [ 6 ] = 0.09	1.48	100.00
D [ 7 ] = 0.00	0.00	100.00
D [ 8 ] = 0.00	0.00	100.00
D [ 9 ] = 0.00	0.00	100.00
D [10] = 0.00	0.00	100.00
D [11] = 0.00	0.00	100.00
D [12] = 0.00	0.00	100.00
D [13] = 0.00	0.00	100.00
D [14] = 0.00	0.00	100.00
D [15] = 0.00	0.00	100.00
D [16] = 0.00	0.00	100.00
D [17] = 0.00	0.00	100.00
D [18] = 0.00	0.00	100.00
D [19] = 0.00	0.00	100.00
D [20] = 0.00	0.00	100.00
D [21] = 0.00	0.00	100.00
D [22] = 0.00	0.00	100.00

## 8 premiers vecteurs propres :

AXE	V1	V2	V3	V4	V5	V6	V7	V8
68	-0.00	-0.13	-0.28	0.28	0.31	-0.22	-0.18	0.03
69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
72	-0.16	0.11	-0.12	0.03	0.12	-0.01	0.40	0.46
73	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
74	0.16	-0.11	0.12	-0.03	-0.12	0.01	-0.22	-0.05
75	0.47	-0.34	0.37	-0.09	-0.36	0.03	0.24	0.40
76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
77	0.25	-0.45	0.04	-0.24	0.54	0.07	-0.38	-0.02
78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
79	0.16	0.51	0.69	0.20	0.39	-0.20	-0.04	0.00
80	0.16	-0.11	0.12	-0.03	-0.12	0.01	-0.31	0.14
81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
83	0.46	-0.17	-0.24	0.59	0.21	-0.06	0.35	-0.00
84	0.16	-0.11	0.12	-0.03	-0.12	0.01	0.32	-0.78
85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
86	-0.02	0.09	0.07	0.35	0.02	0.91	-0.13	0.00
87	-0.61	-0.55	0.42	0.34	0.02	-0.01	0.09	-0.00
88	-0.06	-0.09	0.07	-0.47	0.47	0.26	0.47	0.02
89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

ANNEXES

ANNEXE 20

SUPER-ENTITE SYMPTOMATIQUE SELECTIONNEE AU NIVEAU 2,  
LISTE DES SYMPTOMES

AXE 87

SCORE XIII.II.D FONCTION CEREBRALE

246	ENCEPHALOPATHIE
307	RETARD PSYCHOMOTEUR
643	RETARD PSYCHOMOTEUR SEVERE
353	CONVULSION
526	SYNCINESIES
494	TROUBLE DU LANGAGE
529	QUADRIPLEGIE
550	TROUBLES DE LA COORDINATION
568	MOBILITE FACIALE ABSENTE
600	TROUBLES DU COMPORTEMENT
142	DEBILITE MENTALE PROFONDE
370	DEBILITE MENTALE LEGERE
429	DEBILITE MENTALE MOYENNE
530	TROUBLE DES SENSIBILITES
352	HYPOTONIE MUSCULAIRE
493	HYPERTONIE MUSCULAIRE



## PLAN SYNOPTIQUE DE LA THESE

Cette thèse poursuit un double objectif :

- Dans un premier temps: recenser, analyser et comparer les diverses théories et techniques de modélisation du Raisonnement Incertain et de l'Aide à la Décision, plus particulièrement dans le domaine médical; en ayant soin de préciser à chaque fois les structures mathématiques ainsi que l'orientation philosophique sous-jacentes.

- Dans un deuxième temps: s'attacher à un domaine particulier d'aide au diagnostic, celui de la Génétique Médicale, caractérisé par son étendue, son incertitude et son extrême évolutivité, et présenter une solution située au carrefour d'un certain nombre de grands domaines de l'informatique et des mathématiques. L'originalité du système tient principalement à l'utilisation de techniques d'Analyse des Données comme moteur d'apprentissage d'une connaissance destinée à un système expert.

Ce double objectif se reflète au niveau du plan général de la thèse:

- après une introduction générale,
- le chapitre I correspond à l'étude des modèles mathématiques d'Aide au Raisonnement et à la Décision Incertaine en Médecine,
- le chapitre II est consacré tout d'abord à l'exposé de la problématique de la production d'une connaissance experte dans le domaine considéré, ensuite à l'exposé du modèle théorique proposé, en relation avec les divers aspects du domaine,
- le chapitre III, plus technique, décrit la structure du logiciel correspondant à la mise en application du modèle théorique,
- le chapitre IV est consacré aux conclusions,
- une vingtaine d'annexes termine ce travail.