



**HAL**  
open science

# Diagnostic de pannes des systèmes complexes

Olivier Raoult

► **To cite this version:**

Olivier Raoult. Diagnostic de pannes des systèmes complexes. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 1989. Français. NNT: . tel-00332209

**HAL Id: tel-00332209**

**<https://theses.hal.science/tel-00332209>**

Submitted on 20 Oct 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.

**THESE**

présentée par  
**Olivier RAOULT**

pour obtenir le titre de **DOCTEUR**  
de **L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**  
(arrêté ministériel du 5 juillet 1984)

Spécialité : Informatique

# **Diagnostic de pannes des systèmes complexes**

Thèse soutenue le 20 février 1989 devant la commission d'examen.

Composition du jury :

M. J. Mossière	Président du jury
M. J. Demongeot	Rapporteurs
M. C. Roche	
M. J.M. David	Examineurs
M. A. Lux	
M. W Van de Velde	

Thèse préparée au sein du laboratoire LIFIA sous la direction de M. A. Lux



# INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

Président : Georges LESPINARD

Année 1988

## Professeurs des Universités

BARIBAUD Michel	ENSERG	JOUBERT Jean-Claude	ENSPG
BARRAUD Alain	ENSIEG	JOURDAIN Geneviève	ENSIEG
BAUDELET Bernard	ENSPG	LACOUME Jean-Louis	ENSIEG
BEAUFILS Jean-Pierre	ENSEEG	LESIEUR Marcel	ENSHMG
BLIMAN Samuel	ENSERG	LESPINARD Georges	ENSHMG
BLOCH Daniel	ENSPG	LONGEQUEUE Jean-Pierre	ENSPG
BOIS Philippe	ENSHMG	LOUCHET François	ENSIEG
BONNETAIN Lucien	ENSEEG	MASSE Philippe	ENSIEG
BOUVARD Maurice	ENSHMG	MASSELOT Christian	ENSIEG
BRISSONNEAU Pierre	ENSIEG	MAZARE Guy	ENSIMAG
BRUNET Yves	IUFA	MOREAU René	ENSHMG
CAILLERIE Denis	ENSHMG	MORET Roger	ENSIEG
CAVAIGNAC Jean-François	ENSPG	MOSSIERE Jacques	ENSIMAG
CHARTIER Germain	ENSPG	OBLEMED Charles	ENSHMG
CHENEVIER Pierre	ENSERG	OZIL Patrick	ENSEEG
CHERADAME Hervé	UFR PGP	PARIAUD Jean-Charles	ENSEEG
CHOVET Alain	ENSERG	PERRET René	ENSIEG
COHEN Joseph	ENSERG	PERRET Robert	ENSIEG
COUMES André	ENSERG	PIAU Jean-Michel	ENSHMG
DARVE Félix	ENSHMG	POUPOT Christian	ENSERG
DELLA-DORA Jean	ENSIMAG	RAMEAU Jean-Jacques	ENSEEG
DEPORTES Jacques	ENSPG	RENAUD Maurice	UFR PGP
DOLMAZON Jean-Marc	ENSERG	ROBERT André	UFR PGP
DURAND Francis	ENSEEG	ROBERT François	ENSIMAG
DURAND Jean-Louis	ENSIEG	SABONNADIÈRE Jean-Claude	ENSIEG
FOGGIA Albert	ENSIEG	SAUCIER Gabrielle	ENSIMAG
FONLUPT Jean	ENSIMAG	SCHLENKER Claire	ENSPG
FOULARD Claude	ENSIEG	SCHLENKER Michel	ENSPG
GANDINI Alessandro	UFR PGP	SILVY Jacques	UFR PGP
GAUBERT Claude	ENSPG	SIRIEYS Pierre	ENSHMG
GENTIL Pierre	ENSERG	SOHM Jean-Claude	ENSEEG
GREVEN Hélène	IUFA	SOLER Jean-Louis	ENSIMAG
GUERIN Bernard	ENSERG	SOUQUET Jean-Louis	ENSEEG
GUYOT Pierre	ENSEEG	TROMPETTE Philippe	ENSHMG
IVANES Marcel	ENSIEG	VEILLON Gérard	ENSIMAG
JAUSSAUD Pierre	ENSIEG	ZADWORNÝ François	ENSERG

**Professeur Université des Sciences  
Sociales  
( Grenoble II )**

BOLLIET Louis

**Personnes ayant obtenu le diplôme  
d'HABILITATION A DIRIGER DES  
RECHERCHES**

BECKER Monique  
BINDER Zdenek  
CHASSERY Jean-Marc  
CHOLLET Jean-Pierre  
COEY John  
COLINET Catherine  
COMMAULT Christian  
CORNUEJOLS Gérard  
COULOMB Jean- Louis  
DALARD Francis  
DANES Florin  
DEROO Daniel  
DIARD Jean-Paul  
DION Jean-Michel  
DUGARD Luc  
DURAND Madeleine  
DURAND Robert  
GALERIE Alain  
GAUTHIER Jean-Paul  
GENTIL Sylviane  
GHIBAUDO Gérard  
HAMAR Sylvaine  
HAMAR Roger  
LADET Pierre  
LATOMBE Claudine  
LE GORREC Bernard  
MADAR Roland  
MULLER Jean  
NGUYEN TRONG Bernadette  
PASTUREL Alain  
PLA Fernand  
ROUGER Jean  
TCHUENTE Maurice  
VINCENT Henri

**Chercheurs du C.N.R.S  
Directeurs de recherche 1ère Classe**

CARRE René  
FRUCHART Robert  
HOPFINGER Emile  
JORRAND Philippe  
LANDAU Ioan  
VACHAUD Georges  
VERJUS Jean-Pierre

**Directeurs de recherche  
2ème Classe**

ALEMANY Antoine  
ALLIBERT Colette  
ALLIBERT Michel  
ANSARA Ibrahim  
ARMAND Michel  
BERNARD Claude  
BINDER Gilbert  
BONNET Roland  
BORNARD Guy  
CAILLET Marcel  
CALMET Jacques

COURTOIS Bernard  
DAVID René  
DRIOLE Jean  
ESCUDIER Pierre  
EUSTATHOPOULOS Nicolas  
GUELIN Pierre  
JOURD Jean-Charles  
KLEITZ Michel  
KOFMAN Walter  
KAMARINOS Georges  
LEJEUNE Gérard  
LE PROVOST Christian  
MADAR Roland  
MERMET Jean  
MICHEL Jean-Marie  
MUNIER Jacques  
PIAU Monique  
SENATEUR Jean-Pierre  
SIFAKIS Joseph  
SIMON Jean-Paul  
SUERY Michel  
TEODOSIU Christian  
VAUCLIN Michel  
WACK Bernard

**Personnalités agréées à titre  
permanent à diriger des travaux de  
recherche (décision du conseil  
scientifique)  
E.N.S.E.E.G**

CHATILLON Christian  
HAMMOU Abdelkader  
MARTIN GARIN Régina  
SARRAZIN Pierre  
SIMON Jean-Paul

E.N.S.E.R.G

BOREL Joseph

E.N.S.I.E.G

DESCHIZEAUX Pierre  
GLANGEAUD François  
PERARD Jacques  
REINISCH Raymond

E.N.S.H.G

ROWE Alain

E.N.S.I.M.A.G

COURTIN Jacques

E.F.P.

CHARUEL Robert

C.E.N.G

CADET Jean  
COEURE Philippe  
DELHAYE Jean-Marc  
DUPUY Michel  
JOUVE Hubert  
NICOLAU Yvan  
NIFENECKER Hervé  
PERROUD Paul  
PEUZIN Jean-Claude  
TAIB Maurice  
VINCENDON Marc

**Laboratoires extérieurs  
C.N.E.T**

DEVINE Rodericq  
GERBER Roland  
MERCCKEL Gérard  
PAULEAU Yves

**" Comment pouvez-vous obtenir une traduction  
de termes vagues en des termes précis, vraie  
avec certitude ? "**

**Imre Lakatos, Preuves et Réfutations  
Editions Hermann 1984**



## Remerciements

Je tiens à remercier

Monsieur Jacques Mossière pour avoir accepté de présider ce jury ;

Messieurs Jacques Demongeot et Christophe Roche, rapporteurs, pour l'intérêt qu'ils ont manifestés pour ce travail ainsi que pour leur participation dans sa mise au point finale ;

Monsieur Augustin Lux pour les rares mais avisés conseils dont il m'a fait part tout au long de cette recherche ;

Messieurs Walter Van de Velde et Jean-marc David pour leur participation à ce jury.

Je remercie également Monsieur François Rechenman qui, bien que n'étant pas officiellement rapporteur, a bien voulu accomplir un travail comparable, contribuant ainsi à la mise en forme de ce document.

Une thèse étant essentiellement l'expression individuelle d'une sensibilité collective, je tiens à exprimer toute ma reconnaissance à Christian de Sainte Marie, Jean-michel Lefevre, Annie Luciani et Claude Cadoz qui m'ont accueilli, formé et m'ont inculqué une certaine conception de la science ainsi que du travail de recherche.

Ma gratitude va également à Messieurs Gérard Jourdain, Pierre Simonin et Henri Deboissieu, de la société GSI, pour le rôle moteur qu'ils ont joués tout au long de ces trois années, ainsi que pour leur inestimable soutien financier.

Le Docteur Henri Cloppet, médecin visionnaire aux activités innombrables, trouvera dans ces lignes l'expression de ma reconnaissance pour ses intuitions étonnantes dont découlent, pour une bonne part, les aspects les plus prometteurs de ce travail.

Enfin, je tiens à remercier tout particulièrement Monsieur Madame Mademoiselle ..... pour l'intérêt qu' il elle m'a toujours accordé, pour son soutien dans les moments difficiles ainsi que pour la valeur et l'opportunité des conseils qui m'ont été prodigués.



## Première partie

<b>1. Introduction</b> .....	<b>2</b>
1.1. Outils mathématiques de diagnostic.....	3
1.2. Approche IA .....	6
1.3. Présentation de la démarche.....	7
1.4. Historique de l'étude.....	9
<b>2. Qu'est-ce que le diagnostic ?</b> .....	<b>11</b>
2.1. Tour d'horizon des formulations usuelles.....	12
2.2. Etymologie du mot.....	14
2.3. Précisions sur la notion d'"observation" .....	15
2.4. Pré-requis pour la définition du problème.....	17
2.4.1. Diagnostic et observations.....	19
2.4.2. Diagnostic et référence.....	20
2.4.3. Diagnostic et qualification .....	22
2.4.4. Diagnostic et connaissances a priori.....	24
2.5. Caractérisation du diagnostic .....	26
2.6. Méthodologie.....	30
2.7. Conclusions.....	32
<b>3. Etude bibliographique</b> .....	<b>34</b>
3.1. Outils de base pour le diagnostic .....	34
3.1.1. Conversion FNC-FND. ....	34
3.1.2. Réseaux causaux (causal networks).....	36
3.1.3. Langages de qualification usuels.....	39
3.2. Raisonnement à partir de la structure .....	40
3.2.1. Le système GDE [De Kleer 86].....	41
3.2.2. Diagnosis from first principles [Reiter 86].....	44
3.2.3. Naive causal reasoning for diagnosis [Velde 85].....	47
3.2.4. Forces et faiblesses de ces méthodes.....	48
3.3. Raisonnement associatif.....	51
3.3.1. Parsimonious covering theory [Reggia 85].....	52
3.4. Récapitulatif sur les travaux existants.....	55
<b>4. Système de propagation de propriétés</b> .....	<b>56</b>
4.1. Formalisme de propagation de propriétés .....	56
4.2. Mécanisme d'évaluation.....	59
4.3. Exemple de formule.....	60
4.4. Application au diagnostic.....	61
4.5. Comparaison avec d'autres formalismes .....	64
4.5.1. Règles de production (ordre 0).....	65
4.5.2. Parsimonious covering theory .....	69
4.5.3. Calcul des prédicats (CP1) .....	70
4.5.4. Coefficient numériques ou ens. symboliques ? .....	72
4.6. Complétude .....	74
4.7. Méthodologie de réalisation d'un système .....	76
4.8. Conclusion.....	77
4.9. Récapitulatif.....	77

## Seconde partie

<b>1. Considérations relatives aux S.E. médicaux .....</b>	<b>83</b>
1.1. Le diagnostic, un processus .....	83
1.2. Différentes natures de connaissances .....	84
1.3. Connaissances, méta-connaissances, contrôle.....	86
1.4. Diagnostic et raisonnement incertain.....	86
1.5. Diagnostic assisté, principe et philosophie.....	89
1.6. Communication homme-machine .....	92
<b>2. GBDI, du formalisme au programme.....</b>	<b>95</b>
2.1. Particularités d'implantation .....	95
2.2. Développement et modularité .....	98
2.2.1. Extensions ne modifiant pas les formules .....	98
2.2.2. Extensions complexes.....	98
2.2.3. Extension de la grammaire.....	99
2.3. Aperçu des structures de données.....	100
<b>3. GBDI-IPLSYS, une application médicale.....</b>	<b>103</b>
3.1. Buts poursuivis.....	103
3.2. Données de GBDI-IPLYS .....	104
3.3. Méthodes et calcul ensembliste .....	107
3.3.1. Présentation générale.....	107
3.3.2. Méthodes effectivement implantées.....	112
3.4. Exemples de session.....	114

## Conclusion

<b>Récapitulatif et conclusion.....</b>	<b>122</b>
1. Apport conceptuel .....	122
2. Apport technique.....	123
3. Ouvertures possibles .....	124
3.1. Apprentissage des méthodes.....	124
3.2. Hybridation symbolique-numérique .....	124

## Annexes et bibliographie

<b>Annexe A - Exemple de modélisation .....</b>	<b>128</b>
<b>Annexe B - Représentation minimale.....</b>	<b>131</b>
<b>Annexe C - Exemple d'étude préliminaire .....</b>	<b>133</b>
<b>Annexe D - Réalisation d'une base de données.....</b>	<b>139</b>
<b>Glossaire des termes médicaux.....</b>	<b>143</b>
<b>Bibliographie</b>	

# **Première Partie**

## **Contribution à l'analyse du problème diagnostic**

---

# 1. Introduction

La réalisation de systèmes de diagnostic est une condition nécessaire à toute nouvelle augmentation de la complexité des systèmes construits. C'est également une condition nécessaire pour le développement des méthodes d'analyses de systèmes complexes. Par exemple, dans le domaine médical, on s'intéresse aux dysfonctions de systèmes excessivement complexes : des organismes vivants. Tant que l'on ne disposait que de moyens d'investigations rudimentaires, la complexité de l'organisme était masquée par la simplicité des données accessibles. Ce phénomène est en train de s'estomper. La multiplication et la spécialisation des examens disponibles impose de mener des raisonnements élaborés afin de choisir les examens les plus judicieux.

La complexité du diagnostic est liée à la complexité intrinsèque du système considéré, modérée par la finesse des méthodes d'investigations disponibles. Etant donné un système complexe, il sera éventuellement difficile d'effectuer une bonne interprétation des informations connues. Réciproquement, pour un vaste ensemble d'informations potentiellement accessibles, il peut être délicat de déterminer lesquelles seraient les plus pertinentes, eu égard aux hypothèses de dysfonctions antérieurement établies. A certains égards on peut considérer que le problème n'est donc pas tant de confirmer progressivement une hypothèse jusqu'à la transformer en certitude, mais plutôt de chercher l'information qui, quasiment à elle seule, permettra de conclure.

C'est essentiellement autour de ces deux aspects que s'organise cette thèse. Bien entendu un certain nombre de préalables seront indispensables afin de bien préciser ce que nous entendons par "diagnostic", comment nous le caractérisons, etc. Une fois défini un cadre conceptuel correct nous étudierons les principales idées et techniques qui regroupent une partie de la communauté scientifique autour de ce terme. Enfin nous proposerons un système formel de représentation de connaissances qui, pourvu d'une interprétation adéquate, permet de traiter le problème du choix des meilleures investigations dans le cadre de l'aide au diagnostic.

Avant d'entrer dans le vif du sujet, nous allons présenter un bref historique du problème ainsi qu'une première présentation des spécificités d'une approche de type "Intelligence Artificielle". Dans cette introduction, nous admettrons le terme de "diagnostic" selon son acception courante : recherche de la panne (la maladie) à partir des symptômes observés. De même, nous considérerons que l'IA diffère d'approches plus mathématiques, comme l'approche probabiliste, par les références

qu'elle fait au raisonnement humain. Nous reviendrons en détail sur ces deux points.

Les techniques de programmations, développées à partir des concepts de l'IA, ont été utilisées pour la première fois avec le système MYCIN [Shortl. 76], consacré à l'assistance médicale pour les infections microbiennes. C'est dire si la médecine a toujours été un champ d'application privilégié des recherches en IA, or la médecine est indissociable du diagnostic.

Il existe par ailleurs des formulations mathématiques du problème diagnostic représentées, en médecine, par le théorème de Bayes et, dans le domaine industriel, par les arbres de diagnostic.

## 1.1. Outils mathématiques de diagnostic

Depuis longtemps, les membres des professions médicales ont essayé de formaliser leurs méthodes de raisonnement, afin de rendre leurs pratiques plus systématiques ou d'en faciliter l'enseignement.

Les premières tentatives furent basées sur divers principes mathématiques (voir [Bonnet 84] pour une présentation détaillée), dont l'approche statistique. Cette dernière est essentiellement représentée par la formule de Bayes et ses extensions comme la théorie de Dempster-Shafer [Dempster 67].

Soit  $P(O/M_i)$  la probabilité d'observer  $O$  dans la maladie  $M_i$  et  $P(M_i)$  la probabilité a priori de la maladie  $M_i$  (par rapport à un ensemble fini de maladies  $M_j$ ), alors on peut obtenir  $P(M_i/O)$  par la formule suivante :

$$P(M_i/O) = \frac{P(M_i) P(O/M_i)}{\sum P(M_j) P(O/M_j)}$$

A partir de cette formule et d'un ensemble de symptômes  $O_1, O_2, \dots, O_n$ , il est possible de classer toutes les maladies  $M_i$  par ordre de probabilités décroissantes.

Cette approche est assez limitée. Tout d'abord parce qu'elle est lourde à mettre en oeuvre (probabilités conditionnelles symptômes/maladies) et ensuite parce que le classement d'un ensemble de maladies selon leurs probabilités respectives ne permet pas de traiter le cas d'un malade donné. En effet, l'approche statistique privilégie les maladies courantes au dépend des maladies rares, sans tenir compte de leur gravité respective. Or il existe une règle d'or en médecine qui s'exprime comme suit : s'occuper en priorité des maladies graves, mais curables. La maladie la plus probable peut alors être supplantée par la plus probable des maladies graves. Si l'on accepte cela, il n'y a aucune raison de se limiter à seulement deux critères

(fréquence et gravité). En pratique, il est faux d'assimiler le classement établi par un médecin à un simple classement probabiliste.

De plus, pour qu'une approche statistique soit valide il est nécessaire de mesurer effectivement la fréquence de tel ou tel phénomène. Pourtant, les valeurs de probabilités fournies par les experts traduisent davantage un degré de croyance qu'une pure fréquence observée [Fieschi 84], voir également à ce propos [Fieschi 79, Tversky 74, Rialle 87]. Enfin, les méthodes numériques entraînent souvent une survalorisation des pathologies les mieux connues ou des thérapeutiques les plus explorées (cf "Survalorisation des grands remèdes" [Broussalian 88]).

Les méthodes dites "statistiques" recouvrent plus, à l'heure actuelle, des méthodes de combinaisons de coefficients de croyance selon des théories mathématiques fortes (dont la théorie des probabilités), voire des méthodes d'acquisition de ces coefficients par des relevés de fréquences (non comparables toutefois à des campagnes de mesures telles que les requiert la statistique).

Une mutation récente de la médecine passe par l'usage de telles méthodes, non en pratique médicale courante, mais pour l'établissement même des connaissances médicales [Rialle 88, Robert 88]. Une description axiologique des maladies (cas typiques et déviations possibles) fait progressivement place à une description statistique (composantes principales, ellipsoïde de confiance, etc). Les données statistiques sont ensuite ré-interprétées sous une forme mémorisable pour l'étudiant en médecine. On peut alors se demander quel est le statut d'un médecin utilisant un ordinateur pour manipuler des masses de données statistiques ; est-il encore un médecin ou doit-on considérer comme tel l'association homme-machine ? Dans ce dernier cas, où situerait-on l'expertise ? A priori, l'approche numérique ne correspond pas au champ de notre recherche, mais nous essaierons néanmoins de nous situer par rapport à celle-ci (usages possibles, performances comparées, hybridation des approches, etc).

Une autre approche du diagnostic, nettement symbolique, est la constitution d'*arbres de diagnostic*. Cette méthode fige définitivement le protocole de recherche d'une panne à l'aide d'une structure arborescente (voir figure 1.1). Chaque noeud de l'arbre est attaché à une question et la réponse conduit vers un autre noeud, et ce jusqu'à ce que l'on aboutisse à une feuille portant le nom de la panne.

Outre les problèmes que cela pose si l'utilisateur ne sait pas répondre à une question, il est très difficile de simuler ainsi le phénomène de faisceau d'arguments<sup>1</sup> sans affronter une énorme combinatoire.

On pourra trouver une présentation détaillée du principe des arbres de décision dans [Quinlan 83]. La mise au point du système ID3 [Quinlan 86] a donné lieu à des études très complètes concernant l'acquisition et la minimisation de ces arbres. Enfin, parmi les développements les plus récents, [Quinlan 87] présente une méthode pour engendrer des règles de production à partir d'un arbre de décision, l'arbre de décision étant lui-même construit à partir d'un ensemble de cas (méthode

---

<sup>1</sup>Voir glossaire des termes médicaux

numérique de minimisation de l'entropie). C'est dire que ces différentes approches sont moins éloignées que l'on ne le croit généralement.

Les arbres de décision et les méthodes probabilistes sont généralement assez peu satisfaisantes dès que l'on s'attaque à des systèmes un tant soit peu complexes.

Le théorème de Bayes est peu apprécié des médecins même s'il fournit souvent de bons résultats. Sa valeur explicative est totalement nulle. De même, les arbres de décision figent le diagnostic en spécifiant un protocole unique, fût-il arborescent et complexe. D'autre part, l'usage des arbres de décision est une manière d'obtenir un résultat (une panne, une maladie) sans passer par un processus diagnostic ; le problème est davantage supprimé que résolu.

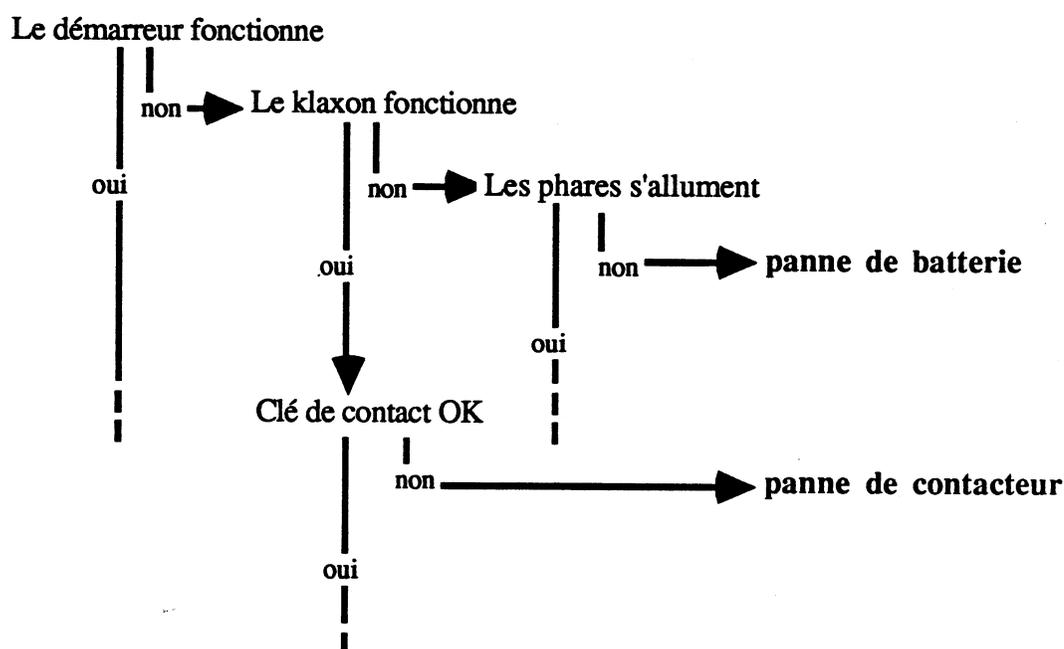


figure 1.1 Arbre de diagnostic

Toutes ces méthodes constituent néanmoins une base solide qui réapparaît dans certains travaux plus modernes. Nous avons déjà parlé de [Quinlan 83] pour les arbres de décisions, nous pourrions également citer [Peng 86] et [Robert 88] pour des extensions intéressantes de la formule de Bayes, [Yen 86] pour un développement de la théorie de Dempster-Shafer. A l'heure actuelle de nombreux travaux réhabilitent, s'il est besoin, les méthodes statistiques. Retenons simplement que, sous réserve d'être associées à un environnement approprié, ces méthodes sont toujours susceptibles d'être utiles. Elles peuvent connaître de nouveaux développements si, comme c'est le cas en médecine, de nombreuses connaissances de bases sont établies directement par ces méthodes.

## 1.2. Approche IA ; les systèmes de diagnostic

A la lumière de ce qui précède, nous pouvons fournir une première caractéristique de l'approche IA : elle cherche à éviter les écueils de la rigidité et de l'opacité. Un moyen d'y parvenir est de rendre explicite les connaissances utilisées. Ces connaissances ne sont pas nécessairement homogènes (nous séparerons les connaissances livresques, académiques et les connaissances expertes, relatives à l'usage de critères économiques ou opérationnels). Il est donc probable que, pour chaque type de connaissance, il faille développer une représentation adaptée.

Un système de diagnostic travaillant sur des systèmes réellement complexes est destiné à collaborer, au moins dans un premier temps, avec des opérateurs humains (médecins, surveillants de centrales nucléaires, etc). Il doit donc produire des résultats dans le langage des experts du domaine. De même, les informations d'entrées sont de format imposé. Par exemple, on ne doit pas demander à l'opérateur une probabilité conditionnelle qu'il ne manipule pas usuellement.

Un des postulats techniques de l'IA est que, dans la mesure où la nature des données d'entrée et de sortie est imposée par les experts, il est plus économique de chercher à reproduire le comportement des experts. Il serait d'ailleurs probablement impossible de développer une théorie correcte sur des données non strictement objectives. Alors que les méthodes mathématiques sont uniquement guidées par le problème initial, l'approche IA impose des contraintes supplémentaires.

Il n'est pas de notre propos de qualifier les programmes actuels (souvent dénommés systèmes experts) de "programmes intelligents", tant il est vrai que la sémantique contenue dans les structures de données, ainsi que les algorithmes de manipulation reste, sinon inexistante, du moins très sommaire.

L'Intelligence Artificielle, discipline largement critiquée par ses propres adeptes, se donne pour but d'établir les concepts et méthodes sous-tendant le comportement humain, relativement à un problème donné, en vue de le reproduire à l'aide de machines. Elle vise donc à simuler une réactivité par rapport à l'environnement que nous qualifions, chez l'être humain, de comportement intelligent.

D'un point de vue plus pragmatique, on peut y voir un ensemble de techniques permettant de s'attaquer à des problèmes complexes, peu ou mal formalisés, en essayant d'en éviter la combinatoire. Nous retiendrons donc que l'IA peut désigner à la fois l'Intelligence Artificielle et l'Informatique Avancée.

Après avoir vu brièvement ce qu'est l'IA, voyons maintenant ce qu'elle n'est pas. Les systèmes existants, ainsi que celui que nous présenterons ici, restent tributaires d'un opérateur humain tant pour l'acquisition effective des données que pour l'exécution des processus d'intervention (figure 1.2). C'est donc en partie grâce à l'intelligence de cet opérateur et à ses capacités d'interprétation que les informations produites par un système de diagnostic vont acquérir une signification. En deçà de cette interprétation, un système de diagnostic reste un simple programme

de manipulation de symboles. Le fait que ces données soient significatives pour un opérateur et que les algorithmes de manipulations aient été conçus en collaboration avec des experts ne font pas de ce programme un programme intelligent.

Nous pourrions admettre, paraphrasant H. Dreyfus [Dreyfus 72], que tant qu'il n'existera pas de robot, pourvu de capteurs et d'effecteurs, capable de mener à bien une fonction de maintenance sur un système complexe, et ce, de manière totalement autonome, on ne pourra pas parler de diagnostic automatique et/ou intelligent.

Retenons donc que l'IA est à la fois une approche issue des difficultés rencontrées par les méthodes mathématiques, une certaine idée de l'intelligence dont le mot clé pourrait être "comportement" et un rassemblement de techniques informatiques. C'est dans ce cadre que nous nous placerons pour étudier le diagnostic.

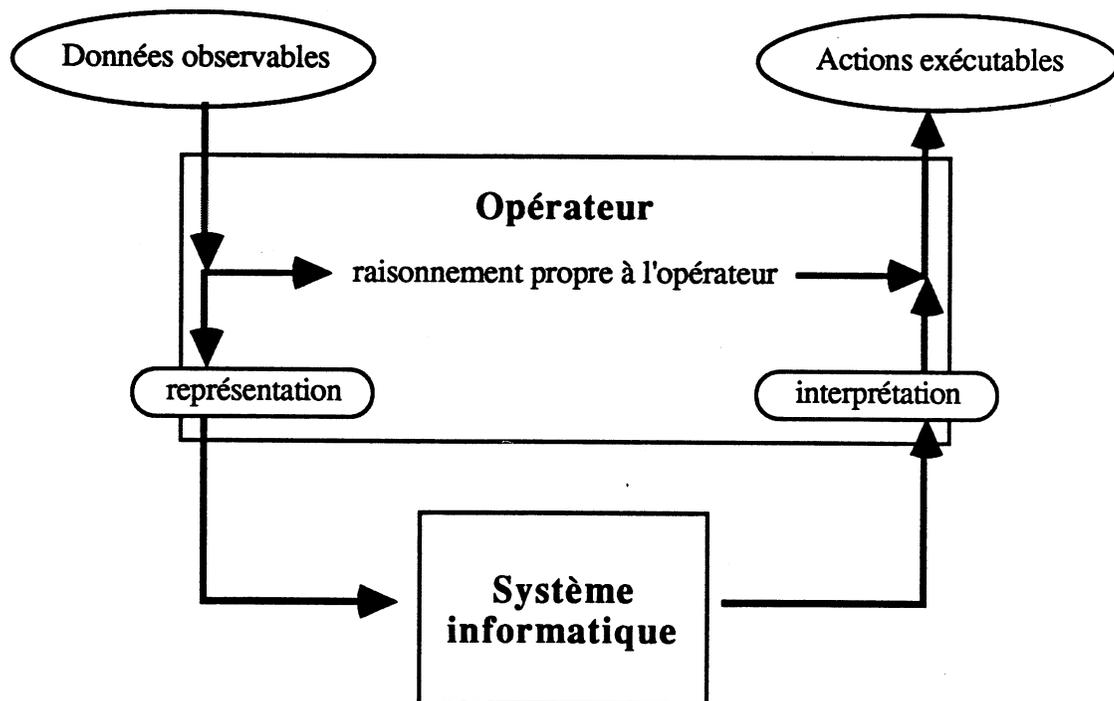


figure 1.2 Interaction opérateur-système

### 1.3. Présentation de la démarche

Nous tenterons de faire le point sur ce qu'est le diagnostic, quelles en sont les composantes principales.

Il est possible de répondre à ces questions selon deux points de vue : le plan technique, qui relève essentiellement des méthodes informatiques (représentation de données, algorithmique) et le plan que nous qualifions de "conceptuel". Ce dernier

point de vue est d'ailleurs une préoccupation ancienne des médecins (pour le diagnostic médical) et des ingénieurs (pour la maintenance de matériel). Il s'agit essentiellement d'explicitier les composantes et les phases d'un processus diagnostic. Les techniques informatiques liées à l'I.A. nous permettrons, de plus, une certaine expérimentation des concepts dégagés.

Comme nous ne nous attaquons pas à des problèmes concernant des entités mathématiques (les entiers naturels, les ensembles, les fonctions, etc) il est indispensable de définir, avant toute chose, les objets sur lesquels nous travaillerons. Nous pouvons citer, par exemple, des objets dénommés "pannes", des relations dites "causales". Que sont ces objets ? Comment les caractériser précisément sans rompre le pont qui les relie à la réalité d'une pratique du diagnostic ? Ces questions feront l'objet du chapitre 2. Afin d'éviter l'écueil philosophique, cadre où nous ne sommes nullement compétent, nous mènerons cette étude à partir de l'ensemble des idées et travaux qui caractérisent la communauté scientifique regroupée autour du thème "diagnostic". En conséquence, il est probable que le lecteur devra effectuer quelques aller-retour entre ce second chapitre et le chapitre 3 qui constitue notre étude bibliographique. Les concepts principaux présentés au chapitre 2 sont d'ailleurs extraits pour l'essentiel des consensus qui existent, ou n'existent justement pas, dans cette communauté scientifique en pleine formation. Bien que la synthèse globale soit de notre fait, nous ne prétendons nullement être à l'origine de toutes les notions présentées, prises individuellement.

L'ensemble de ce travail nous amènera à reformuler des appellations pour le moins ambiguës comme "modèles profonds" et "modèles de surface".

Suite à cette spécification du problème, nous proposons un nouvel outil permettant la construction de systèmes de diagnostic dans un cadre restreint : l'aide à l'interprétation et aux choix d'exams. Nous serons très attentifs à bien expliciter les hypothèses simplificatrices liées à la modélisation afin de bien circonscrire son domaine d'application. Cet outil, ici décrit de manière formelle, a donné lieu à une implantation, le logiciel GBDI, qui sera le principal protagoniste de la seconde partie de ce rapport.

Retenons dès maintenant que ces travaux sont résolument "verticaux", en cela que l'on s'intéresse à l'ensemble de la chaîne, de la spécification informelle au programme informatique, plus qu'à l'élucidation totale d'un unique niveau dans cette chaîne (comme le seul niveau formel par exemple). De ce point de vue, ces travaux ne relèvent pas exclusivement de l'informatique, mais bien de cette discipline qu'est l'intelligence artificielle.

Note : afin d'éviter de surcharger le présent rapport, nous avons reporté de nombreuses considérations et illustrations diverses en annexe. De nombreuses idées, exprimées dans les divers chapitres, semblent parfois être de simples tautologies. Une lecture des annexes associées peut alors aider à percevoir pourquoi, malgré leur évidence, elles sont difficiles à comprendre. Le plus délicat est souvent de ne pas leur ajouter de contraintes supplémentaires qui nous paraissent aller de soi. Par exemple, un médecin pense généralement en termes de maladies, un électronicien en

---

termes de composants fautifs, mais, lorsque nous emploierons le mot "panne", il ne s'agit ni nécessairement de l'un, ni nécessairement de l'autre.

## 1.4. Historique de l'étude

La première phase de cette étude a consisté en une étude approfondie du problème de l'aide au diagnostic médical.

Un système, même très simple, repose sur les capacités d'un opérateur à décrire un malade selon des termes convenus à l'avance. Ce mode de description peut être proche de données brutes (lecture d'une température sur un thermomètre, relevé d'une vitesse de sédimentation, etc) ou, au contraire, déjà très élaboré (affirmation de la présence d'un type d'état fébrile, estimation de l'état général). Le choix du niveau d'abstraction de ce langage de description est vite apparu comme un des choix cruciaux pour la réalisation d'un système. Partant de là, nous nous sommes interrogés sur la nécessité d'un tel langage et sur son rôle dans la définition même d'un problème diagnostic.

De même, personne n'attend d'un système informatique une prise en charge complète du malade. Un système va produire des listes d'hypothèses, éventuellement ordonnées, des propositions d'examens complémentaires, voire des ébauches de thérapeutiques. Là encore, nous avons dû examiner dans quelle mesure le choix a priori de ces contraintes en sortie engendrait un problème de type "diagnostic".

L'application évolua naturellement vers une base de données pourvue de mécanismes de filtrage "intelligents" (entendre par là des mécanismes conditionnés par l'évaluation d'un contexte). Une première tentative de généralisation nous a conduis à poser un formalisme de "propagation de propriétés" qui s'est avéré posséder un comportement intéressant pour la représentation des méthodes de raisonnement propres à certaines formes de diagnostic.

Une seconde tentative de généralisation nous a conduis à l'étude des systèmes de diagnostic consacrés, non plus au défaillances du corps humain, mais aux pannes de matériel. A première vue, bien que ces deux branches se réclament également du diagnostic, elles ne présentent que peu de similitudes. Par comparaison de ces deux approches, nous avons extrait un certain nombre de pré-requis à l'existence même d'un problème diagnostic. Nous avons également défini un cadre conceptuel où apparaissent nettement les choix qui entraînent le glissement vers l'une ou l'autre des classes de diagnostic (modèles profonds et modèles associatifs).

Ce cadre, pour peu productif qu'il soit d'un point de vue purement technique, permet de mettre en lumière des limites intrinsèques à tout processus diagnostic. Il permet simultanément de lever un certain nombre d'ambiguïtés qui subsistent dans les termes pourtant fondamentaux de "pannes" et "d'observations".

---

Ce rapport est l'image de ce cheminement. Nous avons mené de front la réalisation d'une application spécifique, un système d'aide à la prescription et à l'interprétation d'examens médicaux, et l'étude d'un cadre conceptuel très général permettant de réunifier les approches antithétiques du diagnostic (diagnostic médical et dépannage de matériels).

## 2. Qu'est-ce que le diagnostic ?

Dans le présent chapitre, notre but sera d'établir et d'argumenter une spécification générale de ce qu'est le diagnostic. En d'autres termes, nous allons tenter de répondre aux questions suivantes : quels sont les pré-requis à l'existence même d'un processus diagnostic ? Si ces conditions sont réunies, en quoi consiste cette activité ? Les diverses réponses à ces questions nous conduiront à préciser une caractérisation constructive des systèmes de diagnostic.

Comme pour la plupart des termes du langage courant, nous attachons à "diagnostic" une signification dépendante du contexte dans lequel il est employé et qui, dans chaque cas, nous paraît suffisamment claire pour éviter toute confusion. Par exemple, on retrouve l'usage de ce terme aussi bien dans le domaine médical, le dépannage de circuits électroniques que dans l'étude des dysfonctions de processus complexes (centrales nucléaires). Notre but étant de progresser en direction du diagnostic automatique, il est nécessaire d'objectiver la nature et les principaux composants de ce processus. Nous tenterons de construire une définition qui corresponde également à la perception intuitive que nous avons du diagnostic. Il va de soi qu'une telle spécification se place sur le plan conceptuel et ne représente en aucun cas un système formel.

En préambule, nous examinerons un ensemble de formulations présentées dans les principaux articles relatifs au domaine. Ces formulations s'avéreront très variées et peu compatibles. Certains auteurs fixent les buts, d'autres les moyens, et ils n'y consacrent généralement que fort peu de lignes. Or l'automatisation du processus diagnostic est un travail de longue haleine qui réclame un vocabulaire précis. Ce vocabulaire doit nous permettre d'identifier les différents composants conceptuels d'un tel processus. Il doit définir ce qu'est une *observation*, une *panne*, un *modèle*. Bien sur, il est impossible de fournir des définitions absolues de tels termes, au mieux nous espérons énoncer des propositions qui situent ces différents termes les uns par rapport aux autres. Par exemple, nous spécifierons la notion de panne par rapport à celle de modèle (ou plus exactement de *référence*). Nous établirons dans ce qui suit un ensemble de relations entre différents concepts. Ces relations nous permettront de mieux comprendre pourquoi tant de travaux portant sur le diagnostic semblent traiter de problèmes différents. Nous montrerons qu'il existe effectivement une unité de sujet ; cette unité est généralement masquée par les trop nombreuses hypothèses implicites.

Nous étudierons également l'étymologie du terme dans la langue française. Cette étude sera brève, car il est évident que nous aurons besoin de redéfinir ce mot dans le cadre technique qui est le nôtre. Elle sera néanmoins riche d'enseignements.

Ces deux approches (bibliographique et étymologique) nous fourniront des matériaux bruts. Nous tenterons d'en extraire les pré-requis indispensables à l'existence même d'une activité de type diagnostic, indépendamment de la nature de l'acteur (opérateur humain, ordinateur judicieusement programmé ou machine dite "intelligente"). Entre autre, nous définirons le diagnostic comme un sous-processus particulier d'un processus de maintenance.

## 2.1. Tour d'horizon des formulations usuelles

Avant de poursuivre plus avant, nous nous proposons d'explorer les différentes formulations du problème diagnostic. Nous verrons qu'il existe deux pôles entre lesquels s'articulent ces définitions. L'un est relatif aux applications visant essentiellement le dépannage automatique des cartes électroniques, l'autre fait davantage référence aux problèmes médicaux.

L'existence de ces deux pôles peut s'expliquer par le phénomène suivant : du fait de l'absence d'une spécification claire du processus diagnostic, les hypothèses simplificatrices restent implicites. Or la différence de nature d'une carte électronique (conçue dans un but spécifique) et d'un organisme vivant (résultat d'une évolution mal connue à finalité indéterminée) interdit probablement, respectivement dans ces deux domaines, de faire des hypothèses simplificatrices identiques.

Voyons quelques formulations du problème extraites des principaux articles relatifs au diagnostic.

Johan De Kleer [De Kleer 86] présente la définition suivante :

*La pratique du diagnostic requiert la capacité d'évaluer les différences entre le modèle d'un système et le système lui-même. Les différences entre le comportement effectif du système et le comportement prédit guide la recherche des différences entre le système et son modèle.*

Remarquons que cette proposition ne prétend pas donner une définition du diagnostic mais un des ses sous-buts, à savoir la capacité d'évaluer le fonctionnement effectif d'un système par rapport à un modèle. Selon cette formulation l'existence d'un modèle serait donc un pré-requis du diagnostic.

Dans le même ordre d'idée, Raymond Reiter [Reiter 86] définit le problème ainsi :

*Supposons connue une description d'un système ainsi qu'une observation d'un comportement de ce système qui diffère du comportement nominal. Le problème diagnostic consiste à*

*déterminer les composants dont la dysfonction expliquerait les différences observées.*

Prenons maintenant une formulation intermédiaire présentée dans [Hayes 83] :

*Le diagnostic est le processus de détection des pannes affectant un système (ou la détermination de la maladie affectant un système vivant) basé sur l'interprétation de données potentiellement bruitées. Ceci implique une interprétation des données. Le diagnostic sous-entend la compréhension de l'organisation du système observé (de son anatomie) ainsi que des relations et interactions entre les sous-systèmes qui le composent.*

Soit maintenant une application spécifiquement médicale : le système CSRL développé par B. Chandrasekaran. On trouve, dans un des articles associés [Chand. 86], la définition suivante :

*Le but du diagnostic est l'étiquetage d'un cas donné par un nom situé dans une hiérarchie pré-déterminée de diagnostics possibles. L'idée d'une telle hiérarchie est bien établie en médecine sous la forme de classification de maladies.*

On constate ici une disparition complète de la notion de modèle décrivant la structure, par contre nous reprendrons cette idée "d'étiquetage" sous le terme plus général de *qualification de pannes* .

Enfin J.M. David [David 87] caractérise le diagnostic comme sous-processus d'une action de maintenance :

*Le diagnostic est une partie d'un processus plus général de maintenance. Selon le contexte de maintenance, nous considérerons différentes sortes de causes pour un même dysfonctionnement. Par exemple, pour un système électronique, une cause de panne pourra être un composant ou une carte complète, selon que la réparation consiste à remplacer un composant ou une carte.*

Là encore, nous reprendrons cette idée d'inclusion dans un processus de maintenance en la précisant quelque peu.

Ces diverses formulations sont néanmoins peu ressemblantes. J. De Kleer se limite à la description d'un sous-but. R.Reiter réduit le problème à un processus d'identification et considère que le choix de l'observation ne relève pas du diagnostic (c'est une donnée initiale). B. Chandrasekaran ne précise ici que le résultat d'un processus diagnostic, sans spécifier la méthode, ni les données initiales, etc.

Bref, nous devons constater que ces différents auteurs ont plus fait porter leurs efforts sur des problèmes spécifiques, souvent avec succès, que sur une clarification de la nature du problème. Nous soutenons qu'une telle clarification, pour gratuite qu'elle puisse paraître, n'en est pas moins nécessaire. Son absence entraîne un éclatement du vocabulaire en fonction des applications visées et nuit à la compréhension.

Ces formulations sont insuffisantes, mais on y trouve un certain nombre de points communs ainsi qu'un ensemble de mots que nous considérons comme des mots-clés. Nous reviendrons en détail sur ces différents points. Avant cela, étudions un peu le terme même de diagnostic, tel qu'il est défini et compris dans la langue française.

## 2.2. Etymologie du mot

Le Nouveau Petit Larousse contient la définition suivante :

*Diagnostic : n. m. (du grec diagnôsis, discernement). Identification d'une maladie par ses symptômes.*

L'Encyclopaedia Universalis propose la définition et les commentaires suivants :

*Diagnostic : partie de l'acte médical qui vise à déterminer la nature de la maladie observée. Le diagnostic est indispensable à l'établissement du pronostic<sup>1</sup> et de la thérapeutique. Il est moins une phase de l'examen qu'une conclusion décisive de celui-ci. Facile et même évident dans certains cas, il peut être très difficile dans d'autres : devant l'urgence thérapeutique, le médecin se doit alors de décider si et quand les documents qui s'accumulent permettent l'arrêt de l'investigation et le passage à une conclusion, qui reste forcément révisable, mais n'en est pas moins nécessaire, car une recherche de plus en plus raffinée des signes pourrait parfois se poursuivre indéfiniment. C'est en ce sens qu'on comprendra que le même mot grec (diagnôsis) ait signifié à la fois discernement, décision et diagnostic médical.*

*[...] La médecine est devenue scientifique le jour où les découvertes de Morgagni et Laennec ont permis le diagnostic anatomopathologique, celles de Virchow le diagnostic microscopique, celles de Claude Bernard le diagnostic physiologique, celles de Pasteur le diagnostic bactériologique. S'y sont ajoutés depuis lors les diagnostics immunologiques, génétiques et biochimiques.*

---

<sup>1</sup>pronostic : conjecture sur l'évolution d'une maladie.

*On appelle plus particulièrement "diagnostic positif" le moment où sont confrontés les signes recueillis dans la description typique d'une maladie, et "diagnostic différentiel" (par analogie avec la définition par genre proche et différence spécifique) celui où ils sont confrontés avec les signes des maladies qui sont les plus voisines par leur symptomatologie.*

Première constatation : "diagnostic" est un terme médical. Son extension en un concept plus général comme le *diagnostic de pannes* et son utilisation dans le monde industriel et technique sont récentes.

On peut en voir une première émergence dans les travaux de J. De Kleer [De Kleer 76] où n'apparaît pas explicitement le terme de diagnostic. Les travaux de Randal Davis [Davis 82, 84] seront les premiers explicitement consacrés au diagnostic de pannes. En gros, il est possible de faire coïncider cette apparition du diagnostic non-médical avec le développement des techniques de propagation de contraintes, TMS et ATMS, et d'en fixer la date d'émergence vers les années 80.

Cette extension, pour audacieuse qu'elle soit, se justifie parfaitement tant il est vrai que le raisonnement lié au diagnostic, tel qu'il est intuitivement perçu, ne se limite pas aux systèmes biologiques. Il peut être étendu à tout objet observable pour lequel une réflexion plus ou moins complexe peut suppléer à une perception limitée. Toutefois, il est à notre avis fort dangereux, d'un point de vue conceptuel, de proposer une définition, à l'instar de celle de R. Reiter, qui exclut de son champ d'application la pratique médicale et renie l'historique du mot.

Un examen rapide des formulations présentées ci-dessus permet d'extraire des mots-clés comme :

- système observé, patient, malade ;
- modèle, comportement ;
- observation, examen ;
- interprétation, comparaison, identification.

Parmi les concepts sur lesquels nous reviendrons, nous pouvons citer principalement : l'existence d'un état nominal, la capacité à exhiber des divergences, le choix des observations et la qualification de panne. Nous verrons que ce dernier point, souvent manquant, assure le lien entre diagnostic et maintenance.

### **2.3. Précisions sur la notion d'"observation"**

Nous allons être amenés à utiliser très souvent les termes d'observations, de modèles. Ces termes peuvent être interprétés de manière variée, aussi nous allons présenter ici le contexte dans lequel nous les utiliserons. Tout ceci ne constitue pas une définition stricte ; par bien des aspects, nous simplifierons volontairement certains problèmes.

L'essentiel de notre caractérisation se trouve présentée dans la figure 2.1.

Cette figure présente les rapports entre un *objet* et un *acteur* susceptible d'observer cet objet. Considérons l'objet comme un *support d'informations* [CSMA 88]. Seules certaines de ces informations sont perceptibles par le système perceptuel de l'acteur. On peut considérer que l'acteur ne peut percevoir la totalité des informations portées par l'objet mais seulement un sous-ensemble. Il existe donc un objet simplifié que nous appellerons *description de l'objet* qui porte ces informations perceptibles. Le système perceptuel assure un transfert unidirectionnel des informations perceptuelles de l'objet réel vers sa description.

Dans un second temps, l'acteur *observe*, non pas l'objet, qui lui est inaccessible, mais la description de l'objet. Il s'agit là d'une acquisition contrôlée d'informations (ou perception descendante). Cette observation permet à l'acteur d'extraire des informations de la description. Ces informations ne prennent de sens que par intégration dans un ensemble de connaissances. On peut considérer ceci comme une phase *d'interprétation* de données observées. L'acteur peut alors construire un modèle contenant des informations effectivement possédées par l'objet, mais qui ne sont pas directement perceptibles (par la perception ascendante). Ce modèle constitue la *représentation* de l'objet (au sens fort de re-présentation, nouvelle présentation).

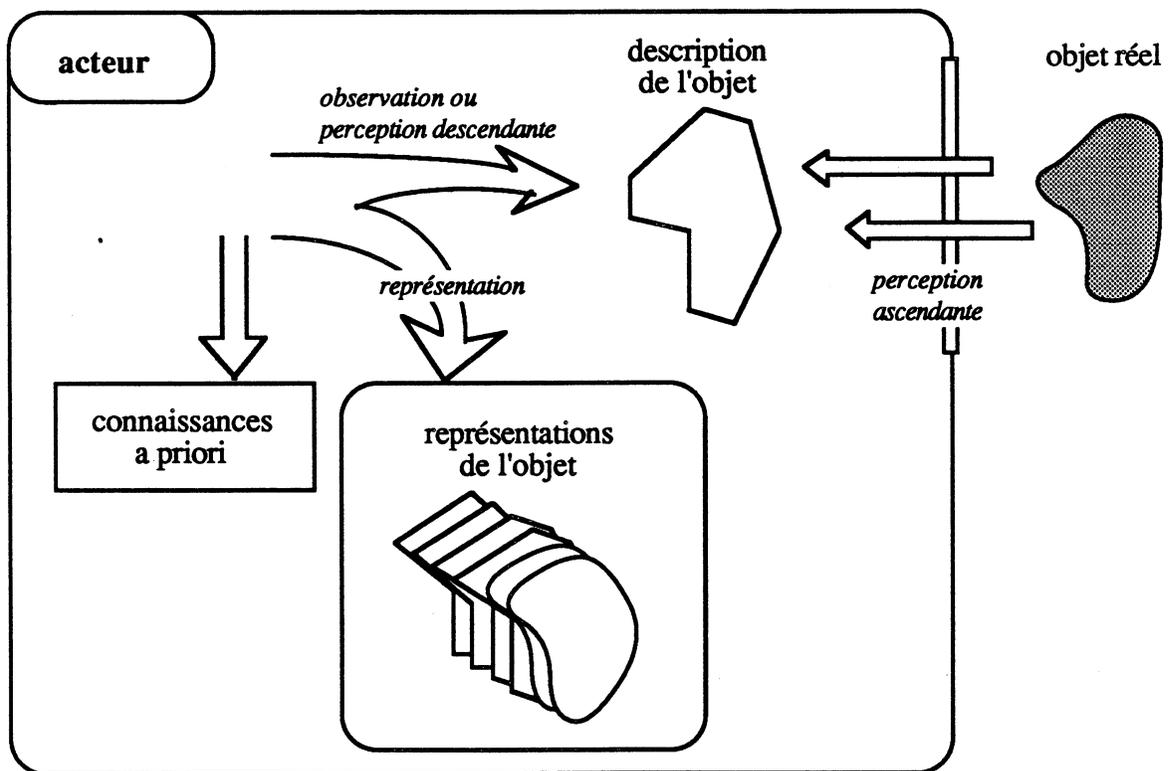


figure 2.1 : perception et observation

---

C'est en cela que le mécanisme d'interprétation, associant l'observation et l'enrichissement d'un modèle, peut être vu comme un moyen dont dispose un acteur pour "étendre" sa perception. D'autre part, sur la figure 2.1 nous avons dessiné plusieurs représentations. En effet, il est peu probable qu'existe une unique interprétation des informations perçues, et chaque interprétation donne lieu à une représentation distincte.

Prenons un exemple. L'œil voit de manière permanente et automatique, il s'agit là de perception ascendante. Un acteur peut observer ce que voit son œil ; c'est dans ce sens que l'on dit qu'il "regarde". En tout état de cause il est impossible à l'acteur de regarder ce que son œil ne peut percevoir. L'acteur regarde donc un certain signal porteur d'informations. Ces informations peuvent être transformées en *objet observé* sous réserve que l'acteur soit capable de qualifier ce qu'il regarde, de l'intégrer dans un certain modèle. Supposons que l'acteur observe une image et la qualifie de "tache rouge sur la peau d'un enfant" ; à partir de ses connaissances, de mise en correspondance de plusieurs observations, il est possible à cet acteur d'interpréter cette tache rouge comme un symptôme de la rougeole. Si nous revenons alors à l'objet initial, nous pouvons considérer que l'acteur "voit" la rougeole chez l'enfant. Cette information, la présence de la maladie, l'acteur ne l'a ni perçue, ni observée, mais elle fait partie intégrante de sa représentation.

Cette forme d'extension de la perception pourrait éventuellement être qualifiée "d'opération diagnostique". De notre point de vue, c'est élargir un peu trop la portée de ce mot. Si nous nous référons à la définition de l'Encyclopaedia, nous constatons que cette extension de la perception doit servir un double but :

- 1 - la recherche des observations pertinentes ;
- 2 - la production d'hypothèses de pannes.

Nous restreindrons donc le terme de "diagnostic" à ce mécanisme de construction d'une représentation d'un objet observé, uniquement si un tel enrichissement est dirigé vers les buts cités ci-dessus.

## **2.4. Pré-requis pour la définition du problème**

Lors de la rédaction de cette thèse, nous nous sommes longuement posé le problème de l'ordre de présentation. Fallait-il d'abord présenter les travaux les plus connus du domaine, puis reprendre ces travaux point par point pour en extraire une synthèse ? Fallait-il d'abord effectuer une analyse du problème afin de définir un cadre qui serait ensuite illustré par les travaux existants ? L'analyse que nous faisons du diagnostic n'est pas totalement nouvelle en cela que ses divers éléments se retrouvent, de manière éparse, dans de nombreuses publications. A notre connaissance, le regroupement que nous en avons fait, ainsi que la spécification de

nombreux points obscurs, est nouveau. Nous avons donc reporté l'étude bibliographique après la phase d'analyse. La conséquence de ce choix est que, pour un non-initié, l'analyse conceptuelle du problème peut paraître bien abstraite. De plus, elle est émaillée de remarques anticipant l'étude bibliographique. Nous ne pouvons que conseiller une lecture rapide de ces deux chapitres, même si quelques notions restent en suspens, suivie d'une relecture. Pour les initiés, peut-être un peu trop habitués à l'évidence des suppositions usuelles, un rapide survol de l'annexe A serait également profitable.

Dans ce qui suit, nous allons tenter l'énumération des différents composants nécessaires pour une spécification complète d'un problème diagnostic. Ceci nous permettra de définir un vocabulaire sur lequel nous nous appuyerons par la suite, tant pour analyser les systèmes existants, que pour définir une méthodologie de construction.

La caractérisation de l'activité diagnostique requiert la satisfaction des propositions suivantes :

- Il existe un objet "accessible". Le diagnostic est une activité d'un acteur sur un objet perceptible et observable. L'acteur doit disposer de mécanismes de perception ascendante, applicables à l'objet considéré, ainsi que de mécanismes de perception descendante<sup>1</sup> ;
- Il existe une spécification d'un comportement nominal qui permet de définir les notions mêmes de *bon fonctionnement* et de *panne* (ou bonne santé et maladie). Le concept de "comportement nominal" impose l'existence de comportements non-nominaux. L'objet observé est susceptible de diverger par rapport à la norme et la recherche des divergences constituera la première phase du diagnostic.
- L'acteur dispose d'une motivation précise. Son objectif est d'effectuer une interprétation de ses observations en vue d'acquérir certaines des informations portées par le système qui ne lui sont pas directement perceptibles. On sous-entend donc un désir de qualifier, selon un code défini a priori, ce qui est observé. C'est-à-dire que l'acteur veut fournir une description autre que perceptuelle et énumérative, mais plutôt relative à un codage d'un écart par rapport à la norme ("je vois un ensemble d'effets qui dénotent une panne nommée P, je vois P !").

Ces différents pré-requis pourraient définir un cadre général lié à la notion d'*interprétation*. On conçoit bien que cette notion est centrale dans le concept de

---

<sup>1</sup>Dans le cas le plus simple, il y a identité entre les informations perçues et les informations observées ; en d'autres termes, l'objet est "totalement" observé (au sens des limitations perceptives) avant toute analyse des observations. Le problème du choix de nouvelles observations ne se pose alors plus.

comportement intelligent. Nous en avons effectué ici une spécialisation en introduisant, en divers endroit, une référence explicite à la notion de panne. En effet, selon la compréhension intuitive que nous avons du mot, l'idée spécifique de diagnostic n'apparaît que lorsque cette interprétation est contrainte selon un but précis : la maintenance.

Nous allons maintenant reprendre ces points plus en détail et nous verrons en quoi ils permettent d'éclaircir le problème tout en étant opératoires pour la construction d'un système. Ils permettront également de mieux cerner l'unité effective des travaux relatifs au domaine.

### *2.4.1. Diagnostic et observations*

C'est une évidence que tout diagnostic repose sur une observation d'un objet réel. Il faut donc, implicitement ou explicitement, disposer de **capacités de description** de ce qui est observé.

C'est, parmi l'ensemble des problèmes abordés par l'IA, un de ceux qui n'a de sens que par son interaction avec le monde réel (autre exemples : l'apprentissage, le raisonnement incertain, la vision par ordinateur et la robotique) Un système de diagnostic doit donc disposer de facultés perceptives. Ces facultés peuvent être réelles ou virtuelles (capteur ou opérateur). Elles seront le support de la description de l'objet (valeur lue sur le capteur ou réponse acquise auprès d'un opérateur).

Tout système de diagnostic comporte un langage de description, implicite ou explicite, qui définit et *limite* les informations que le système utilise pour effectuer le diagnostic. Nous avons déjà explicité ce point dans le paragraphe précédent.

Pour une carte électronique simple, ce langage peut être du type "variable-valeur" ( $U=5v$  aux bornes de  $x$ ). Pour une carte plus complexe on peut utiliser des triplets "variables-valeur-contexte" ( $U=5V$  aux bornes de  $x$ , à  $T=t$  avec un vecteur de test  $V$ ). Dans le cas d'un circuit comprenant des interrupteurs (installations électriques domestiques, voir annexe C), on peut considérer qu'une observation est le relevé d'un graphe d'état des transitions constatées (une action sur l'interrupteur  $I$  allume ou éteint la lampe  $L$ ).

Enfin, pour des objets observés très complexes ou mal connus, on peut se limiter à une description booléenne (le circuit  $C$  chauffe, le malade a de la fièvre) ou multi-valuée (le malade à une fièvre faible, modérée, forte). Dans les travaux de V. Rialle [Rialle 87] le corps humain est divisé en un ensemble d'entités symptomatiques, qui représentent des noms de variables observables, et qui prennent leur valeur dans une liste de symptômes.

Le choix de ce langage est fondamental, il peut être complexe (valeur, graphes, chronogramme, etc) et son adéquation va conditionner les performances finales. En effet, le système tentera d'établir un diagnostic sur la seule base de phrases de ce

langage. Ce dernier doit être suffisamment complet pour que le diagnostic soit le plus facile possible, mais pas trop complexe afin de simplifier le processus d'observation (il est plus simple de mesurer une tension que d'établir un chronogramme). En d'autres termes, le langage de description doit être pertinent pour le diagnostic. Nous verrons par la suite que ce problème est loin d'être évident.

Il est certain que si la description est constituée par les réponses d'un opérateur, nous pourrions fort commodément évacuer le difficile problème de la perception effective. Un tel système de diagnostic utilisera, dès le premier abord, des informations de haut niveau, déjà qualifiées, ou re-présentées, par l'opérateur. Nous pourrions alors nous dispenser de doter un tel système d'un mécanisme de qualification bas-niveau (c'est-à-dire pour l'interprétation des informations de niveau "signal"). Plus exactement, nous nous intéresserons à un second mécanisme de qualification, de plus haut niveau, qui consiste à nommer une maladie plutôt que d'énoncer la liste de ses symptômes.

En pratique, la totalité des systèmes de diagnostic médicaux utilisent un tel raccourci et demandent à un opérateur de s'assurer de la présence, ou de l'absence, de tel ou tel symptôme. Pour les cartes électroniques complexes, on demande à un opérateur de "faire la mesure", ce qui peut également être passablement complexe (extraction d'un composant, positionnement de sondes, sélection d'un vecteur de test, etc). Ce passage du signal observé à la qualification symptomatologique n'est que rarement abordé et seulement dans des cas très restreints (interprétation automatique de radiologies du rachis, interprétation d'images angiographiques [Tsuyi 81], débogage de VLSI à l'aide d'un microscope électronique à balayage [Marzouki 88]). Il s'agit là de micro-mondes très simples où les processus d'observation effective peuvent être aisément automatisés.

Ajoutons brièvement que le langage d'observation ne se limite pas nécessairement à la description de l'objet observé, conçu comme une entité physique, mais qu'il peut confondre l'objet et de son environnement immédiat. Par exemple, une description peut contenir une information comme "le matériel radiologique est en panne", ce qui n'est pas une caractéristique du malade en tant que personne, mais qui est néanmoins le produit d'une perception concernant ledit malade (i.e. qui a son importance dans le processus diagnostic en cours).

#### *2.4.2. Diagnostic et référence*

La question de base que l'on va ensuite se poser est la suivante : "ces observations sont-elles normales ou impliquent-elles la présence d'une panne, et, si oui, laquelle ?". La phase d'interprétation des données observées fait donc apparaître la notion de **référence de bon fonctionnement** (ou description du comportement nominal). L'objet observé doit être comparé à un modèle nominal.

---

Cette contrainte justifie partiellement l'idée présentée dans [De Kleer 87] où l'existence d'un modèle du système observé est considérée comme indispensable à la pratique du diagnostic. Ceci est vrai sous réserve d'identifier les deux notions de *modèle* et de *référence de bon fonctionnement*.

Dans le cas d'une carte électronique, la référence de bon fonctionnement peut être un agrégat d'équations que doivent vérifier les grandeurs mesurables (s'il se trouve que ces grandeurs sont justement les entrées et les sorties de chaque composant élémentaire, on peut effectivement voir les équations comme une sorte de modèle). Par simulation, on pourra obtenir l'état nominal de la carte, étant donné un vecteur de test. La simple comparaison terme à terme des valeurs théoriques et des valeurs mesurées permet alors d'affirmer la présence d'une panne (mais non de nommer cette dernière). La recherche d'une divergence est donc simple, aux problèmes de simulation près.

Autre exemple : pour une installation électrique domestique, la référence peut être un graphe d'état complet (l'annexe C reprend cet exemple de manière détaillée). Si les observations constituent elles-mêmes un graphe d'état partiel, la recherche d'une divergence sera une fois de plus relativement aisée<sup>1</sup>.

Pour des systèmes plus complexes et non-décomposables on peut utiliser un modèle symptomatologique. On dira d'un malade qu'il a de la fièvre, sachant que l'état normal est l'absence de fièvre, d'une voiture qu'elle broute, sachant que l'état normal est un régime régulier. Le terme même de "symptôme" fait donc déjà appel à l'observation d'un état anormal. Si les capacités de description d'un objet se limitent à des symptômes, la référence (implicite) de bon fonctionnement est définie par l'absence de tout symptôme. C'est une forme de référence négative. Elle a principalement pour avantage d'éviter la phase de recherche d'une divergence, ce qui est une forme extrême de simplification.

Revenons un instant sur cette notion de "modèle". Dans le cas du dépannage de cartes électroniques, l'expression d'une liste de contraintes relatives aux seuls rapports entrées/sorties pour chaque composant est une méthode simple, voire la plus simple, pour caractériser le bon fonctionnement. Nous effectuons alors très naturellement un amalgame entre cette liste de contraintes et une idée de "modèle décrivant la structure". Mais cela n'est nullement nécessaire pour le diagnostic. Par exemple, on peut très bien utiliser comme référence des équations qui concernent les sorties de plusieurs composants distincts, même si ces derniers ne sont pas interconnectés entre eux. Une fois données les équations, nous pouvons les réécrire, selon les règles de l'algèbre, sans changer la référence. Les travaux de R. Reiter [Reiter 87] sont révélateurs de ce propos en cela que l'ensemble des contraintes spécifiant l'état nominal n'est pas assimilé à un modèle composants-connexions du système observé (voir étude bibliographique correspondante).

---

<sup>1</sup>Nous considérons dans cette première partie qu'un problème est (conceptuellement) "simple" dès lors qu'il peut être associé à un énoncé formel et/ou résolu par un algorithme glouton. La recherche de l'inclusion de deux graphes est de ce type.

En pratique, une référence de bon fonctionnement, pour une carte électronique, va définir des relations nominales entre des grandeurs mesurables, sans plus. Il n'est absolument pas indispensable de décrire les composants électroniques qui assurent ces relations. C'est par un artifice plutôt discutable que l'on s'arrange généralement pour affecter chaque relation à un unique composant, renforçant ainsi l'idée que la référence doit être une sorte de modèle "réaliste" de l'objet.

Cette confusion entre *spécification de comportement nominal* et *modèle* est très courante et relativement difficile à percevoir. Cela est en partie dû à l'absence de définition claire de ce qu'est un modèle. Dans cette étude, nous utilisons plutôt ce terme selon son acception courante, à savoir qu'un modèle d'un objet est une certaine description de l'objet, étant donné un observateur extérieur chargé d'évaluer le taux de ressemblance. Sinon, nous lui préférons systématiquement le terme de "référence". Nous aurons souvent l'occasion de revenir sur ce point.

Dans le domaine médical, le bon fonctionnement se définit plutôt comme l'absence de toute manifestation pathologique. Un système de diagnostic, alimenté par un opérateur, et utilisant la notion de symptôme comme mode de perception, dispose, dès l'abord, d'informations décrivant un écart par rapport à une norme. La simple présence d'un symptôme est déjà, en soi, un indice de panne. Il s'agit là d'une forme de spécification qu'il nous est plus difficile de mettre en correspondance avec la notion intuitive de modèle. De là vient principalement cette dichotomie entre modèles profonds et modèles associatifs. Nous pouvons donc dire que cette dichotomie apparente est la conséquence d'une interprétation erronée de certaines formes de référence de bon fonctionnement.

Cette unification des deux pôles usuels du diagnostic (modèles profonds et modèles associatifs) peut sembler artificielle au premier abord. Nous nous devons d'insister sur le point précédemment exposé : il n'existe pas de différence conceptuelle entre ces deux types d'approches. Il existe certes des différences techniques importantes, mais ce n'est pas ici notre propos.

### 2.4.3. *Diagnostic et qualification*

Poursuivons notre raisonnement : pour effectuer un diagnostic, on se basera sur une description de l'état de l'objet et la comparaison, explicite ou implicite, à une référence. Reste alors le problème de la description de la panne elle-même, de ce que nous avons nommé la **qualification de panne**. Supposons que l'on dispose d'un ensemble d'observations, d'une référence, et que l'on soit capable de vérifier si la référence est, ou non, respectée. Si elle ne l'est pas, on pourrait se contenter d'exhiber les mesures et les équations incompatibles et de déclarer : "la panne est décrite par ces observations et cette référence". L'exécution d'un processus diagnostic, tel qu'il est intuitivement perçu, doit aller plus loin.

---

Par exemple, pour une carte électronique, on désire exhiber, non pas une description de divergences, mais un nom de composant auquel on attribue la responsabilité de ces divergences. Il s'agit d'une qualification de pannes par identification de composants. On se fixe donc, a priori, une contrainte supplémentaire qui est de ne décrire les pannes qu'en termes de composants.

De même, en médecine, une maladie pourrait être définie comme une liste de symptômes, puisque chaque symptôme est une anomalie en soi. A la question : "docteur, qu'est-ce que j'ai ?" votre médecin pourrait alors vous répondre "vous avez de la fièvre, des troubles du rythme cardiaque et des boutons, tout ça n'est pas normal ! ". Une telle réponse ne vous satisferait pas. Vous attendez une qualification de votre état en terme de nom d'une maladie ("vous avez une hépatite virale", "la tuberculose", etc). Il s'agit, dans ce cas, d'une qualification de panne par identification de maladie.

Remarquons à ce point que le système de diagnostic ne se charge pas, a priori, de l'exécution du dépannage. La description des pannes, donnée par le système, peut alimenter un autre processus chargé d'effectuer les remplacements ou les réparations. Si nous supposons que ce processus nous est imposé ou tout au moins est très contraint selon d'autres critères (capacités d'action), cela signifie que les informations de sortie du diagnostic doivent respecter un certain format. Nous pouvons appeler ce format "langage de description de pannes" ; il importe néanmoins de remarquer que cette dénomination est arbitraire : les informations produites peuvent s'éloigner de la notion de panne telle qu'elle est intuitivement conçue. Par exemple une description de panne pourrait être : "le potentiomètre P doit être ajusté à 10k $\Omega$ ", "l'état de ce malade requiert une radio pulmonaire", "ce malade peut être traité par Nux Vomica", etc.

C'est généralement la simplicité des processus de dépannage implicites (échange standard de composants, thérapeutique pré-établie) qui incline à penser que le diagnostic n'est qu'affaire d'identification (identification de composants fautifs, identification de maladies).

Toutefois, on ne peut assimiler, conceptuellement parlant, le diagnostic de matériel à un processus d'identification de composants ni le diagnostic médical à un processus d'identification de nom de maladie. La notion de panne doit garder toute sa généralité : une panne est une phrase d'un langage de qualification.

Si le langage de qualification est un langage sans constructeur, on se ramène effectivement à une simple classification. Si, à l'inverse, ce langage contient des constructeurs comme  $\{\wedge, \vee, \neg\}$  alors il doit y avoir construction d'une description de panne (pour une approche générale de ces notions de construction-classification, voir [Velde 87], voir également le paragraphe 3.1.3).

Dans le même ordre d'idées, le choix des nouvelles mesures est souvent négligé [Davis 84, De Kleer 86, Reiter 86] au profit d'une stratégie d'exploitation maximale des informations déjà connues.

Ce choix relève-t-il ou non du diagnostic ? Nous affirmons que oui tant il nous semble qu'exploiter des mesures et repérer les informations manquantes relève du même mécanisme [Rialle 87]. Il serait effectivement possible de limiter le diagnostic à l'unique phase d'exploitation, mais cela irait à l'encontre de l'usage courant du terme. De plus, cela impliquerait la création d'un processus intermédiaire entre la maintenance et ce type de diagnostic qui serait justement chargé de piloter les diagnostics successifs et de régir les observations.

On peut voir un avis de ce type dans [De Kleer 86] qui écrit : "Le travail que nous décrivons ici représente un autre pas vers le diagnostic automatique, quoi qu'il en soit, il reste beaucoup à faire". Une partie du travail restant a d'ailleurs été accomplie par un retour à des méthodes numériques [De Kleer 88]. J.M. David [David 84] présentait également le diagnostic comme un problème d'acquisition-interprétation d'informations en vue de maintenance.

Nous considérerons donc, dans ce qui suit, que le choix de nouvelles observations fait partie des informations que doit produire un système de diagnostic. En toute rigueur, on devrait définir un langage de description d'observations qui serait l'exact pendant du langage de qualification de panne. Nous avons préféré distendre un peu ce dernier pour lui faire intégrer toutes les productions du système.

Dans l'application médicale qui fera l'objet de la deuxième partie, on retrouvera cette séparation dans la production de noms d'examens, sachant que ces examens permettent d'accéder à un, ou plusieurs, des symptômes discriminants. Il existe donc bien un langage de description de l'objet observé (symptômes constatés) distinct du langage décrivant les observations pertinentes (examens conseillés).

#### 2.4.4. Diagnostic et connaissances a priori

En résumé, la description d'un problème diagnostic passe donc par la spécification :

- de capacités perceptives (que suis-je capable de voir ?) ;
- d'une référence de bon fonctionnement (qu'est-ce qui est anormal ?) ;
- d'un langage de qualification des pannes (comment décrire la panne ?).

L'environnement, en entrée et en sortie, du processus diagnostic est alors parfaitement défini. **Le concepteur d'un système est totalement libre du choix du format de ces entités.** D'un point de vue plus formel, nous considérons qu'il est dépourvu de sens de parler d'un diagnostic, avant de préciser les trois entités énoncées ci-dessus. Le diagnostic n'est pas un problème qui se pose "naturellement", il est la conséquence de contraintes que l'on s'impose relativement à l'observation et à la description d'un objet donné.

---

Ces pré-requis étant définis, il reste à spécifier les relations qui unissent ces entités, puis à en extraire l'algorithme qui va manipuler ces données et qui, à des observations, va faire correspondre des hypothèses de pannes (des phrases du langage de qualification) fondées sur un modèle nominal de l'objet observé. Nous dénommerons l'ensemble des données complémentaires utilisées par cet algorithme "connaissances a priori". Ces données sont distinctes de la spécification de bon fonctionnement, même si, dans le cadre d'une implantation sur ordinateur, spécification et connaissances a priori se situent au sein d'une même structure de données.

Ainsi, dans l'approche General Diagnostic Engine [De Kleer 87], un système d'équations algébriques définit le bon fonctionnement. Ce système fixe des relations dites "nominales" entre paramètres mesurables. Nous insistons sur le fait que cette référence, en tant que système d'équations, est invariable pour toute réécriture respectant les règles de l'algèbre.

Ce système est ensuite découpé en une somme de contraintes bien précises où chaque contrainte est nommée. La plupart du temps les noms évoquent d'ailleurs un composant physique (additionneur n°1, multiplieur, etc). Ces pseudo-composants découlent en fait du découpage arbitraire du *système d'équations* en une *somme d'équations*. Enfin, cette "dénomination" des contraintes est un biais qui va permettre de qualifier les divergences en termes de conjonctions de noms d'équations transgressées (violated constraints). L'étiquetage des contraintes est donc une connaissance a priori. On peut voir ce processus comme une introduction des termes du langage de qualification dans une structure de données. Un algorithme adapté va ensuite "presser" cette structure et produire les descriptifs de pannes.

La théorie de la couverture parcimonieuse (*parsimonious covering theory* [Reggia 87]) fait référence à une liste exhaustive des symptômes, qui permet de définir l'état nominal (absence de tout symptôme). Une base de données décrit l'association des symptômes et des états pathologiques ainsi que des mécanismes d'exploitation ad-hoc (voir également l'étude bibliographique). Ainsi la référence est complètement implicite, alors que les données complémentaires nécessaires à la qualification sont, elles, totalement explicites. La situation est donc exactement inverse à la précédente où toute l'information semblait être contenue dans la référence.

Il est indispensable de séparer nettement la spécification de bon fonctionnement de la base de données sous-tendant les mécanismes de qualification. La spécification de bon fonctionnement est un pré-requis à l'existence du diagnostic, à l'existence même de la notion de panne, alors que la base de connaissances a priori est utilisée pour passer de l'expression des divergences à la qualification des pannes.

La spécification de bon fonctionnement permet de répondre à la question : le système observé est-il en panne ? La qualification est un processus qui permet de

décrire la panne selon un code pré-défini (distinct du codage des divergences). Selon les modes de représentation, un nom de panne peut être une conjonction de contraintes violées, un nom de maladie, ou tout autre langage qu'il nous plaira d'imaginer. Par exemple, les travaux de E. Broussalian [Broussalian 88] décrivent le problème de la prescription homéopathique comme une forme de diagnostic où le langage d'observation est constitué des symptômes ressentis par le malade, où la référence est implicite (absence de symptômes) et où le langage de qualification consiste en une description d'un traitement homéopathique. L'homéopathie n'utilise pas le concept de pathologie, qualification et thérapeutique sont donc confondues. Les connaissances a priori contiennent, entre autres, le principe de similitude et les répertoires de matière médicale.

Nous avons relevé dans ce qui précède deux points particulièrement importants. Il s'agit de deux différenciations, l'une non-justifiée (modèles profonds et modèles associatifs), l'autre non-perçue (référence de bon fonctionnement et base de connaissances a priori). En premier lieu, il n'existe pas de différences, autres que techniques, entre les modèles profonds et les modèles associatifs. Dans les deux cas, ces modèles sont, en fait, des modes de spécification du fonctionnement nominal qui permettent la détection de divergences. Ensuite il ne faut pas confondre cette spécification et l'ensemble des données qui vont permettre la qualification de la panne. Ces confusions sont d'autant plus faciles à faire que la référence et les données complémentaires sont souvent confondues au sein de la même structure informatique. Pour les modèles profonds, le système d'équations est aussi considéré comme une réunion d'équations "personnalisées" et non seulement comme un système. Pour les modèles associatifs, c'est la référence qui se trouve implicitement incluse dans la définition de symptômes alors que les relations causales symptômes-maladies servent de base aux mécanismes de qualification (cf. paragraphes 3.2.1 et 3.2.4).

## 2.5. Caractérisation du diagnostic

Paraphrasant l'Encyclopaedia Universalis nous proposons la formulation suivante :

*Diagnostic : partie d'une opération de maintenance qui vise à qualifier un dysfonctionnement (ou écart par rapport à une norme de fonctionnement) affectant un système observé. Le diagnostic produit toutes les informations nécessaires à l'activation du processus de dépannage ou, si cela n'est pas encore possible, les spécifications de nouvelles observations.*

Cette caractérisation peut être représentée par le paradigme suivant :

$$\text{MAINTENANCE} = [ [\text{OBSERVATION} + \text{DIAGNOSTIC}] * + \text{DEPANNAGE} ] *$$

Le diagnostic doit donc piloter l'acquisition des informations, afin d'établir la nature de la maladie ou de la panne, puis, lorsqu'une conclusion est possible, produire les informations d'entrées du processus de dépannage. De cette manière, nous définissons le format des entrées et des sorties du processus diagnostic, ainsi que le cadre dans lequel il s'inscrit. L'organigramme présenté figure 2 situe le rôle du diagnostic au sein d'une fonction de maintenance.

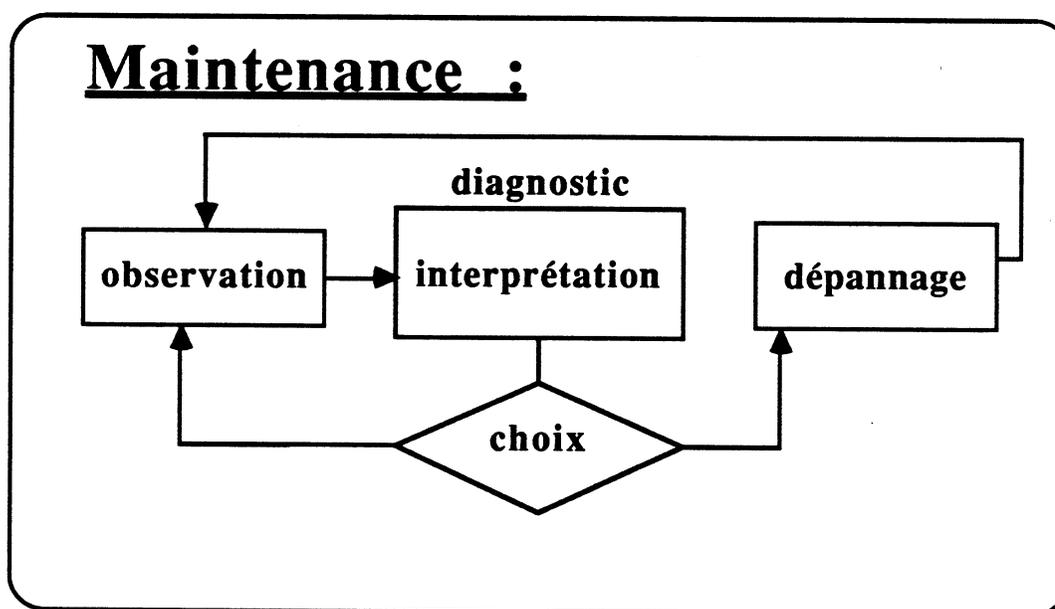


figure 2.2 : Maintenance et diagnostic

Le terme "maintenance" sous-entend l'existence d'une spécification de bon fonctionnement et définit la notion de dépannage. Le terme "observation" décrit cette notion d'observabilité partielle du système considéré. Notion qui, nous l'avons vu, rend nécessaire une interprétation des données par un processus diagnostic (cette interprétation s'appuie sur une base de connaissances non représentée ici).

On retrouve ainsi la triade PERCEPTION, RAISONNEMENT, ACTION. De plus, dans la mesure où l'observation n'est pas totale ni spontanée, un raisonnement fondé sur les observations initiales sera nécessaire afin de piloter les observations ultérieures. Ce mécanisme est souvent dénommé *prédiction-vérification* [Garvey 76, Lux 85] et peut se placer dans le cadre de la *perception active*.

Il est certain que, si les données sont très nombreuses, la phase d'analyse et d'interprétation de ces données va représenter l'essentiel du temps alloué au raisonnement diagnostic ; inversement, si les données sont rares (en début de processus), le bouclage acquisition-interprétation va être rapide et la durée de la phase d'interprétation négligeable devant le temps nécessaire à l'exécution des mesures. Selon les circonstances on percevra donc le diagnostic comme une activité essentiellement intellectuelle (interprétation des observations, évaluations d'hypothèses) ou pragmatique (choix et exécution de mesures).

La spécification présentée ci-dessus est donc très générale ; c'est une restriction de la problématique perception, raisonnement, action en vue d'une finalité précise : la maintenance. Cette généralité entraîne l'existence d'une grande diversité parmi les outils conceptuels et formels utilisables pour la réalisation d'applications. Chaque classe d'applications possède une affinité plus ou moins grande avec certains de ces outils, d'où la diversité apparente des approches.

Nous pouvons résumer ce qui précède sous la forme de deux propositions aux termes soigneusement pesés :

- une panne est la qualification d'une divergence entre une observation et une référence ;
- le diagnostic est un processus itératif de qualification partielle visant à énoncer une qualification terminale. Une qualification partielle contient une description des pannes possibles et des observations pertinentes ; une panne terminale est la description d'une panne unique (aucune discrimination n'est nécessaire) ou une disjonction de pannes indiscernables (aucune discrimination n'est possible).

Reprenant un vocabulaire défini dans [CSMA 88] nous pouvons dire que la recherche d'une panne est un cas particulier de résolution de problème. Selon ce vocabulaire, la maintenance se définit comme un *processus productif*, un mécanisme réflexe de haut niveau (le but est connu, c'est la référence de bon fonctionnement, et il faut agir pour l'atteindre). Par contre, le diagnostic est l'archétype du *processus cognitif* (le but est de caractériser un état inconnu, il faut acquérir de l'information).

Ces considérations permettent de dégager quelques limites intrinsèques à un système de représentation donné. Tout d'abord, les capacités d'observations doivent être suffisamment étendues pour permettre effectivement un diagnostic. Par exemple, pour une carte électronique, si notre seule perception se limite à la couleur et au poids de la carte, il est peu probable que nous arrivions à identifier les composants fautifs. La description, nécessairement partielle, dont nous disposons doit être pertinente eu égard au diagnostic voulu. Ensuite, étant donné ces capacités

d'observations et une référence, nous pouvons affirmer qu'il n'existe pas plus de pannes qualifiables que de divergences descriptibles. Notre capacité à exprimer les divergences entre les observations et la référence conditionne directement le nombre de pannes exprimables. En pratique, on peut également admettre qu'il n'y a pas de d'injection de l'ensemble des divergences et l'ensemble des pannes ; la qualification possède un effet "aggrégateur" en cela que des divergences distinctes peuvent être identiquement qualifiées (par exemple, plusieurs divergences distinctes peuvent, au cours de divers processus diagnostic, être finalement imputées à un même composant). On devrait donc préciser une des propositions précédente en : une panne est la qualification *d'une classe* de divergences ...

La figure 2.3 résume les rôles respectifs du langage d'observation, de la référence et du langage de qualification. Ce type de schéma est trompeur en cela que l'ensemble des objets descriptibles n'est pas nécessairement fini, ni même énumérable (par exemple, pour un langage d'observation du type "variable=valeur") de même que le nombre de classes "qualifiables".

Le langage d'observation définit un ensemble de situations descriptibles. Dans cette ensemble, la référence de bon fonctionnement isole un sous-ensemble de situations dites "normales" (éventuellement réduit à un singleton). Enfin, le langage de qualification partitionne l'espace restant en classes d'équivalences (chaque classe est associée à une unique qualification). Nous parlons ici de qualifications partielles et/ou terminales (c'est-à-dire qui contiennent, ou non, un énoncé des observations pertinentes). Chaque observation effectuée va modifier l'objet décrit (la position dans l'ensemble 1) et le but du processus est, par des "déplacements successifs", d'atterrir dans une classe terminale.

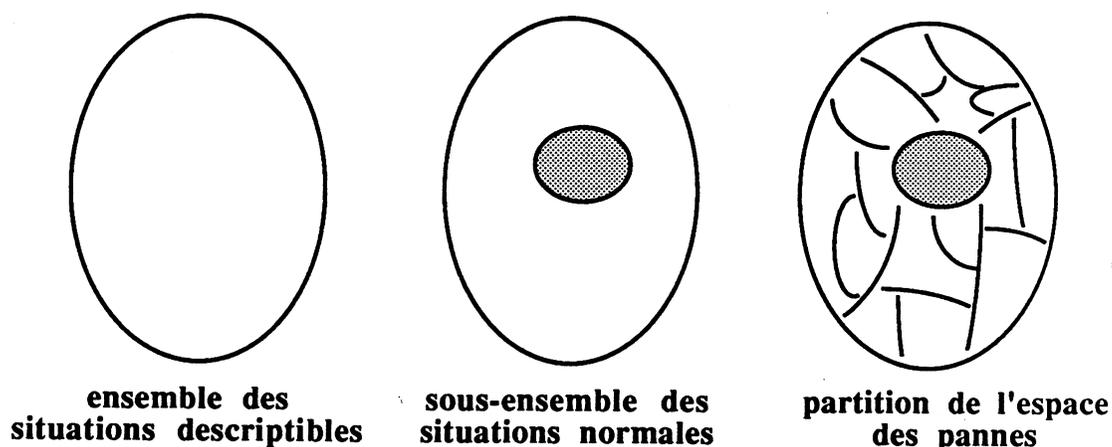


figure 2.3 : observation, référence, qualification

Fallait-il définir un *langage de description de divergences* ? Nous ne le pensons pas. L'expression des divergences est une capacité interne d'un système de

diagnostic, elle découle directement des capacités perceptives et de la nature de la référence. Au mieux, cela nous donne une alternative à la spécification d'un problème diagnostic. On pourrait remplacer le triplet "observation - référence - qualification" par "observation - divergence - qualification", la référence devenant une conséquence de la notion de divergence. Cette dernière formulation serait d'ailleurs assez pertinente pour décrire le diagnostic fondé sur l'usage de symptômes (diagnostic médical), à ceci près qu'il faudrait alors se limiter à "divergence - qualification", puisque l'énoncé des symptômes est à la charge de l'opérateur.

Nous allons maintenant préciser une méthodologie de construction de système de diagnostic. Nous abordons ici un aspect déjà plus "technique" que précédemment. Il s'agit de nommer différents composants d'un programme de diagnostic tels que le choix d'une expression pour chacun de ces composants fasse clairement ressortir les limitations conséquentes.

## 2.6. Méthodologie

Nous proposons ci-dessous une démarche permettant de mieux cerner les différentes étapes de la construction d'un système (un programme) de diagnostic. A chaque étape, correspond un ensemble de choix simplificateurs qui contraignent les choix suivants et restreignent la portée globale du système.

Nous nous bornerons ici à la simple présentation de cette méthodologie. Le lecteur pourra avantageusement revenir sur cette partie, une fois lue l'étude bibliographique. En effet nous ferons référence, à titre d'illustration, aux choix implicites faits dans un ensemble de travaux classiques [De Kleer 87, Reggia 87]. Ces illustrations peuvent être un peu obscures pour toute personne peu accoutumée aux techniques propres à ce domaine.

Nous proposons donc, dans le cadre de la conception d'un système de diagnostic, les différentes étapes suivantes :

**1 - Spécification de la référence de bon fonctionnement.** Cette spécification va donner un sens à la notion de panne. Elle peut être réalisée de manière extrêmement variée. Dans une approche "associative" [Reggia 87] on utilise une liste exhaustive de symptômes pathologiques. Le bon fonctionnement est défini comme l'absence de tout symptôme. Dans une approche "modèle profond" [De Kleer 87, Reiter 87] on construit une liste de contraintes entre paramètres mesurables. Le bon fonctionnement est alors défini comme la satisfaction de toutes les contraintes. L'annexe C présente un exemple de référence plus exotique.

**2 - Spécification des capacités perceptives.** Il s'agit ici des capacités perceptives du système ; en d'autres termes, c'est la description de toutes les

informations représentables dans le système à propos de l'objet observé. Cette spécification fige l'interface entre le processus d'observation et le processus diagnostic. Ces capacités peuvent être représentées, par exemple, sous la forme d'une liste de noms de symptômes effectivement constatés (ou d'une application de l'ensemble des symptômes dans {vrai, faux}). Dans un modèle de type GDE la perception est une liste d'association de paramètres mesurables (U, R, I, etc) et de valeurs perçues.

Selon les cas, il y a une interpénétration plus ou moins grande entre les deux concepts précédemment cités (spécification de bon fonctionnement et capacités perceptives) et les objets informatiques censés les représenter. Par exemple, l'approche associative utilise une liste de symptômes comme référence (une liste exhaustive) et une sous-liste comme perception. Une "observation" d'un malade va consister en une sous-liste particulière de symptômes effectivement constatés. Il s'agit ici, nous l'avons vu en détail au paragraphe 2.3.1, de "perceptions pré-interprétées" s'appuyant sur les capacités d'un opérateur. Pour l'approche GDE, la liste de contraintes définit la référence de bon fonctionnement et la liste des paramètres mesurables unissant ces contraintes spécifie les capacités perceptives du système. Une observation sera une association paramètre-valeur obtenue auprès d'un opérateur ou d'un appareil de mesure. Il y a également ici un phénomène de pré-interprétation, mais de plus bas niveau.

**3 - Description des données *a priori*** . On peut également regrouper ces données sous les termes de connaissances *a priori* , données complémentaires, base de données, etc. Citons par exemple la définition de relations causales entre maladies et symptômes, la connaissance de classes d'objets comme "les maladies graves", etc. Dans l'approche "modèles profonds", on peut également citer la connaissance du caractère individuel des contraintes, et la connaissance de vecteurs de test pré-établis.

**4 - Spécification du langage de qualification.** Ce point permet de spécifier ce que doit produire le processus diagnostic. Pour un dépannage par échange standard, la simple dénomination des éléments à changer suffit. S'il existe la possibilité de réglages (potentiomètres), des informations supplémentaires peuvent être jugées nécessaires. On peut également considérer que le diagnostic doit piloter, si nécessaire, de nouvelles observations. Dans ce cas, le processus diagnostic doit produire des informations destinées au processus de dépannage et/ou des informations destinées au processus d'observation. La qualification d'une panne est donc une description pertinente de l'état de l'objet observé, du point de vue d'un processus de maintenance considéré comme connu.

**5 - Choix d'un mécanisme d'interprétation.** Ce dernier point est plus technique que conceptuel. Les choix appliqués aux critères précédents vont contraindre de manière très forte les solutions possibles. Par exemple, une fois choisie la forme "liste de contraintes" comme spécification de bon fonctionnement, on débouche naturellement sur des mécanisme de propagation de contraintes. A

l'inverse, une approche symptomatologique conduit à l'usage de méthodes associatives (qualification par étiquetage). Les différents outils informatiques disponibles à l'heure actuelle sont d'ailleurs peu nombreux et il est rare que plus d'un permette de s'attaquer à une application donnée ; la possibilité de choix est donc assez théorique, du moins pour l'instant.

Cette démarche en cinq points permet de faire apparaître clairement les divers choix, arbitraires du point de vue de la spécification du problème. Bien entendu, chacun de ces choix correspond à une limitation des capacités du système final ; certaines pannes seront non-représentables, certains paramètres seront inobservables (par rapport aux capacités d'observations d'un expert du domaine), enfin certaines pannes représentables pourront n'être pas inférables du fait des limites du mécanisme d'interprétation. En d'autres termes, le diagnostic, vu comme une application du langage de description dans le langage de qualification, n'est généralement pas surjectif.

En pratique, nous pensons qu'il est préférable de commencer par la spécification de la référence. Nous pourrions alors nous pourvoir des capacités d'observations nécessaires pour mettre en évidence un ensemble de divergences. Les divergences exprimables seront alors un guide et une limite pour la spécification du langage de qualification.

## 2.7. Conclusions

La volonté de proposer une caractérisation du diagnostic qui englobe les différentes perceptions intuitives du problème, relativement à de grandes classes d'applications, nous a conduit à préciser un ensemble de pré-requis. Citons par exemple l'existence d'un système réel partiellement observable, d'une référence de bon fonctionnement et d'une volonté de maintenance.

De fait, la définition proposée est une spécialisation, dans le cadre de la maintenance, de la problématique perception-raisonnement-action. Cette généralité explique en grande partie la diversité des approches conceptuelles et formelles. Nous avons néanmoins extrait de ces considérations une méthodologie de construction de systèmes de diagnostic. Le caractère opératoire de cette méthodologie est encore limitée, du fait du petit nombre de mécanismes d'interprétations disponibles à l'heure actuelle. Une autre partie de notre travail consistera à proposer des outils ouvrant de nouveaux champs d'applications.

Retenons la différence conceptuelle séparant la spécification de bon fonctionnement, le support même du concept de panne, et la base de connaissances a priori permettant la qualification. Cette différence est généralement difficile à percevoir du fait du mélange des représentations associées au sein d'une même structure informatique (contraintes et réseau de contraintes, ensembles et sous-

---

ensembles). Cette structure est souvent perçue comme un modèle de l'objet. En d'autres termes, nous associons intuitivement à la notion de "modèle" bien plus qu'une simple référence. Le modèle est décrit par C. Cadoz [Cadoz 88] comme une entité *simplificatrice* (elle évite d'énoncer toutes les conclusions de toutes les observations) et *productive* (elle permet de calculer les conclusions à partir des observations). Le modèle, ainsi décrit, est alors effectivement l'union de la référence, des données sous-tendant l'interprétation et de l'algorithme d'interprétation lui-même. On retrouve alors le fait que le modèle est complètement finalisé vers un but productif donné.

Nous avons abondamment utilisé les termes de "modèles profonds" et "modèles associatifs", souvent avec des guillemets, toujours avec réticence. Nous préférierions remplacer "modèles profonds" par "spécification de bon fonctionnement par liste de contraintes et qualification de pannes par identification de contraintes transgressées". De même, nous préférierions remplacer "modèles associatifs" par "spécification de bon fonctionnement par une liste de symptômes et qualification par identification de pathologies".

En raccourci, cela donnerait "diagnostic par contraintes individualisées" et "diagnostic symptomato-pathologique" (cette deuxième appellation est un pléonasme étant donné l'origine médicale du mot diagnostic). L'imagination nous a fait défaut pour découvrir des termes plus agréables aussi nous ne les imposerons pas au lecteur. Autant que faire se peut, nous préciserons la nature de la référence de bon fonctionnement lorsque cela sera nécessaire.

L'annexe C donne un exemple de modèle qui n'est ni profond, ni associatif, pour le dépannage d'installation électrique domestique. Les observations sont des graphes d'états, la référence est le graphe nominal et la qualification se fait par énoncé de connexions erronées.

Enfin, nous avons insisté sur le caractère bouclé d'un processus diagnostic. Les systèmes informatiques classiques (systèmes experts) étant dépourvus de moyens d'actions effectifs, nous considérerons que la boucle peut être fermée par l'utilisateur. Tout système devrait, pour sa part, produire des hypothèses de maladie (de panne), ainsi que des conseils relatifs aux examens pertinents.

### 3. Etude bibliographique

Ce chapitre est consacré aux modèles formels employés pour la réalisation de systèmes de diagnostic et aux techniques qui en découlent. Notre préoccupation principale sera d'effectuer une présentation synthétique des réussites techniques réalisées dans ce domaine.

Dans un premier temps, nous nous intéresserons à quelques concepts généraux que l'on retrouve dans de nombreuses études. Pourtant l'usage implicite, et parfois contourné, d'outils de base comme l'algèbre de Boole entraîne souvent une complexification excessive des problèmes traités.

Ensuite nous examinerons plus spécifiquement les travaux de J. De Kleer (General Diagnostic Engine), de R. Reiter (Diagnosis from first principles) puis de W. Der Velde (Naive causal reasoning for diagnosis). Nous avons choisi d'essayer de mettre en valeur les hypothèses implicites liées aux représentations adoptées afin de mieux illustrer les idées présentées dans le chapitre précédent. Rappelons que, selon notre analyse, il doit être possible d'exhiber une référence de bon fonctionnement, des capacités perceptives, une interface avec un module de dépannage (qualification et/ou observations complémentaires) ainsi qu'un mécanisme d'interprétation de données observées basé sur un ensemble de connaissances a priori. Ces travaux sont usuellement rangés dans la catégorie "modèles profonds". Nous montrerons qu'ils comportent des divergences subtiles mais importantes. Nous insisterons sur les atouts de ces méthodes ainsi que sur leurs principales faiblesses.

Nous examinerons ensuite la seule théorie "associative" bien posée, à notre connaissance, qui est due à J. Reggia (Parsimonious covering theory). Cette théorie nous servira de point de comparaison pour le formalisme que nous présenterons dans les chapitres suivants.

#### 3.1. Outils de base pour le diagnostic

##### 3.1.1. Conversion FNC-FND.

La conversion d'une forme normale conjonctive (conjonction de disjonctions) en une forme normale disjonctive (disjonction de conjonctions) occupe une place importante dans la problématique du diagnostic.

Par essence même, une hypothèse de panne va s'ajouter, de manière disjonctive, à une liste d'hypothèses antérieures. S'il se trouve que toutes les hypothèses concevables incriminent un élément commun (une maladie commune), alors nous pourrions considérer cet élément comme "certainement" fautif. Un tel comportement peut être facilement représenté grâce aux opérateurs logiques de l'algèbre de Boole : la conjonction et la disjonction. Nous serons par la suite amené à utiliser deux formes usuelles d'algèbres de Boole : un espace de prédicats muni de la conjonction et de la disjonction, et l'ensemble des parties d'un ensemble muni de l'union et de l'intersection.

Du point de vue plus pragmatique du dépannage de circuits, ce problème s'exprime de la façon suivante : supposons (cf figure 3.1) deux observations  $O_1$ ,  $O_2$  incompatibles entre elles au regard d'une chaîne de contraintes les unissant. Nous serons naturellement amenés à postuler que l'une, au moins, des contraintes n'est pas respectée. Si, à chaque contrainte  $C_1, C_2, C_3$ , nous associons un prédicat  $F_{C_1}, F_{C_2}, F_{C_3}$  signifiant que la contrainte correspondante est "fautive" (transgressée) alors nous pourrions énoncer la proposition :

$$F_{C_1} \vee F_{C_2} \vee F_{C_3}$$

Il est alors facile de voir que, s'il existe un autre cheminement possible<sup>1</sup> entre  $O_1$  et  $O_2$ , par exemple  $C_4, C_5$ , alors la proposition deviendra :

$$(F_{C_1} \vee F_{C_2} \vee F_{C_3}) \wedge (F_{C_4} \vee F_{C_5})$$

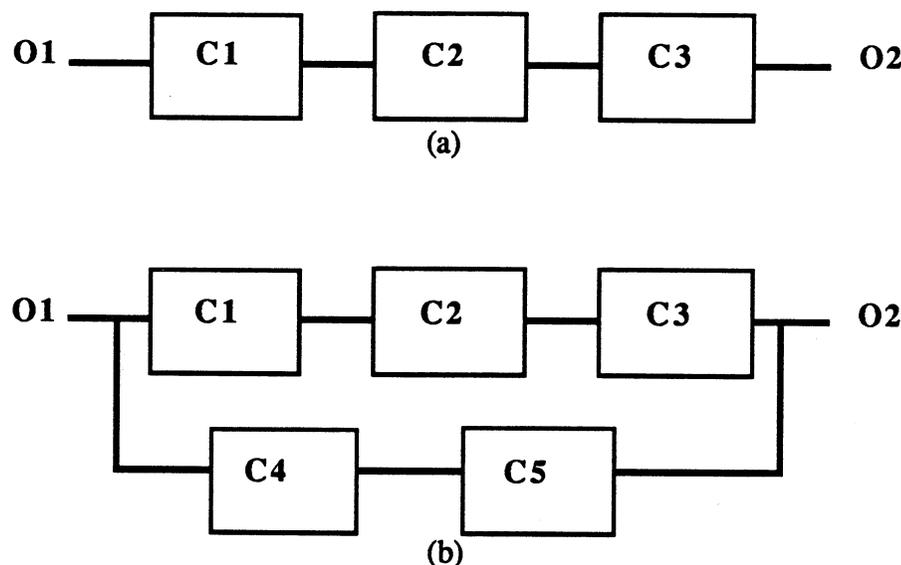


figure 3.1 : chemins possibles et disjonction

<sup>1</sup>Le caractère étrange d'une variable alimentée par deux circuits distincts disparaît si l'on renonce à une interprétation en termes de circuits, de connexions et d'entrées-sorties ... ce qui est indispensable ici. Il s'agit uniquement d'une variable apparaissant dans l'expression de plusieurs contraintes.

L'existence d'une contradiction sur un chemin possible entraîne l'expression d'une disjonction de défauts élémentaires ; l'existence de plusieurs chemins possibles unissant deux observations contradictoires fait alors apparaître une conjonction de disjonctions (forme normale conjonctive).

La transformation d'une telle forme en une forme normale disjonctive revient à introduire la notion de "pannes composées" et à décrire le dysfonctionnement en termes de disjonctions de pannes composées. Ici, cela donnera :

$$(F_{c1} \wedge F_{c4}) \vee (F_{c2} \wedge F_{c4}) \vee (F_{c3} \wedge F_{c4}) \vee \dots$$

Le parallèle avec le domaine médical est immédiat. Si un symptôme  $S_1$  est une manifestation possible de trois maladies connues  $M_1, M_2, M_3$  (à l'exclusion de toute autre maladie) et qu'un symptôme  $S_2$ , simultanément présent, est une manifestation possible de  $M_4$  et  $M_5$ , alors une description possible du malade est donnée par :

$$(M_1 \vee M_2 \vee M_3) \wedge (M_4 \vee M_5)$$

La paternité de ces notions est difficile à établir. De fait, elles apparaissent dans la plupart des articles relatifs au diagnostic des pannes multiples, sous une forme plus ou moins explicite [De Kleer 86, Reiter 85, Geffner 87, etc].

### 3.1.2. Réseaux causaux (causal networks)

L'idée de représenter un objet comme un enchaînement de relations de cause à effet est également fort ancienne. Elle a donné naissance au principe de suspension de contraintes énoncé par R. Davis [Davis 84].

Soit un composant (au sens large : un composant électronique, une fonctionnalité, une propriété) de l'objet observé. Ce composant est plongé dans un environnement et assure une certaine propriété. Ce composant peut être, ou non, défectueux, ce que nous noterons avec le prédicat  $F$ . Considérons que l'environnement soit observable et que le prédicat  $E$  indique sa validité. De même,  $P$  indiquera si la propriété est vérifiée ou non.

Exemple : pour un additionneur,  $E$  sera vrai si le circuit est bien alimenté et si la nature des signaux d'entrées est conforme à une certaine norme électrique ;  $P$  sera vrai si la valeur de sortie est bien la somme des valeurs d'entrées.

La proposition minimale que l'on puisse faire sur un composant s'exprime :

$$E \wedge \neg F \Rightarrow P \quad (1)$$

Elle s'interprète comme suit : "Si un composant est plongé dans un environnement adéquat et qu'il n'est pas défectueux alors il doit assurer sa fonction nominale". Remarquons que seuls E et P sont susceptibles d'être connus à partir d'observations, alors que F est non-observable.

Une conséquence de cette proposition est que, dès lors qu'un composant est défectueux ou qu'il est plongé dans un environnement anormal, toute déduction quand à son comportement devient impossible. Si la propriété P est une contrainte unissant des variables, alors la prise en compte de  $(\neg E \vee F)$  consiste à supprimer cette contrainte ; les variables redeviennent libres.

De fait, cette proposition minimale est extrêmement faible. Par exemple, elle n'interdit pas le phénomène de la "panne invisible" (représenté par  $E \wedge F \wedge P$ ). On considère alors qu'un composant peut être défectueux, même si ce défaut n'entraîne aucune conséquence observable, ce qui pose pour le moins le problème quasi-philosophique de la signification du terme "défectueux".

Autre faiblesse de cette proposition : sa sensibilité à l'environnement. Dans un problème diagnostic, on est confronté à des divergences entre des observations et une référence. Si la notion "d'environnement correct" s'appuie sur cette même référence, l'apparition d'une divergence nous privera simultanément de tout moyen d'en inférer la cause. C'est donc une règle générale de tout système de diagnostic que la description de l'environnement doit être dissociée de la description des propriétés (on ne doit pas avoir  $\neg P \Rightarrow \neg E$ , sinon F devient indémontrable). Par exemple, toujours pour le même additionneur, l'environnement exprime des contraintes de type électriques alors que les propriétés concernent des valeurs numériques.

Une méthode fort simple pour s'affranchir de cette sensibilité à l'environnement est sa suppression pure et simple (en d'autres termes E est considéré comme vrai en toutes circonstances). C'est la méthode retenue par De Kleer [De Kleer 86] dans G.D.E. (voir plus loin). Cette première extension de la proposition précédente s'écrit :

$$\neg F \Rightarrow P \quad (2)$$

Remarquons que cette proposition ne permet toujours pas de disculper un composant ( $\neg F$  est indémontrable).

Une autre modification, nettement plus aventureuse mais pleine d'intérêt, a été donnée par Van De Velde [Velde 85, 88]. Elle consiste en :

$$(\neg P \wedge E) \vee (P \wedge \neg E) \Leftrightarrow F \quad (3)$$

Cette proposition exprime qu'un composant est défectueux dès lors qu'il ne produit pas la propriété prévue, dans un environnement correct ou inversement, qu'il produit cette propriété dans un environnement pourtant incorrect.

L'intérêt de cette seconde partie peut être illustré par l'exemple suivant : les phares d'une voiture sont alimentés par la batterie ; si cette dernière est défectueuse, on ne doit pas constater un fonctionnement normal des phares. On sait qu'il n'existe pas dysfonction des phares qui puisse les faire briller en l'absence d'alimentation électrique.

En d'autres termes, lorsque les composants possèdent déjà une certaine complexité interne, ils ne sont pas capables de corriger une panne antérieure dans une chaîne causale. Le principe de suspension de contraintes, dans sa forme la plus pure, renonce à l'usage de cette information.

Nous nous sommes livrés à un travail de comparaison des propositions (1) et (3) et nous avons abouti au résultat suivant : il est possible de passer de (1) à (3) en ajoutant des propositions élémentaires que l'on peut interpréter de manières diverses. La différence brute s'exprime par :  $(F.\neg E.\neg P) \vee (\neg F.\neg E.P) \vee (F.E.P)$ .

Nous avons réexprimé cette différence par trois propositions (a), (b), (c), telles que  $(1) \wedge (a) \wedge (b) \wedge (c) = (3)$ .

principe de manifestation :	$E \wedge P$	$\Rightarrow \neg F$	(a)
principe de non-correction :	$\neg E$	$\Rightarrow \neg P \vee F$	(b)
principe d'optimisme :	$\neg E \wedge \neg P$	$\Rightarrow \neg F$	(c)

Le principe de manifestation impose que  $(E \wedge F \wedge P)$  soit toujours faux, donc on rejette la possibilité de pannes invisibles.

Le principe de non-correction exprime l'idée que les composants ne peuvent compenser une panne antérieure.

Le principe d'optimisme consiste à disculper un composant dès lors qu'il est en environnement anormal (on attribue  $\neg P$  à un effet de l'environnement et non à une dysfonction du composant).

Nous devons insister sur le fait que toutes ces propositions sont arbitraires ; pour chacune d'elles, existent des modèles (i.e. des problèmes réels tels que, pour une décomposition donnée, elle soient toujours vraies).

Bien entendu, devant une telle situation, on serait tenté de choisir la proposition la plus générale. Pour autant que nous le sachions, cela n'a jamais été fait. L'attitude usuelle serait même plutôt inverse : comment enrichir cette proposition de manière à tenir compte de ce que l'on sait d'utile sur l'objet observé ? L'usage a priori de telle ou telle proposition consiste à introduire un biais dans le système d'une manière comparable à l'introduction du *learning bias* [Utgoff 86] dans les règles de généralisation d'un système d'apprentissage.

### 3.1.3. Langages de qualification usuels

Dans la caractérisation du problème diagnostique que nous avons présentée, le langage de qualification occupe une place prépondérante. Nous avons bien insisté sur le caractère arbitraire de ce langage. Toutefois, force nous est de constater que les langages utilisés dans les formalismes "classiques" s'articulent quasi-exclusivement autour d'un langage propositionnel.

Pour le dépannage de cartes électroniques, dès lors que l'on recourt à l'échange standard, l'identification des composants fautifs est suffisant.

Appelons  $C$  l'ensemble des noms de composants connus du système. Un premier langage de qualification de pannes serait  $\{C, \vee\}$  ( $C$  muni de la disjonction). Ce langage induit l'hypothèse de la panne unique, puisque l'on ne peut décrire la défaillance simultanée de deux composants.

L'usage du langage  $\{C, \vee, \wedge\}$  permet de lever cette hypothèse. On peut alors exprimer une disjonction de pannes multiples (expressions conjonctives).

Si la disculpation de composants est possible (voir paragraphe ci-dessus) on peut élargir encore ce langage par inclusion de la négation :  $\{C, \vee, \wedge, \neg\}$ .

Pour les problèmes médicaux, on remplacera simplement l'ensemble des noms de composants  $C$  par l'ensemble des noms de maladies  $M$ . Du fait de la nature du problème, on se restreint classiquement à l'hypothèse de la maladie unique, parfois étendue à l'association morbide de deux maladies. La pleine puissance de la conjonction n'est, en pratique, jamais requise.

Il est nécessaire de bien comprendre le rôle particulier de la disjonction dans ce problème. En dehors de tout aspect de logique formelle, le problème du diagnostic contient intrinsèquement une disjonction du fait de l'existence nécessaire de pannes distinctes (un objet ne possédant qu'une unique panne concevable n'est pas justifiable d'un processus diagnostique).

D'autre part, la persistance d'une disjonction dans la qualification d'une panne doit être considérée comme un échec partiel du processus diagnostique. Ce processus consiste, d'un certain point de vue, à réduire une expression disjonctive maximale (disjonction de toutes les pannes concevables) à une unique expression conjonctive (la panne diagnostiquée). Si la disjonction subsiste, c'est qu'il existe plusieurs hypothèses de pannes entre lesquelles le système ne peut trancher.

En ce qui concerne cet aspect du diagnostic, l'isomorphisme avec l'apprentissage (et plus exactement la généralisation) est totale [Raoult 87].

Enfin, par analogie avec les moteurs d'inférences "d'ordre zéro", on pourrait parler de diagnostic d'ordre zéro (langage de qualification sans variables). Dans ce cadre, et seulement dans celui-ci, le problème du diagnostic s'apparente fortement avec un problème de classification. Le nombre de classes est le nombre de phrases

possibles du langage. Il peut se ramener au nombre d'expressions non-disjonctives, si l'on accepte une classification partielle (une disjonction de classes possibles). Ce dernier point consiste simplement à déplacer la disjonction au-delà, ou en-deçà, du problème de classification.

Par exemple, une situation donnée peut être partiellement classifiée comme étant "de classe  $C_1$ " ou "de classe  $C_2$ ". Si, à l'inverse, on s'autorise à considérer  $C_1 \vee C_2$  comme un unique classe (ie. la disjonction appartient au langage de qualification), alors la même situation sera classifiée comme étant de classe " $C_1 \vee C_2$ ".

Ces langages, que nous qualifierons de "simples" malgré leur forte combinatoire, perdent de leur intérêt dès lors que l'on désire une description des pannes non réduite à une énumération. Par exemple, un médecin pose classiquement un diagnostic de "tuberculose..." associé à un commentaire "...forme miliaire à bas-bruit". Le passage du diagnostic au pronostic impose déjà une description de pannes en termes de "nom + paramètres". De tels langages existent parmi la multitude de systèmes experts centrés sur telle ou telle application spécifique mais, à notre connaissance, aucun n'a fait l'objet de formalisation bien posée. A l'heure actuelle, aucun autre consensus que celui des langages "d'ordre zéro" ne s'est encore dégagé (à l'exception peut-être de formulations probabilistes difficiles à situer).

Après ces quelques considérations sur un état de l'art plutôt morcelé, nous allons explorer plus en détails des travaux qui, selon nous, représentent le mieux l'état d'avancement actuel de l'approche I.A. du diagnostic. Ils sont représentatifs, tant par les résultats qu'ils permettent d'atteindre, que par les ambiguïtés qu'ils perpétuent. Nous tenterons à ce propos de maintenir un difficile équilibre entre la reconnaissance et la critique.

### 3.2. Raisonnement à partir de la structure

La notion de diagnostic à partir de modèles profonds, ou diagnostic basé sur la connaissance de la structure et le comportement, est due, pour l'essentiel, aux travaux de Randall Davis de 1980 à 1984 (un récapitulatif complet se trouve dans [Davis 84]).

Des travaux du même type ont également été poursuivis par Johan De Kleer [De Kleer 76, 84, 86] et, plus récemment, par Raymond Reiter [Reiter 86].

La conception du diagnostic portée par ces travaux constitue une référence pour une partie de la communauté scientifique IA et ils ont bénéficié d'un développement spectaculaire. De notre point de vue, ils seraient toutefois plus justement dénommés "outils d'identification d'éléments fautifs dans un montage", dénomination d'ailleurs utilisée dans [De Kleer 76]. En effet, ils représentent le noyau central

---

d'un processus de diagnostic automatique, mais demandent encore à être fortement complétés.

L'idée directrice est relativement simple. Soit un montage possédant un ensemble de paramètres mesurables qui permettent de caractériser son état. Soit maintenant un modèle de ce montage constitué par le réseau de contraintes qui lient ces paramètres lorsque ce montage fonctionne correctement. L'usage de ce modèle est double : il définit le bon fonctionnement (satisfaction de toutes les contraintes) ; il fournit un outil de simulation permettant de prévoir la valeur d'un ou plusieurs paramètres.

On constate bien le phénomène précédemment cité relatif à la fusion de la spécification de bon fonctionnement, de la spécification des capacités perceptives et de la base de connaissances nécessaire à l'interprétation.

Soit maintenant un ensemble d'observations, dont nous avons vu qu'elles sont indispensables à l'activation d'un processus diagnostic. Considérons, pour la clarté de l'exposé, que chaque contrainte est associée à un élément (réel ou virtuel) du montage. Projets maintenant les valeurs mesurées dans le réseau de contraintes et propageons ces valeurs en respectant les mécanismes logiques associés aux contraintes ; s'il apparaît des incohérences, le processus "diagnostic" doit produire toutes les combinaisons minimales de contraintes qu'il faut relâcher pour retrouver un état global cohérent. Ces différentes combinaisons seront considérées comme autant de pannes possibles.

La notion de minimalité vient du principe de parcimonie : soient  $n$  éléments constituant une panne possible (c.a.d que la suppression des  $n$  contraintes associées restitue un état cohérent), alors tout sur-ensemble de ces  $n$  éléments sera également une panne possible, mais sans apport significatif.

La relaxation de la contrainte associée à un élément présumé fautif a été présentée formellement dans [Davis 84] sous le nom de *suspension de contrainte* et découle du postulat qu'un élément défaillant est susceptible de lier les paramètres qui le concernent de manière quelconque. En supprimant simplement la contrainte, on ne fait aucune hypothèse quant à ses nouvelles fonctionnalités.

### 3.2.1. Le système GDE [De Kleer 86]

Nous reprenons l'exemple désormais archétypique d'un circuit composé de trois multiplieurs et deux additionneurs (défini dans [Davis 84] et repris par [De Kleer 86]).

Nous modifierons légèrement la présentation faites par De Kleer du système GDE (General Diagnostic Engine) afin de faciliter la comparaison avec les autres travaux existants.

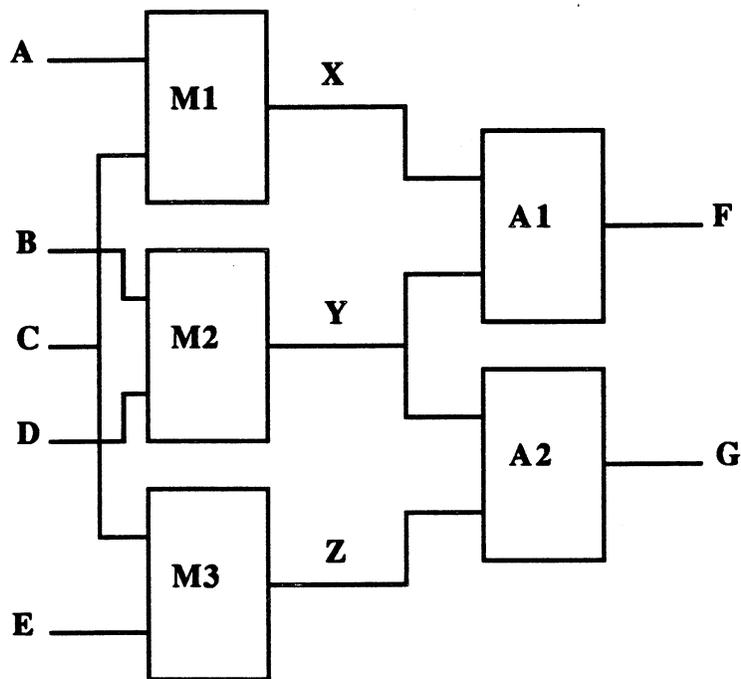


figure 3.2 : Exemple de montage

Il est facile d'exprimer les contraintes liées à chaque élément :

$$\begin{aligned} M1 : A * C - X &= 0 \\ M2 : B * D - Y &= 0 \\ M3 : E * C - Z &= 0 \\ A1 : X + Y - F &= 0 \\ A2 : Y + Z - G &= 0 \end{aligned}$$

C'est volontairement que nous avons exprimé les contraintes ainsi, plutôt que sous la forme  $A * C = X$ , afin de bien marquer que la notion de direction de propagation est absente du modèle. On trouve une fort belle formulation de cette idée dans [Velde 85] où elle est désignée sous le terme d'abstraction de la causalité.

*Abstraction de la causalité : la relation de cause à effet ignore la nature physique de la causalité. Elle exprime seulement qu'il existe une relation entre des propriétés, mais ne propose pas d'interprétation opérationnelle ou fonctionnelle.*

Une interprétation de la contrainte M1 serait ici : le circuit M1 produit X à partir des valeurs d'entrées A et C. Cette interprétation n'est pas contenue dans l'expression de la contrainte.

D'autre part, il est important de noter que ces contraintes sont explicitement nommées. Ces noms vont servir de support à la qualification de panne (i.e. une panne est une liste de noms de contraintes conjointement violées). Par conséquent, toute réécriture de ce système est prohibée.

Ce réseau étant défini, nous pouvons y projeter une observation  $O_1$  telle que celle présentée ci-dessous :

$$\begin{array}{lll} A = 3 & B = 2 & C = 2 \\ D = 3 & E = 3 & F = 10 \end{array}$$

De  $A=3, C=2$ , on peut déduire que  $X=6$ , sous réserve que la contrainte  $M1$  soit satisfaite, ce que De Kleer note selon une formulation conditionnelle :

$$X=6\{M1\}$$

De même  $B=2, D=3$  permet d'établir  $Y=6\{M2\}$ . Puis les valeurs de  $X$  et  $Y$ , par application de  $A1$ , donnent  $F=12\{A1, M1, M2\}$  (il y a cumul des conditions de propagation). Or  $F$  vaut 10, donc la condition implicite  $\{A1, M1, M2\}$  n'est pas réalisée :  $\langle A1, M1, M2 \rangle$  est un *conflit*, c'est-à-dire qu'il n'est pas possible que les trois contraintes citées soient simultanément réalisées, un au moins de ces éléments étant "en panne" (disjonction logique).

La saturation du mécanisme de propagation (voir l'article [De Kleer 86] pour le détail des conditions d'arrêt) produit la liste de tous les conflits détectables pour un ensemble donné d'observations. De cette conjonction de conflits, il est alors facile d'extraire les *candidats*. La correspondance entre conflits et candidats est celle qui existe entre les termes d'une forme normale conjonctive et ceux d'une forme normale disjonctive (cf paragraphe 3.1). Donc, l'originalité de ces travaux ne réside pas dans les propriétés unissant conflits et les candidats, propriétés qui découlent directement de l'algèbre de Boole. Le point essentiel réside dans la technique qui permet de propager les valeurs conditionnelles en évitant les bouclages et les branches infinies.

L'ensemble du processus peut être compris comme une méthode de résolution partielle, par des techniques de type ATMS (Assumption Truth Maintenance System), de systèmes d'équations. Ceci sera utile pour comprendre la similitude des travaux de De Kleer et R. Reiter. Anticipons donc sur le paragraphe suivant en définissant un prédicat  $AB(e)$  qui est vrai sur un élément  $e$  si ce dernier est anormal (défectueux). Il est alors facile de réécrire le problème précédent sous la forme :

$$\begin{array}{l} \text{axiomes : } \neg AB(m1) \supset (A * C - X = 0) \\ \neg AB(m2) \supset (B * D - Y = 0) \\ \dots \\ \neg AB(a2) \supset (Y + Z - G = 0) \end{array}$$

$$\text{observations: } A=1 ; B=2 ; C=2 ; \dots G=12 ;$$

Le mécanisme de GDE produit bien des séquences d'affectations des  $AB(e_i)$  dans {vrai,faux} qui satisfont ces équations.

On trouvera dans [De Kleer 87] une méthode due à J. De Kleer et B.C. Williams qui permet le choix des "meilleures" observations. Ce choix est basé sur la recherche d'un minimum d'entropie, avec la supposition que toutes les mesures ont le même coût. Cette entropie dérive des probabilités que chaque contrainte a d'être déféctueuse, modulée par la possibilité de réduire au maximum le nombre de candidats avec une unique mesure. Sans entrer dans les détails techniques de la détermination et de l'utilisation de cette entropie, remarquons que ces méthodes produisent des résultats intuitivement acceptables.

L'exemple présenté dans [De Kleer 87] est celui de la cascade d'inverseurs (cf figure 3.3).

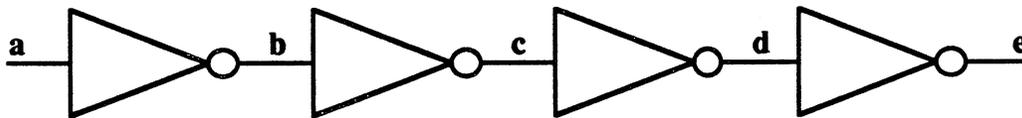


figure 3.2 : Cascade d'inverseurs

Si nous constatons une contradiction, par exemple  $a=1$  et  $e=0$ , alors un au moins des inverseurs est fautif. La recherche du minimum d'entropie, en supposant des probabilités de pannes uniformes, permet d'établir que la meilleure mesure est celle de la variable  $c$ . On retrouve donc une approche dichotomique classique (mesure à mi-chemin entre deux mesures antérieures). De manière complémentaire, si un des inverseurs possède une probabilité de panne élevée, la meilleure mesure aura tendance à se rapprocher de ce circuit.

Ce résultat est donc fort intéressant même si on peut regretter l'usage de méthodes fortement numériques et leurs conséquences habituelles : la perte de tout caractère explicatif. Il serait également important d'étendre ces méthodes au cas où le coût des mesures est extrêmement variable.

### 3.2.2. *Diagnosis from first principles [Reiter 86]*

Etant certainement moins liée aux techniques de type ATMS, cette approche se veut plus générale. Elle se situe toutefois dans la même ligne directrice, à savoir l'identification d'éléments fautifs.

La spécification de bon fonctionnement est fournie directement sous la forme d'un système d'équations comportant, entre autres, des prédicats "d'anormalité"

$\text{AB}(x)$ . Les équations ne sont pas nommées, à l'inverse de l'approche GDE, et ce sont ces prédicats qui seront la base de la qualification<sup>1</sup>.

Par exemple, un additionneur  $X$  sera décrit en calcul des prédicats du premier ordre par la proposition suivante :

$$\text{ANDG}(X) \wedge \neg \text{AB}(X) \supset \text{out}(X) = \text{and}(\text{in1}(X), \text{in2}(X))$$

Ceci se traduit par : un additionneur fonctionnant normalement (not-ABnormal) réalise une fonction liant ses paramètres  $\text{in1}(X)$ ,  $\text{in2}(X)$  et  $\text{out}(X)$ . Les "éléments" (réels ou virtuels, nous le rappelons) sont explicitement connectés entre eux par des équations d'égalités entre paramètres (en lieu et place des variables partagées de GDE). Les observations sont des assignations de paramètres à des valeurs données. Aux équations génériques, telles que celle présentée ci-dessus, il faut donc ajouter un ensemble de formules de natures diverses. Par exemple :

description d'un additionneur A1 :	$\text{ANDG}(A1)$ ;
connexion A1/M1 :	$\text{in1}(A1) = \text{out}(M1)$ ;
observation :	$\text{in1}(A1) = 10$ .

Le problème est alors de trouver des affectations des  $\text{AB}(x_i)$  telles que l'ensemble des formules forme un tout cohérent. Une différence importante avec l'approche GDE est l'usage explicite d'un processus d'abstraction exprimé dans [Velde 85] : l'abstraction des composants.

*Abstraction des composants : il n'existe pas d'information spécifiant quelle propriété décrit un certain composant du système réel. Donc les primitives du modèle sont les propriétés et non les composants.*

Dans cette approche, un prédicat d'anormalité peut figurer dans plusieurs équations, une équation peut comporter plusieurs  $\text{AB}(x)$ , etc. La tentation d'imaginer un isomorphisme entre le réseau de contraintes et une structure physique disparaît totalement. L'utilisateur a ici toute latitude pour écrire un système d'équations. Un circuit réel peut être associé à plusieurs équations ; une équation peut faire référence à plusieurs circuits. En d'autres termes, les équations du système ne sont pas individualisées. Toute forme de réécriture respectant le CP1 sera sans conséquence du point de vue de la référence de bon fonctionnement. C'est d'ailleurs cette propriété qui permet de faire appel à un démonstrateur. Le fait qu'il ne soit pas nécessaire de personnaliser les équations est dû à l'usage explicite des prédicats  $\text{AB}(X)$ .

Une qualification n'est pas une conjonction de noms d'équations, comme c'est le cas de l'approche GDE, mais une conjonction de  $e_i$  tels que  $\text{AB}(e_i)$ . Il y a donc une plus grande dissociation entre les formules et la représentation des noms de pannes.

<sup>1</sup>plus exactement une panne sera une liste d'éléments  $x_1, x_2, \dots, x_n$  tels que  $\text{AB}(x_1), \text{AB}(x_2), \dots$ .

Les équations étant posées, R. Reiter suppose qu'il dispose d'une fonction TP prenant comme paramètres :

- SD : la description du système (un ensemble de formules) ;
- COMPONENTS : une liste de termes  $x_i$  pour lesquels on postule  $\neg AB(x_i)$  ;
- OBS : un ensemble de valeurs observées.

Cette fonction rend une liste de termes  $x_j$  tels que, si l'on introduit l'axiome  $AB(x_j)$  dans le système global (SD + COMPONENTS + OBS), alors ce système est cohérent (relativement au CP1). Si cela n'est pas possible, elle rend " $\sqrt$ ". En d'autres termes, cette fonction calcule un conflit basé sur le postulat que les composants de COMPONENTS fonctionnent correctement, ou "conflit conditionné par une liste d'éléments excusés".

L'essentiel du travail de R. Reiter consiste à démontrer formellement que, par des appels successifs de TP, il est possible d'obtenir tous les candidats (que l'on ait affaire à des pannes simples ou multiples). De plus, il propose une méthode fort astucieuse afin de limiter au maximum le nombre d'appels de cette fonction qu'il qualifie lui-même de "probablement très coûteuse". Cette méthode découle assez directement des propriétés du développement booléen : une certaine séquence d'appels de TP retournant systématiquement  $\sqrt$  signifiera que l'on a atteint le maximum d'informations déductibles de OBS. Cette considération, ajoutée à des notions de *value-caching*<sup>1</sup>, permet de réduire très fortement le nombre d'appels de TP.

On est ici en présence d'une méthode de haut niveau qui renvoie les difficultés d'implantation au niveau de l'écriture de la fonction TP.

La comparaison avec l'approche GDE est simple : comme seule une résolution partielle (i.e. l'obtention des conflits) nous intéresse, la totale puissance d'un démonstrateur n'est en fait pas nécessaire. Les méthodes de propagations conditionnelles constituent une méthode de résolution à partir de laquelle on pourrait définir le prédicat TP. Il est toutefois intéressant de noter que l'usage de l'ATMS, comme implantation de TP, éviterait d'avoir recours à un processus de combinaison extérieur (les *H-S trees* de R. Reiter). La supposition de l'existence d'une fonction TP permet donc de spécifier un processus complémentaire, mais la réalisation effective de cette fonction risque de rendre ce processus inutile. Cette approche possède donc un intérêt certain du point de vue formel, mais est peu opératoire pour la conception d'applications.

---

<sup>1</sup>Méthode consistant à mémoriser les requêtes et leur résultats. Si le système reconnaît une requête déjà calculée il fournit le résultat mémorisé et ne relance aucun calcul.

### 3.2.3. Naive causal reasoning for diagnosis [Velde 85]

Présentons brièvement les principes généraux, avant de tenter une comparaison avec les travaux ci-dessus. W. Velde définit un *réseau causal* comme un ensemble  $P$  fini, non-vidé, de noeuds pourvus d'une valeur dans  $\{n, a\}$  (normal, anormal) et reliés par une relation " $-->$ ". Le réseau  $(P, -->)$  peut être vu comme un graphe orienté, que l'on suppose sans boucle ( $-->$  est antiréflexive et sa fermeture transitive également). Une observation est une affectation particulière des éléments de  $P$  dans  $\{n, a\}$  représentée par une fonction  $\S$ . Soit, de plus, la notation " ${}_p$ " pour désigner les prédécesseurs de  $p$  selon  $-->$ , et " $p_{}$ " pour désigner ses successeurs.

Nous reproduisons (figure 3.4) un système très simple associé au réseau qui le représente. Ce système a seulement valeur d'exemple, car il n'a pas le niveau de complexité pour lequel cette approche est intéressante.

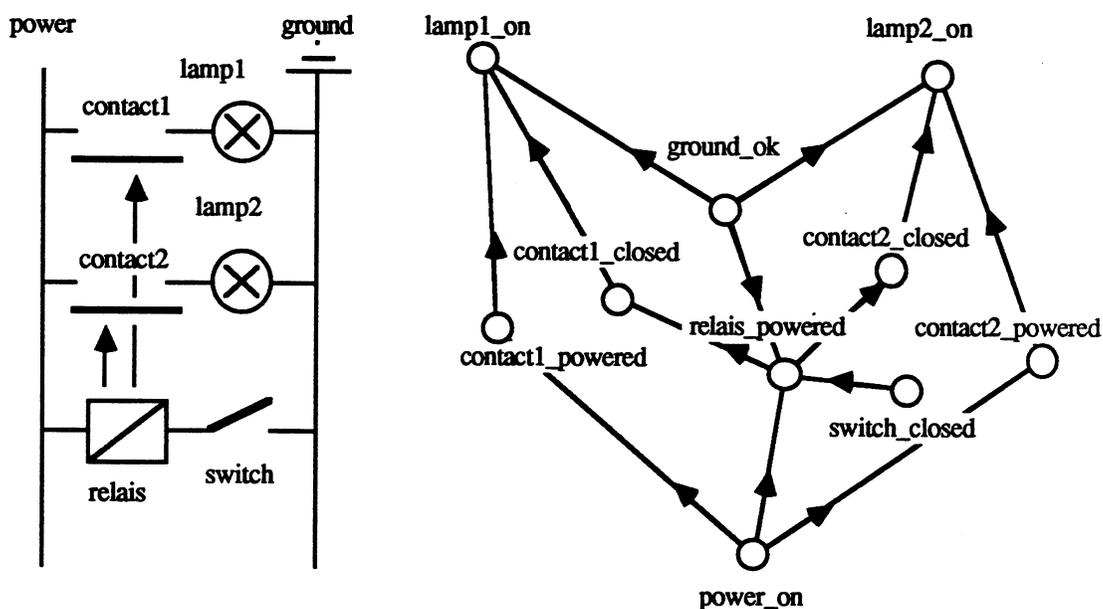


figure 3.4 : Exemple de réseau fonctionnel

Une dysfonction, sur un noeud  $p$ , peut se produire selon deux configurations :

$$\begin{aligned} \forall c \in {}_p : \S(c)=n \wedge \S(p)=a ; \\ \exists c \in {}_p : \S(c)=a \wedge \S(p)=n . \end{aligned} \quad (1)$$

Ce que l'on peut interpréter comme suit :

- avec des entrées normales, le noeud  $p$  produit une sortie anormale ;
- malgré une entrée anormale, le noeud  $p$  produit une sortie normale.

Ce cadre étant posé, W. Velde propose, et prouve formellement, un algorithme de propagation d'hypothèses sur ce réseau qui construit des *explications* (terme équivalent à *candidats*).

Le plus grand mérite de W. Velde est d'avoir clairement exprimé des principes d'abstractions déjà esquissés dans [Kuiper 84]. Nous reproduisons ici les trois principes qu'il définit :

*Abstraction des composants : il n'existe pas d'information spécifiant quelle propriété décrit un certain composant du système réel. Donc les primitives du modèle sont les propriétés et non les composants.*

*Abstraction des propriétés : toute variable correspondant à une propriété est remplacée par une référence qualitative qui exprime si la propriété est satisfaite ou non. Ainsi l'espace des quantités contient seulement les valeurs [ normal (n), anormal (a) ].*

*Abstraction de la causalité : la relation de cause à effet ignore la nature physique de la causalité. Elle exprime seulement qu'il existe une relation entre des propriétés mais ne propose pas d'interprétation opérationnelle ou fonctionnelle.*

Un élément original dans l'approche de W. Van de Velde est l'utilisation d'une relation directionnelle qui donne à l'ensemble du modèle une interprétation causale plus compréhensible. Rappelons que, dans GDE, un additionneur se décrit par " $X+Y -Z =0$ " et non par " $X+Y$  produit  $Z$ ". Dans l'approche présentée ici, il est possible de distinguer deux types de dysfonctions et de postuler qu'un élément, soumis à des conditions anormales, doit produire un état anormal. On rejette implicitement la notion qu'une erreur puisse être corrigée et cela modifie profondément la nature du problème diagnostic. Cette hypothèse est ici admissible (et même quasiment nécessaire), parce que l'on se place à haut niveau d'abstraction et serait peu raisonnable concernant, par exemple, une carte électronique.

### 3.2.4. Forces et faiblesses de ces méthodes

On peut être étonné par la rapidité avec laquelle ces méthodes ont été mises au point, que ce soit à partir d'adaptations de principes utilisés par ailleurs ou à partir d'idées originales. Ces travaux provoquent un vif intérêt et résultent d'une somme d'efforts considérable.

Ces méthodes sont bien formalisées et donc théoriquement bien assises. D'autres part, elles sont très formelles et bien adaptées à une classe assez large de problèmes : tout ce qui relève du diagnostic totalement automatique de systèmes simples. Enfin, elles s'appuient sur des outils génériques (ATMS, démonstrateurs

---

en CP1) dont on est en droit d'attendre par ailleurs des progrès qui réduiraient le coût global de ces approches.

Dans les inconvénients, on peut remarquer que ce sont des méthodes "à un seul niveau", conséquence probable de l'extrême formalisation du problème posé. En tant que telles, elles semblent peu appropriées à l'introduction d'heuristiques plus humainement compréhensibles. Citons, par exemple, la recherche initiale de pannes "simples" suivie, en cas d'échec, par des méthodes plus lourdes. Dans les travaux ci-dessus, toutes les hypothèses sont produites simultanément et aucune procédure n'est décrite qui viserait à ordonner cette production selon des critères de simplicité. Etant donné les méthodes, on peut d'ailleurs douter qu'un tel ordonnancement soit possible sans une remise en cause complète. Un algorithme développé dans cette direction pourra se trouver dans [Geffner 87].

Nous avons également vu que ces approches restreignent implicitement le diagnostic à un problème d'identification de composants fautifs, avec la supposition que chaque composant peut être représenté par une unique contrainte.

De plus, le principe de suspension de contraintes, s'il est valide, peut être inadapté dans certain cas (court-circuits, entrées grillées imposant la valeur d'une sortie, etc). La suppression pure et simple de toute contrainte est une approche "maximaliste" (cf paragraphe 3.1.2) qui peut entraîner une multiplication importante des candidats, mêlant les candidats significatifs et les candidats peu significatifs. Par exemple, sur un chemin assez long, il peut effectivement exister deux éléments défectueux dont le second corrige l'erreur introduite par le premier. En pratique c'est toutefois moins probable qu'une panne simple (cf. principe de persistance [Velde 87]). Dès que le nombre de candidats est potentiellement important, il est primordial de définir un ordre partiel sur les candidats plus strict que celui lié au simple principe de parcimonie (minimalité booléenne).

A notre connaissance, la meilleure analyse de ces techniques a été présentée dans [David 86]. Cette analyse est fondée sur l'expérience, dans le cadre de réalisations en vraie grandeur.

Nous présentons brièvement ci-après les points fondamentaux décrits dans cet article. Selon J.M. David, un système sur lequel on désire pratiquer un diagnostic "de type GDE" doit satisfaire quatre critères :

- 1 : le système peut être décrit comme un réseau de contraintes ;
- 2 : il existe une description dans laquelle un dysfonctionnement global peut être expliqué par un dysfonctionnement d'un "composant" ;
- 3 : une dysfonction élémentaire n'affecte pas la structure du réseau ;
- 4 : les dysfonctionnements que l'on cherche à détecter sont statiques.

Les critères (1) et (2) tendent à montrer que cette approche (reasoning from structure and behavior) est bien adaptée pour les systèmes de type "traitement de flot de données" dont un exemple typique se trouve décrit à la figure 3.2.

Pour les systèmes analogiques, le fonctionnement devrait être décrit comme un équilibre concernant plusieurs composants plutôt que comme une propagation d'états. Selon J.M. David, de tels systèmes sont, en pratique, très délicats à modéliser (voir l'annexe A pour un exemple commenté de modélisation d'un système analogique).

Le critère (3) indique qu'il est impossible de détecter des courts-circuits à partir d'un tel modèle. En effet, la suspension de contraintes ne concerne que les composants du modèle. Dans l'approche GDE, les connexions sont implicites du fait de l'usage de variables (figure 3.2 : M1 et A1 sont "connectés" par la variable X, mais la connexion n'est pas représentée) et, d'autre part, le langage de qualification de pannes ne permet pas de décrire des courts-circuits. Dans le formalisme de R. Reiter les connexions sont plus explicites, il est donc possible d'introduire des éléments virtuels associés aux connexions et de les pourvoir d'un prédicat AB(..).

Par exemple, la connexion "in(A1)=out(M1)" pourrait s'écrire :

$$\neg AB(cn\_A1\_M1) \supset in(A1)=out(M1)$$

En pratique, une méthode aussi brutale rendrait le système totalement inopérant en multipliant les candidats possibles. Chaque hypothèse de dysfonction d'un "élément" donné serait accompagnée de toutes les possibilités de mauvaises connexions.

Enfin, le critère (4) résulte de la considération suivante : les observations que l'on va accumuler à propos de l'objet observé vont, en pratique, s'étagérer dans le temps (chaque mesure possède un temps d'exécution) ; pour que deux mesures soient susceptibles d'être comparées il est nécessaire de supposer que l'objet observé n'a pas varié dans l'intervalle (ni du point de vue de sa structure, ni du point de vue de son état global, panne comprise). Ce dernier point laisse bien peu d'espoir quand à une modélisation réaliste de systèmes biologiques, bien que J. De Kleer et R. Reiter déclarent tous deux que cela est possible. On trouve par exemple dans [Reiter 86] la proposition :

$$Adult(X) \wedge Heart-of(X,H) \wedge \neg AB(H) \supset Rate(H) \geq 70 \wedge Rate(H) \leq 90$$

D'une telle proposition à un modèle utilisable, par exemple, pour la reconnaissance de dysfonctions cardiaques, il existe une distance que nous nous refuserons à franchir sans examiner une application en vraie grandeur.

---

En conclusion, nous tenons à insister sur deux points :

- l'idée d'un modèle profond plus ou moins proche de la structure physique doit être rejetée. Les modèles doivent être conçus à partir des pannes que l'on désire détecter et en fonction des paramètres significatifs mesurables (voir annexe A). En d'autres termes, le "modèle" n'est qu'une manière de mettre en rapport un langage de description et un langage de qualification. Ces deux langages définissent clairement les capacités maximales du système. Le diagnostic peut alors être vu comme une application de l'ensemble des objets descriptibles dans l'ensemble des pannes qualifiables. L'usage d'un modèle est alors assimilable à la recherche d'un mécanisme productif, d'une fonction calculable.

- les approches présentées sont d'un haut niveau d'abstraction et fournissent *a contrario* des outils de bas niveau (des modules d'identification). Ces outils ne constituent pas des systèmes de diagnostic automatique et demandent à être enrichis d'un environnement encore peu spécifié. Ils sont d'un intérêt pratique inestimable sous réserve de ne pas les charger d'un pouvoir de représentation injustifié.

### 3.3. Raisonnement associatif

Cette théorie a été présentée par J.A. Reggia et Y. Peng dans un ensemble de travaux liés à l'usage de l'abduction pour la représentation du raisonnement causal. Ces travaux, réalisés à partir de 1983, décrivent formellement un certain type d'approche fort adapté aux problèmes médicaux. Partant d'un problème pratique selon une approche de type IA, nous avons abouti à des conclusions similaires (voir [Raoult 85]). A l'instar des modèles profonds qui s'imposent de manière naturelle dès lors que l'on traite de cartes électroniques, les modèles de surface (modèles associatifs, modèles de couverture) découlent naturellement de la pratique médicale.

On pourra également trouver des approches de ce type de modélisation dans [Milne 85, Chand. 85].

L'idée de base en est simple : dès que l'on est confronté à un système si complexe que toute modélisation (au sens classique : "représentation des interactions internes") en est impossible, il paraît avantageux de l'envisager d'un point de vue totalement externe. Le système observé sera donc décrit comme une boîte noire soumise à des perturbations (des maladies, des pannes nommées) et qui, sous l'action de ces perturbations, va produire des signes de dysfonctionnement (des symptômes). Connaissant les signes que produit usuellement telle ou telle perturbation, et étant confronté à un ensemble de signes peut-on en déduire la ou les perturbations qui les ont produits ? La pratique médicale (la pathologie) montre que, même si la réponse n'est pas toujours positive, une telle exploitation de l'information est généralement suffisante pour identifier les maladies et/ou déterminer une thérapeutique.

Cette approche est fondée sur le postulat suivant : pour toute perturbation, il existe, soit un signe spécifique (symptôme pathognomonique<sup>1</sup>), soit une combinaison spécifique de signes (i.e. les maladies peuvent se distinguer les unes des autres par la spécificité de leur symptomatologie). Le problème d'identification d'une maladie se ramène alors à l'interprétation d'un groupe de symptômes simultanément présents.

### 3.3.1. Parsimonious covering theory [Reggia 85]

Egalement dénommée "*diagnostic based on set covering model*" [Reggia 83], cette approche peut se présenter de la façon suivante (nous emprunterons la notation de [Peng 86]). Soit :

- un ensemble fini de symptômes : S ;
- un ensemble fini de maladies : M ;
- une relation C de {SxM} ;
- un ensemble S+ inclus dans S.

Ces divers éléments seront pourvus de l'interprétation suivante : S est l'ensemble des noms de symptômes que la perception peut fournir ; M est l'ensemble des noms de pannes pouvant affecter le malade ; la relation C indique qu'un symptôme  $s_i$  est "causalement lié" à une maladie  $m_i$  dès lors que  $C\langle s_i, m_i \rangle$  ; le sous-ensemble S+ représente les symptômes effectivement constatés sur une instance donnée du système (i.e. un système défaillant, un malade).

Ces données peuvent se représenter graphiquement comme sur la figure 3.4 :

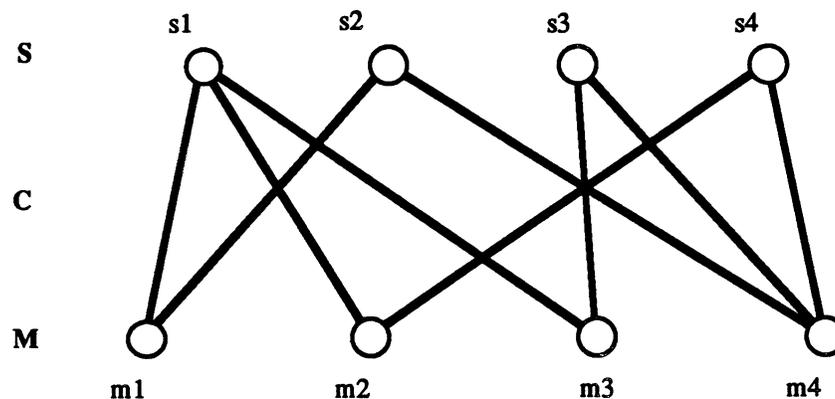


figure 3.4 : Eléments de base d'un modèle associatif

<sup>1</sup> Voir glossaire des termes médicaux

J. Reggia définit deux fonctions, *causes* (causes de) et *man* (manifestation de), qui synthétisent sous une forme ensembliste la relation C.

$$\begin{aligned} \text{causes}(sj) &= \{mi \in M / \langle sj, mi \rangle \in C\}; \\ \text{man}(mi) &= \{sj \in S / \langle sj, mi \rangle \in C\}. \end{aligned}$$

A partir de ces fonctions il décrit le problème diagnostic selon le schéma présenté figure 3.5.

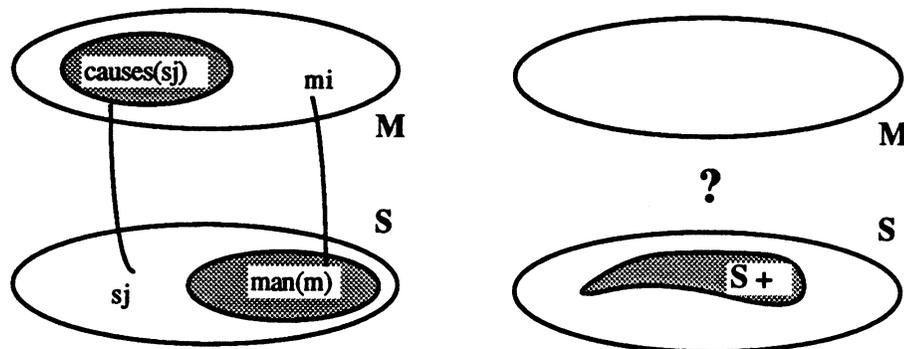


figure 3.5 : Organisation des connaissances et problème

Il s'agit donc de construire une "couverture d'un ensemble de symptômes" à partir de la relation C. On peut y voir une introduction de la notion de contexte. En effet, le but est de fabriquer, à partir d'associations causales hors-contexte (intrinsèques à un symptôme donné), une association causale tenant compte de la présence simultanée de plusieurs symptômes. S+ est alors considéré comme la représentation d'un contexte.

Du fait de l'interprétation causale qui est faite de l'ensemble du problème, il est possible de définir des propriétés que peut, ou doit, vérifier la couverture construite ; c'est ce que fait Y. Peng [Peng 86] qui propose les critères suivants :

(1) *Restriction à une panne unique* : une couverture  $M_i$  de  $S+$  est une explication si elle ne contient qu'un unique désordre ;

(2) *Minimalité* : une couverture  $M_i$  de  $S+$  est une explication si elle est de cardinalité minimale par rapport à toutes les autres couvertures de  $S+$ , c'est-à-dire qu'elle contient le plus petit nombre de désordres nécessaires pour couvrir  $S+$  ;

(3) *Non-redondance* : une couverture  $M_i$  de  $S+$  est une explication si elle ne contient aucun sous-ensemble qui couvre également  $S+$ , c'est-à-dire que la suppression d'un élément quelconque entraîne la non-couverture de  $S+$  ;

(4) *Pertinence* : une couverture  $M_i$  de  $S+$  est une explication si elle ne contient que des désordres présents dans l'union de tous les désordres associés à une quelconque des manifestations de  $S+$ .

Note : les propositions (2) et (3) ne sont pas équivalentes. Supposons que  $\{m_1\}$  et  $\{m_2, m_3\}$  soient des explications non-redondantes (ni  $m_2$ , ni  $m_3$ , ne sont suffisantes isolément pour expliquer les symptômes), alors l'application du principe de minimalité éliminera le doublet au profit du singleton.

Le lecteur pourra trouver dans [Peng 86] des méthodes permettant le choix *a priori* d'une combinaison de ces heuristiques afin de découvrir au plus vite l'hypothèse la plus probable (au sens statistique), étant donné la probabilité absolue de chaque désordre ainsi que la force (la probabilité conditionnelle) de chaque association causale. Cette méthode constitue une de ces extensions de l'approche purement Bayésienne dont nous avons brièvement parlées dans notre introduction.

Nous pouvons émettre plusieurs remarques concernant le principe de modélisation ainsi que la méthode d'obtention des hypothèses :

- le modèle du système est un modèle de dysfonctionnement. La seule perception que l'on a du système est limitée aux signes (symptômes) qui sont *a priori* considérés comme des indications de pannes (de maladies). La spécification de bon fonctionnement est définie par l'absence de toute manifestation ;

- l'ensemble  $S+$  est ici considéré de manière implicite comme décrivant *tous* les signes effectivement présents. On suppose qu'il n'existe pas d'autre signe présent mais non encore observé. Ceci revient à dire que le choix de nouvelles observations n'est pas posé, une fois de plus. Nous proposerons des méthodes permettant de lever cette limitation ;

- toutes les hypothèses engendrées à partir de la notion de couverture (selon C) sont indiscernables entre elles. Si le réseau (S,M,C) est tel que ces hypothèses sont trop nombreuses, il est indispensable de rajouter de l'information. Y.Peng y ajoute une dimension statistique. Nous ferons de même avec un abord entièrement symbolique (fréquence, gravité, coût, etc).

Reprenons le schéma de la figure 3.5. Il définit une forme fixe de représentation des connaissances initiales qui est relativement pauvre (S,M,C + coefficients). L'approche que nous défendrons au chapitre suivant revient à prendre le problème dans l'autre sens. Nous partirons d'un ensemble de diagnostics, posés *a priori* par un expert, pour essayer de construire une structure de données plus diversifiée (d'un type globalement semblable à  $\langle S,M,C \rangle$ ) qui permette de reproduire les diagnostics imposés. Historiquement, notre approche fut originale et ce n'est que par la suite que nous avons pu constater à la fois sa proximité et son opposition par rapport à celle présentée dans [Reggia 83].

---

Bien que l'abord statistique soit parfois plus opératoire qu'une approche entièrement symbolique, nous le rejeterons a priori comme ne correspondant pas au champ de notre recherche présente.

### **3.4. Récapitulatif sur les travaux existants**

Nous venons de détailler les principes de base des deux pôles principaux autour desquels se situent la plupart des travaux sur le diagnostic. Malgré les différences techniques indéniables qui existent entre ces deux approches, on ne peut que constater une grande similitude conceptuelle.

La spécification de bon fonctionnement est donnée, soit par un ensemble de contraintes (le système observé est correct si toutes les contraintes sont satisfaites), soit exhaustivement comme une liste de manifestations pathologiques (le système observé est correct si aucune manifestation n'est présente). Dans les deux cas il subsiste la même difficulté pour prouver le bon fonctionnement. Pour le système GDE, si un ensemble limité d'observations n'engendre aucun conflit, on pourrait supposer, ou non, que le système est correct. De même dans une approche Set Covering Model, si parmi tous les symptômes examinés aucun n'est "effectivement présent", alors on pourrait supposer que le patient n'est pas malade. Dans les deux cas, il reste nécessaire d'effectuer toutes les observations possibles (toutes les mesures) avant de pouvoir certifier le bon fonctionnement.

La détermination des hypothèses de pannes se fait dans les deux cas via un ensemble de connaissances a priori associées à des mécanismes d'exploitation ad hoc. Ces connaissances sont intimement liées aux informations dont nous disposons relativement au système considéré (connaissance des mécanismes internes ou connaissance plus expérimentale décrivant les comportements possibles).

Par exemple, si l'on connaît des lois de propagation à l'intérieur du système observé on peut décider d'utiliser ces lois pour guider l'interprétation. Il sera alors nécessaire de représenter ces lois et d'y associer un mécanisme d'évaluation. A l'inverse, si l'on connaît des associations directes entre symptômes et pannes on peut exprimer ces informations sous une forme "symptôme <-- cause --> maladie" et définir un mécanisme d'exploitation de ces relations.

## 4. Système de propagation de propriétés

Notre propos est de définir un formalisme applicable au diagnostic. Etant donné la généralité du problème, nous avons vu qu'il sera difficile de définir un unique système formel compatible avec les deux pôles du diagnostic ("modèles profonds" et "modèles associatifs"). Le système que nous définissons ci-dessous peut être vu comme une application entièrement symbolique de l'approche de J. Reggia. A ce titre, il peut être considéré comme une approche "associative".

Il va de soi que ce formalisme existe par lui-même et serait susceptible de supporter d'autres interprétations. Par exemple, il présente de fortes similitudes avec le calcul des prédicats du premier ordre (CP1 sur des ensembles finis) appliqué à une structure de type "schéma" (*frames*). Nous adopterons donc une démarche directe en présentant initialement le formalisme puis, dans un second temps, de quelle manière nous le rattachons au problème diagnostic.

Nous tenterons également une évaluation de sa pertinence. Pour ce faire, nous établirons une comparaison avec d'autres systèmes comme les règles de production (moteur d'inférence d'ordre 0) et le système de J. Reggia. Enfin, nous définirons une notion de complétude d'un système de diagnostic et nous montrerons que, relativement à cette définition, notre système de propagation de propriétés est complet.

Ce formalisme a été effectivement implanté dans le cadre de l'application GBDI (Gestion de Base de Données Intelligente). La seconde partie de ce rapport sera consacrée à GBDI en tant que générateur d'applications. Nous y décrirons de nombreux aspects pratiques : spécificités d'implantation ; évaluation de la combinatoire ; vérification de la cohérence ; méthodes de développement d'une application précise ; etc. Dans un premier temps, nous nous intéresserons uniquement à une description formelle.

### 4.1. Formalisme de propagation de propriétés

Afin de faciliter la lecture de ce qui suit, pour des lecteurs peu habitués à ce genre d'expressions (des médecins par exemple), nous avons inséré quelques "apartés" qui anticipent sur les paragraphes suivants. Ces apartés peuvent être omis en première lecture.

Le formalisme est le suivant :

Soit, tout d'abord, un ensemble fini d'objets  $O = \{o_1, o_2, \dots, o_n\}$ .

Nous allons définir plusieurs classes d'applications sur  $O$  auxquelles nous donnerons le nom de "propriété". Ces applications sont également en nombre fini.

Nous distinguerons :

- les propriétés statiques :  $SP_1, \dots, SP_p$ , applications de  $O$  dans  $\{\text{vrai, faux}\}$  ;
- les propriétés binaires :  $BP_1, \dots, BP_q$ , applications de  $O \times O$  dans  $\{\text{vrai, faux}\}$  ;
- les propriétés initiales :  $IP_1, \dots, IP_r$ , applications de  $O$  dans  $\{\text{vrai, faux}\}$  ;
- les propriétés propagées :  $PP_1, \dots, PP_s$ , applications de  $O$  dans  $\{\text{vrai, faux}\}$ .

Les propriétés statiques, binaires et initiales sont données a priori (par exemple, en extension). Chacune des applications  $PP_1, \dots, PP_s$ , ou plus simplement chacun des ensembles  $e_{PP_i} = \{x / PP_i(x) = \text{vrai}\}$  correspondant est défini par une formule, respectivement  $F_{pp1}, F_{pp2}, \dots, F_{pps}$ . Autrement dit, nous imposons que les propriétés propagées soient définies en compréhension.

Nous utiliserons le terme de "propriété" pour parler des applications ci-dessus (unaires et binaires). Le terme "classe" serait néanmoins plus adapté pour les ensembles associés aux propriétés unaires ( $SP$ ,  $IP$  et  $PP$ , à l'exclusion des  $BP$ ). Comme, à l'implantation, nous uniformiserons tout ceci avec des relations (pointeurs), nous avons préféré conserver ce terme peu connoté de "propriété".

Situons rapidement le rôle de ses différentes applications. Prenons un objet du domaine médical comme "radio-pulmonaire". Cet objet possède la propriété d'être un examen ; c'est une caractéristique intrinsèque et statique. Nous pourrions représenter cela par une propriété statique :

$EXAMEN(\text{radio-pulmonaire}) = \text{vrai}$

Autre exemple : la radiographie pulmonaire permet de détecter la tuberculose. Nous avons ici une propriété binaire statique unissant deux objets :

$DETECTE(\text{radio-pulmonaire, tuberculose}) = \text{vrai}$

Supposons maintenant que, pour un malade donné, cet examen a déjà été pratiqué (information entrée par l'opérateur). L'examen acquerra donc une nouvelle propriété que nous avons dénommée "propriété initiale" :

$DEJA-FAIT(\text{radio-pulmonaire}) = \text{vrai}$

Les propriétés propagées sont des propriétés non-intrinsèques aux objets (comme le fait d'être "possible", pour une maladie) qui sont déterminés par le système, en fonction des formules  $F_{ppi}$  :

$POSSIBLE(\text{tuberculose}) = \text{vrai}$

Si nous ne disons rien de plus sur les formules  $F_{ppi}$ , il est clair que ce formalisme est bien trop général pour avoir un intérêt quelconque. Nous avons

donc restreint l'expression des  $F_{ppi}$  à une formulation de type ensembliste (voir la grammaire ci-dessous) faisant intervenir :

- les ensembles  $e_{SP_i} = \{x / SP_i(x)=vrai\}$  ;
- les ensembles  $ch_{BP_i}(o_j) = \{x / BP_i(x,o_j)=vrai\}$  ;
- les ensembles  $e_{IP_i} = \{x / IP_i(x)=vrai\}$ .

La grammaire définissant les formules bien formées  $F_{ppi}$  est la suivante :

```

Fppi -->      ensemble

ensemble -->    e_SPi | e_IPi |
                  intersection( ensemble, ensemble ) |
                  union( ensemble, ensemble ) |
                  exclusion( ensemble, ensemble ) |
                  tout_intersection1( ensemble, ch_BPi) |
                  tout_union( ensemble, ch_BPi) |
                  Si val_booléenne Alors ensemble Sinon ensemble.

val_booléenne -->(val_entière op_entier val_entière) |
                  (val_booléenne op_booléen val_booléenne) |
                  NOT val_booléenne.

val_entière -->  constante entière |
                  cardinal( ensemble ) |

op_booléen -->  AND | OR .

op_entier -->   = | ≠ | < | > .
  
```

Selon la philosophie de formalisme, si l'on désire distinguer explicitement un objet "except" des autres objets du domaine, il est nécessaire de nommer une propriété qui ne soit vraie que pour cet objet :

EXCEP-PROP(excep) = vrai

En d'autres termes, le caractère exceptionnel d'un objet n'est pas contenu dans son nom mais dans le fait qu'il est le seul (à un instant donné) à posséder une certaine propriété. Nous verrons par la suite que cette caractéristique est très précieuse pour toute extension du domaine. Par exemple, l'objet "except" qui paraissait initialement exceptionnel peut, suite à un ajout de nouveaux objets, apparaître comme un simple représentant d'une classe d'objets caractérisée par EXCEP-PROP. Quoi qu'il en soit, aucune modification des mécanismes de propagations ne sera nécessaire, puisque l'on ne faisait déjà référence qu'à la propriété et non aux objets. Nous reviendrons sur ce point.

Tout ceci signifie en particulier qu'il est interdit de faire explicitement référence à un objet donné pour inférer la valeur d'une propriété propagée. La détermination des propriétés propagées ne peut se faire que

<sup>1</sup> cette fonction effectue l'intersection des ensembles  $ch_{BP_i}(o_j)$  pour tous les objets  $o_j$  appartenant à  $ens\_calculable$ .

selon d'autres propriétés. Par exemple, supposons que l'on dispose de deux objets "chat" et "chien", ils ne pourront être distincts pour le système que s'il existe au moins une propriété SP, IP ou BP qui est vraie de l'un et non de l'autre. Dans le cas contraire, ils sont indiscernables, et toute propriété propagée qui sera vraie de l'un sera également vraie de l'autre. Un exemple de formule est présenté au paragraphe 4.3.

Nous emploierons souvent le terme "interprétation" dans ce qui suit. Il ne s'agit aucunement de l'interprétation formelle (isomorphisme d'un système formel dans un autre système formel) mais bien de "l'interprétation intuitive". Par exemple, lorsque nous dirons que  $SP_1$  peut être interprétée comme une propriété statique unaire, nous ferons l'économie d'un cadre formel pour définir ce qu'est une propriété statique unaire. En réalité, nous pourrions nous dispenser totalement de toute interprétation, qu'elle soit intuitive ou formelle, et dériver du formalisme une application effective. Cette application sera utilisée par un être humain qui lui associera une "signification" à sa convenance.

Si nous avons adopté, cette attitude minimale, l'exposé de ce qui suit aurait été relativement sec et de peu d'intérêt, laissant systématiquement dans l'ombre le "pourquoi" des différents éléments du formalisme et de leurs interactions. Nous avons donc pris le parti de nous placer fréquemment en-dehors du formalisme pour mieux expliquer ses raisons d'être.

## 4.2. Mécanisme d'évaluation

Etant donné le "langage" défini ci-dessus, il nous faut encore préciser les mécanismes d'évaluation associés à ce langage, son fonctionnement effectif et son interaction avec un utilisateur.

Sont données de façon permanente, statique et non-modifiable, les éléments suivantes : toutes les applications  $SP_i$ , toutes les applications  $BP_i$ , toutes les formules  $F_{ppi}$ . Nous regrouperons cette somme d'informations sous le sigle BD (base de données).

Un système S de propagation de propriété est alors décrit comme la fonction qui, à la base de données BD, fait correspondre une fonction  $S_{BD}$  :

$$S(BD) = S_{BD}$$

La fonction  $S_{BD}$  décrit un système effectivement activable par un utilisateur (i.e. un système productif). L'utilisateur fournit une liste de valeurs pour les applications  $IP_i$  et le système leur fait correspondre, par évaluation des  $F_{ppi}$ , la liste des valeurs des applications  $PP_i$  (c'est-à-dire des ensembles  $e_{PP_i}$  correspondants).

$$S_{BD}(IP_1, IP_2, \dots, IP_k) = (PP_1, PP_2, \dots, PP_v)$$

De ce point de vue, S est une sorte de générateur de système de diagnostic qui produit une fonction calculable des propriétés initiales vers les propriétés propagées.

### 4.3. Exemple de formule

Afin de fixer les idées, nous présentons ci-dessous un bref exemple de formule  $F_{ppi}$ .

Soit des objets représentant des symptômes et des maladies. Les symptômes vérifient la propriété statique SYMPTOME(x), les maladies vérifient la propriété statique MALADIE(x). De plus, symptômes et maladies sont unis par une propriété binaire CAUSE(s,m). Enfin un certain nombre de symptômes ont été effectivement constatés par l'opérateur, ce qui va se traduire par l'affectation d'une propriété initiale OBS(x) à un certain nombre d'objets (en l'occurrence, uniquement des symptômes).

A partir de ces propriétés, on désire établir : les maladies évoquées, les maladies possibles.

Nous supposons que les maladies évoquées sont "toutes les maladies en rapport avec un quelconque des symptômes constatés". Ceci se traduit par :

$$e\_MAL\_EVOQ := \text{tout\_union}(e\_OBS, ch\_CAUSE)$$

Nous supposons que les maladies possibles sont "toutes les maladies qui sont des causes communes des symptômes constatés". Ceci se traduit par :

$$e\_MAL\_POSS := \text{tout\_intersection}(e\_OBS, ch\_CAUSE)$$

Ces deux formules sont extrêmement simples (la formule se réduit à l'emploi d'une unique fonction de base, sans pré-condition) et peuvent fournir des résultats étranges. Par exemple  $e\_MAL\_POSS$  est vide dès que les symptômes sont contradictoires entre eux. La mise au point d'un système de diagnostic, pour un domaine donné, implique donc une recherche potentiellement très ardue des *bonnes* formules (i.e. celles qui fournissent un résultat comparable au résultat de l'expert).

Nous pouvons voir sur cet exemple que la complexité potentielle des formules est directement liée au nombre de propriétés SP, BP et IP disponibles. Si nous définissons ici une propriété statique unaire GRAVE(x) nous pourrions coder la proposition "les maladies possibles sont les causes communes des symptômes observés, si ces causes sont trop nombreuses (> 5), on se limitera aux seules causes communes du type maladies graves, sous réserve qu'il en existe". Tout ajout de propriété augmente la complexité potentielle de la rédaction des "bonnes"

formules. En contrepartie, cela augmente la qualité de la reproduction du comportement de l'expert. Un compromis devra donc être fait.

#### 4.4. Application au diagnostic

Nous avons esquissé une interprétation particulière des différentes composantes du formalisme. Cette interprétation n'est pas limitative, mais elle correspond au domaine d'application qui nous a conduit à créer tout ceci. D'autre part, cette interprétation nous sera indispensable pour évaluer la "capacité à représenter" de ce formalisme.

Dans ce paragraphe, nous reprenons d'abord en détail les interprétations que nous pouvons associer aux différentes applications  $SP_i$ ,  $BP_i$ ,  $IP_i$  et  $PP_i$ .

- Les  $SP_i$  dénotent l'existence de propriétés statiques unaires relatives aux objets du domaine. Si, par exemple, les objets du domaine sont des animaux, il peut être utile de disposer de la liste des mammifères présents parmi ces animaux. Cette liste peut être considérée comme un *concept* [CSMA 88] ou une *propriété* (le concept de mammifère, vu comme une classe, ou la propriété d'être un mammifère, vue comme un prédicat). Si nous adoptons un mode de représentation de type "objet" (au sens de SRL [Wright 83]), nous pouvons en effet exhiber deux représentations équivalentes du point de vue formel :

Chat	Mammifère
type : objet de O	type : propriété sur O
mammifère : vrai	concerne : (chat, chien, ...)
...	...

En d'autres termes, la dualité qui existe entre les objets et les propriétés qu'ils supportent (ou par lesquelles ils sont supportés) peut être orientée dans le sens voulu. On peut identifier les éléments de O à des objets (des instances d'un schéma) pourvus d'un champ booléen indiquant s'ils possèdent, ou non, une certaine propriété. A l'inverse, on peut s'appuyer sur des objet-propriétés possédant un champ dont la valeur est une liste de termes de O.

Nous pouvons appeler le premier choix "accès par les objets" et le second "accès par les propriétés". L'implantation GBDI est alors résolument orientée, pour les  $SP_i$ , sur l'accès par les propriétés. Pour des objets de base qui sont des maladies, des examens ou des symptômes, les propriétés sont, par exemple "être un symptôme observé", "être une maladie possible", etc. Chaque propriété sera représentée par une classe d'objets de O.

- Les  $BP_i$  peuvent être perçues comme représentant des propriétés statiques binaires unissant deux objets du domaine. On pourrait citer la propriété, pour un animal, d'être le prédateur d'un autre. Un animal n'est pas intrinsèquement un prédateur, il n'est prédateur que par rapport à un autre animal.

Là encore, on pourrait envisager un accès à partir des objets ou des propriétés.

Chat	Prédateur-de
type : objet de O	type : propriété sur OxO
prédateur-de : (oiseau, poisson,...)	concerne : ( chat, oiseau),
...	(chat,poisson),
	...)
	...

L'implantation GBDI est cette fois-ci orientée, pour les  $BP_i$ , sur l'accès par les objets (cf. partie 2, paragraphe 2.1).

Dans un domaine médical, ces propriétés uniront, par exemple, un examen avec les symptômes qu'il permet de détecter, un symptôme avec les maladies qu'il évoque, etc.

- Les  $IP_i$  traduisent également l'existence de propriétés unaires, mais ces dernières, à la différence des  $SP_i$ , ne sont pas intrinsèques aux objets étudiés et sont donc susceptibles de changer au cours du temps. Par exemple, on pourrait affecter la propriété "être-vu" à tous les animaux visuellement perçus par un observateur à un instant donné. Ces propriétés constituent les paramètres d'entrée du système (le langage d'observation).

La propriété initiale de base concernant le diagnostic "associatif" sera un ensemble de symptômes observés (équivalent du S+ de J. Reggia).

- Enfin les  $PP_i$  peuvent représenter des propriétés unaires susceptibles d'être inférées (selon les  $F_{ppi}$ ) des autres données précédemment citées. Par exemple, tout animal, prédateur de mammifères de plus de 10 Kg, étant en vue peut être considéré comme dangereux. Ces propriétés sont bien définies en compréhension, n'étant pas intrinsèques aux objets connus et ne pouvant faire l'objet d'une perception directe. Le fait qu'un objet donné possède, ou non, cette propriété doit être inféré. Ceci revient à déterminer si, oui ou non, cet objet appartient à la classe des objets possédant la dite propriété. Le plus important est le fait que ces propriétés ne sont définies que relativement à d'autres propriétés (unaires et binaires), toute référence explicite à un objet de O étant interdite.

On voit ici que cette interprétation sous-entend que l'on s'intéresse davantage à l'existence d'objets possédant une propriété donnée, voire à tous les objets la possédant plutôt qu'aux propriétés particulières d'un objet initialement fixé (cf. paragraphe 4.5.3).

L'ensemble du processus d'évaluation des  $F_{ppi}$  peut être perçu comme la volonté de convertir une expression en compréhension de certains concepts, ou propriétés, en une expression en extension (une liste d'objets), sachant bien sûr que cette conversion n'est possible qu'une fois figées les variables du problème (les valeurs des propriétés initiales, paramètres de la fonction  $S_{BD}$ ).

On retrouve bien une approche de type raisonnement associatif (diagnostic à partir de symptômes), dès lors que l'ensemble  $O$  contient un sous-ensemble d'objets dénommés "symptômes", un sous-ensemble d'objets dénommés "maladies" et qu'il existe une application  $IP_i$  décrivant les symptômes constatés.

Si besoin est, nous pouvons compléter ces données de base par des informations de fréquences (ensembles de type  $e_{SP_i}$  de maladies dites "fréquentes", "courantes" ou "rares") voire de gravité (ensembles pré-définis de "maladies graves", "maladies bénignes", etc). Ces informations permettront de gérer diverses heuristiques utilisées par les médecins.

Enfin, les modes d'associations d'un symptôme avec une ou plusieurs maladies peuvent être représentés par des propriétés binaires variées. Dans l'application GBDI-IPLSYS, présentée dans la seconde partie de cette thèse, on trouve par exemple des propriétés comme "évoque" (tel symptôme évoque telle maladie), "évoque fortement" ou "affirme".

Supposons maintenant que l'on ait ainsi codé l'ensemble des informations nécessaires à la pratique d'un diagnostic donné. Nous disposons donc d'un expert qui nous garantit que les objets et les propriétés associées sont suffisants pour produire un diagnostic correct dans un domaine donné (une classe de pathologies). Nous pouvons alors nommer des applications de type  $PP_i$  comme les "maladies possibles" ou les "examens conseillés". La spécification des méthodes de calcul de ces propriétés, c'est-à-dire le calcul des ensembles  $e_{PP_i}$  associés, introduit alors les connaissances de résolution nécessaires à la pratique du diagnostic.

Si l'expert garantit que les données sont suffisantes, il doit être capable d'exploiter ces données à la main, pas par pas, pour retrouver ses conclusions propres. Par exemple "je pense à la maladie X, or EVOQUE(X,S) et S est un symptôme constaté. D'autre part, EVOQUE(Y,S) n'est pas gênant car la maladie X est grave, en effet GRAVE(X), alors que Y est bénigne, PEU-GRAVE(Y), donc, dans un premier temps, la maladie X a bien le pas sur la maladie Y". La transcription de ce discours, dans la grammaire que nous avons définie, va produire une série de formules permettant le calcul des propriétés propagées. Nous présentons des exemples détaillés dans la seconde partie, chapitre 3.

La figure 4.1 présente les grandes lignes de l'application GBDI qui fera l'objet de la seconde partie de cette thèse. Les symptômes observés permettent tout d'abord de déterminer un ensemble de maladies dites "possibles". A partir de ces dernières on déterminera successivement les symptômes pertinents (qui discriminent les maladies possibles) puis les examens les plus adaptés pour acquérir ces symptômes. On retrouve bien cette idée de "propagation" de propriétés sur des objets de base.

Une telle structure décrit bien un processus diagnostique. Les capacités perceptives du système sont constituées des applications  $IP_i$  disponibles. Il existe autant d'objets descriptibles que de combinaisons d'applications  $IP_i$  (les observations effectuées correspondent à des affectations de valeurs à ces propriétés initiales).

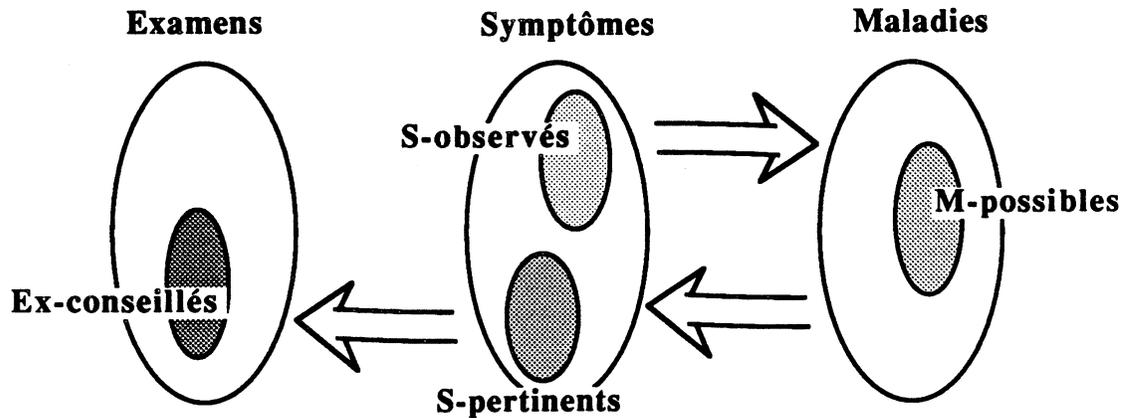


figure 4.1 : diagnostic et propagation

Les informations produites en sortie sont définies par les applications  $PP_i$  calculées. Il existe autant de pannes qualifiables que de combinaisons d'applications  $PP_i$ . Seul un nombre restreint de ces pannes seront effectivement produites par le système (l'image de l'ensemble des objets descriptibles, c.f. paragraphe 3.2). Ces applications peuvent être relatives à des conclusions possibles (hypothèses de maladies) et/ou à des investigations souhaitables, ainsi qu'à toute classe potentiellement utile à l'opérateur. Il faudra produire au minimum ces deux premières classes si l'on veut pouvoir parler de diagnostic tel que nous l'avons caractérisé.

Cette double interprétation (système de propagation de propriétés et système de diagnostic) n'est pas fondamentalement liée au formalisme présenté. Ce formalisme est simple, d'implantation relativement aisée, mais il reste à découvrir l'ensemble des domaines d'applications où il est efficace. En première analyse, on peut estimer qu'il est utilisable dans tous les domaines contenant un grand nombre d'objets, structurés par quelques propriétés statiques (unaires et binaires), et où le problème posé peut se ramener à la détermination de propriétés dynamiques que l'on sait exprimer en compréhension.

#### 4.5. Comparaison avec d'autres formalismes

Nous nous livrons ci-dessous à une comparaison du formalisme proposé relativement à deux autres : les règles de production (ordre 0) et le système de J. Reggia (théorie de la couverture parcimonieuse).

Cette comparaison fera ressortir des inclusions de tel formalisme dans tel autre ; la plupart du temps, cela est simplement dû à l'intégration de la syntaxe d'un système dans l'autre. Ce n'est donc pas ce point qui est réellement important. Par contre nous tenterons d'évaluer les "capacités à représenter" relatives de ces

formalismes. Même si deux systèmes sont formellement inclus (tout ce qui est exprimé dans l'un peut s'exprimer dans l'autre), il peut se trouver que l'un soit subjectivement plus simple que l'autre pour décrire un certain phénomène. Un formalisme donné est donc plus ou moins adapté à la représentation "simple" d'une classe de problème, étant donné une perception subjective, par un être humain, de cette simplicité.

L'évaluation de la simplicité du formalisme décrit ci-dessus est donc un travail assez aléatoire car il repose sur une *compréhension* du système formel et donc sur des considérations subjectives.

Toute évaluation rigoureuse nécessiterait une définition formelle de cette notion de simplicité, une définition de concepts comme la généralité, l'extensibilité, la facilité de mise au point incrémentale, etc. Ce travail dépasse nettement le but que nous nous sommes fixé dans la présente thèse et serait justiciable, à lui seul, d'une thèse relative à l'évaluation des systèmes formels.

#### 4.5.1. Règles de production (ordre 0)

Soit un problème défini par un ensemble fini  $\{A, B, C, \dots\}$  de prédicats connectés par un ensemble de règles décrivant leurs dépendances (exemple : Si A et B Alors C).

Nous allons montrer que notre formalisme permet de reproduire ce système de règles mais également qu'il permet d'en enrichir la représentation du point de vue de l'utilisateur.

Nous pouvons poser :

$$O = \{ a, b, c, \dots \} ;$$

$SP = \{SP_a, SP_b, SP_c, \dots\}$ , où chaque application  $SP_x$  n'est vraie que pour l'objet  $x$  ;

$IP = \{IP_{vrai}, IP_{faux}\}$ , deux applications qui regroupent les objets déclarés "vrais" (resp. "faux") par l'utilisateur ;

$PP = \{PP_{vrai}, PP_{faux}\}$ , deux applications qui regroupent les objets que nous qualifierons de "déductibles par le système global  $S_{BD}$ ".

On observe alors facilement l'équivalence (en terme de quantité d'information) des deux représentations :

- le prédicat A est vrai ;
- $\text{intersection}(e_{SP_a}, e_{IP_{vrai}}) = \{a\}$ .

Dans cet esprit il est facile de transcrire une règle "Si A et B Alors C" dans l'expression de la formule  $F_{ppvrai}$  par :

$$F_{ppvrai} = \text{union} ( \dots , \\ \text{Si } \text{intersection}( e_{SP_a}, e_{IP_{vrai}} ) = \{a\} \text{ et} \\ \text{intersection}( e_{SP_b}, e_{IP_{vrai}} ) = \{b\} \\ \text{Alors } e_{SP_c} \text{ sinon } \{ \} , \\ \dots )$$

L'existence d'un évaluateur nous permettra alors d'obtenir :

$$e_{PP_{vrai}} = \text{eval}( F_{ppvrai} ).$$

Cette formulation du problème reproduit complètement le cas où le réseau de règles fait clairement apparaître les faits initiaux (c'est-à-dire absents de toute partie droite de règle) et où seuls ces faits initiaux font l'objet d'une affectation de la part de l'utilisateur. Dans ce cas, le réseau se décompose en autant d'arbres de déductions qu'il existe de faits non-initiaux. La formule  $F_{ppvrai}$  s'exprime alors comme l'union des expressions décrivant ces arbres ( $n$  expressions pour  $n$  faits non-initiaux, avec uniquement des faits initiaux en partie gauche).

Sous réserve d'accepter l'apparition potentielle de phénomènes d'incohérence et de bouclage, il est possible d'admettre une formulation itérative (l'expression d'une propriété propagée peut alors faire appel à sa valeur antérieure).

Ainsi, les deux règles "Si A et B Alors C" et "Si C et D Alors E" pourraient<sup>1</sup> se traduire par :

$$e_{PP_{vrai}} := \text{union} ( e_{PP_{vrai}} , \\ \text{Si } \text{intersection}( e_{SP_a}, e_{PP_{vrai}} ) \neq \{ \} \text{ et} \\ \text{intersection}( e_{SP_b}, e_{PP_{vrai}} ) \neq \{ \} \\ \text{Alors } e_{SP_c} \text{ sinon } \{ \} );$$

$$e_{PP_{vrai}} := \text{union} ( e_{PP_{vrai}} , \\ \text{Si } \text{intersection}( e_{SP_c}, e_{PP_{vrai}} ) \neq \{ \} \text{ et} \\ \text{intersection}( e_{SP_d}, e_{PP_{vrai}} ) \neq \{ \} \\ \text{Alors } e_{SP_e} \text{ sinon } \{ \} );$$

etc.

<sup>1</sup> s'il était établi que A, B et D sont des faits initiaux cela pourrait s'étendre en :  
Si A et B Alors C    et    Si A et B et D Alors E

Si le graphe est cohérent , pour toute affectation initiale, il existe un et un seul état saturé. La valeur correspondante de  $e_{PP_{vrai}}$  sera obtenue par l'algorithme suivant :

```

n_initial := 0 ;
e_PP_vrai := e_IP_vrai ;
n_actuel := cardinal (e_PP_vrai ) ;
TantQue (n_initial ≠ n_actuel)
    n_initial := n_actuel ;
    ...évaluer toutes les expressions...
    n_actuel := cardinal (e_PP_vrai ) ;
FinTantQue ;

```

Mais ce n'est pas ce point que nous jugeons le plus important. Une formulation ensembliste, telle que définie par notre formalisme, permet de faire disparaître conjonctions et disjonctions. Plus exactement, elle permet de ne conserver que les conjonctions et disjonctions réellement significatives selon une certaine interprétation.

Prenons une règle telle que : Si A ou B ou C ou D ou E Alors H. Il est évident que nous pouvons la reproduire à l'aide d'une formule  $F_{pp}$  disjonctive. Mais considérons maintenant la signification de cette règle dans le domaine d'où elle est extraite. Il est tout à fait possible que cette règle traduise le fait que, s'il existe un objet quelconque de l'ensemble {a, b, c, d, e} qui vérifie la propriété de "véracité", alors on peut effectuer une certaine déduction. Une précondition disjonctive peut être écrite à l'aide d'un quantificateur existentiel.

De ce point de vue, {a, b, c, d, e} peut être compris comme l'expression en extension d'une propriété statique. Représentons cette propriété par  $e_{SP}=\{a,b,c,d,e\}$ . La règle pourra alors être transcrite comme :

Si  $\text{intersection}(e_{SP}, e_{IP_{vrai}}) \neq \{\}$  Alors ...

Cette deuxième forme peut être perçue comme "plus riche" que la forme disjonctive. Elle contient potentiellement l'information : "si A alors H, si B alors H, .. si E alors H et ce pour la même raison  $e_{SP}$ ".

Il peut également se produire que A, B, C, D et E ne soient pas équivalents dans leur relation avec H. Par exemple, il peut exister deux raisons distinctes de faire cette inférence. Notons  $e_{SP_{c1}}$  et  $e_{SP_{c2}}$  ces deux raisons. Ce cas donnerait une formule du type :

Si  $\text{intersection}(e_{SP_{c1}}, e_{IP_{vrai}}) \neq \{\}$  ou  
 $\text{intersection}(e_{SP_{c2}}, e_{IP_{vrai}}) \neq \{\}$   
Alors ...

Donc, l'utilisation d'un langage ensembliste nous fournit un équivalent du quantificateur existentiel, mais, comme nous ne renonçons pas pour autant à l'usage explicite de la disjonction, la puissance de représentation du formalisme est augmentée. Au pire, un agrégat de cas spécifiques sera exprimé sous forme totalement disjonctive. On dispose alors d'un outil de représentation de cette notion de "cas spécifiques/cas général". Il y a équivalence formelle entre ces deux écritures, du point de vue du résultat produit, mais elles ne possèdent pas la même capacité d'expression, du point de vue de l'utilisateur<sup>1</sup>.

On peut tenir le même discours sur une règle "Si A Alors B et C et D et E". L'inférence consiste à affecter une propriété donnée, ici la véracité, à un ensemble d'objets. On peut alors considérer que l'on effectue  $n$  affectations distinctes ou une unique affectation pour tous les objets possédant une autre propriété (la propriété étant définie en extension par  $\{b, c, d, e\}$ ). S'il n'y a qu'une inférence, c'est-à-dire que B, C, E et D sont identiques dans leur relation avec A, alors la représentation ensembliste est préférable. Si chaque affectation découle d'une inférence spécifique il s'agit bien d'un "réelle" conjonction, du point de vue de l'utilisateur, et il est préférable qu'elle soit plus explicitement représentée.

Il est possible de parler *d'inférence de propriété* en cela que l'on manipule préférentiellement des propriétés et non les objets qui les composent. Les préconditions concernent l'existence d'objets vérifiant une combinaison de propriétés ; les inférences consistent en la modification de valeurs de propriétés. Notons que ce mode d'expression permet de manipuler indifféremment des propriétés statiques (des  $SP_i$ ), des propriétés initiales (des  $IP_i$ ), voire, sous certaines réserves, des propriétés dynamiques (des  $PP_i$  précédemment évaluées et figées).

Ce groupement d'objets en des ensembles pertinents dénommés "propriétés" est à rapprocher (sinon à identifier) de la notion de concept utilisée dans de nombreux travaux sur l'acquisition de connaissances. On peut également y voir une manière inhabituelle d'introduire le premier ordre, mais nous préférons nous référer à un concept de propriétés plus qu'à un formalisme strictement logique (un exemple de transformation de formules ensemblistes en expressions de logique du premier ordre est donné au paragraphe 4.5.3. Pour des ensembles finis, les expressions sont formellement équivalentes).

Le caractère "propagation de propriétés" des règles de production (ordre 0) peut être visualisé comme sur la figure 4.2.

On voit bien sur cette figure que les deux propriétés initiales (les faits affirmés vrais ou affirmés faux) se "propagent" jusqu'à saturation en deux propriétés propagées (les faits déduits vrais et déduits faux). Les "formules" qui définissent les mécanismes de propagation sont formés par les règles et le Modus Ponens.

---

<sup>1</sup>La même logique a conduit les informaticiens à inventer nombre de langages plus ou moins expressifs plutôt que se restreindre à la programmation d'une machine de Turing.

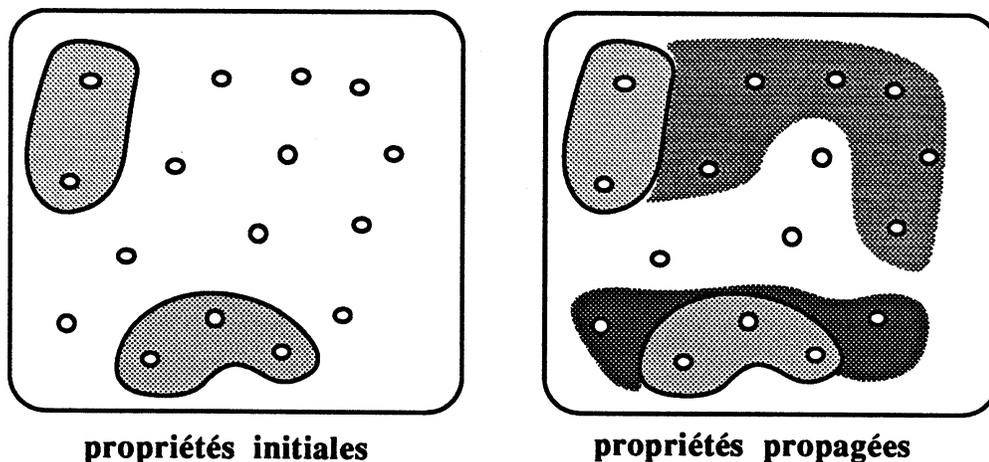


figure 4.2 : règles de production et propagation

Remarquons que nous n'avons pas utilisé jusqu'à présent les applications BPI. Nous allons maintenant établir un parallèle avec la théorie de la couverture qui justifiera leur existence.

#### 4.5.2. Parsimonious covering theory

Rappelons brièvement les différents éléments du formalisme de J. Reggia.

Soit un ensemble fini  $S$  de symptômes, un ensemble fini  $M$  de maladies et une relation  $C$  sur  $M \times S$ . Cette relation exprime qu'un symptôme et une maladie sont "causalement" liés. Tout ceci définit la structure de données initiale avec laquelle on se propose de résoudre le problème suivant : soit un sous-ensemble  $S+$  de  $S$ , représentant une liste de symptômes effectivement constatés chez un malade, peut-on produire un sous-ensemble  $M+$  de maladies qui "expliquent parcimonieusement" les symptômes  $S+$  ?

La notion d'explication peut être exprimée comme une couverture logique par la relation  $C$ . Dans notre formalisme cela pourrait s'écrire :

$O = \{s_1, s_2, \dots, s_n, m_1, m_2, \dots, m_p\}$  ;  
 $BP = \{BP_C\}$  une application de  $O \times O$  dans  $\{\text{vrai}, \text{faux}\}$  équivalente à  $C$  ;  
 $IP = \{IP_{S+}\}$  l'application qui caractérise les symptômes constatés ;  
 $PP = \{PP_{\text{cover}}\}$  l'application définie en compréhension qui définit la "couverture" de  $S+$ .

Si l'on se limite à la notion de "cause commune", il est possible d'écrire :

$$F_{\text{ppcover}} = \text{tout\_intersection}(e_{IP_{S+}}, \text{ch}_{BP_C}).$$

Il est éventuellement possible d'introduire une information similaire à celle créée par l'introduction de coefficients (fréquence absolue d'un symptôme ou d'une maladie, fréquence d'association d'un symptôme avec une maladie). Supposons que ces coefficients varient de 0 à 1 avec une précision de 0,1. Il est alors possible de discrétiser la représentation en créant les applications  $SP_0, SP_{0.1}, SP_{0.2}, \dots, SP_1$ , pour ce qui est des fréquences absolues, et les applications  $BP_0, BP_{0.1}, \dots, BP_1$ , pour ce qui est des fréquences d'associations. Néanmoins cette façon de faire est éminemment artificielle (on retrouve ici la notion d'équivalence formelle qui s'oppose à celle de bonne, ou mauvaise, représentation).

### 4.5.3. Calcul des prédicats (CP1)

Nous allons commenter ci-dessous l'équivalence qui existe entre notre formalisme ensembliste et le calcul des prédicats du premier ordre.

Nous allons montrer, sur un exemple, comment évolue l'écriture d'un problème concret, lorsque l'on passe d'un système à l'autre et nous verrons qu'il est possible d'en tirer quelques enseignements.

Le problème est le suivant : soit un ensemble de symptômes, un ensemble de maladies, un sous-ensemble de maladies dites "graves" et une relation causale unique  $C$  entre symptômes et maladies. On désire obtenir, à partir d'un ensemble  $S+$  de symptômes constatés, l'ensemble  $M+$  défini comme suit :

$M+$  contient les causes communes (selon  $C$ ) des symptômes de  $S+$  ;  
si ces dernières sont trop nombreuses (20 ou plus) on se limitera  
aux seules maladies graves, s'il en existe, sinon on conservera  
l'ensemble complet.

En formulation ensembliste cela se transcrit par : (1)

```
tmp1 = tout_intersection(S+, C)
Si [ Card(tmp1) ≥ 20 et Card(intersection(tmp1, M_graves)) ≠ 0 ]
  Alors M+ = intersection(tmp1, M_graves)
Sinon M+ = tmp1
```

Pour faciliter l'expression du même mécanisme en CP1 nous allons définir une notation préliminaire dénommée "il existe au moins  $n$  éléments qui .." et notée  $\exists_{+n}$ .

On notera l'équivalence :

$$\exists_{+n} x / P(x) \iff \exists x_1 / P(x), \dots, \exists x_n / P(x) \text{ et } x_1 \neq x_2, \dots, x_1 \neq x_n, x_2 \neq x_3, \dots$$

Nous allons également définir quatre prédicats :

$$\begin{aligned}
P_1(x) &= \forall s \in S+ , C(x,s) \\
P_2(x) &= x \in M_{\text{graves}} \\
P_3 &= \exists_{+20} m' \in M / \forall s \in S+ , C(m',s) \\
P_4 &= \exists m'' \in M_{\text{graves}} / \forall s \in S+ , C(m'',s)
\end{aligned}$$

Le premier prédicat indique si son paramètre est une cause commune de S+. Le second indique si son paramètre est une maladie grave. Les troisième et dernier prédicats sont un peu particuliers, en ce sens qu'ils ne font pas intervenir de paramètre. P<sub>3</sub> est vrai s'il existe plus de vingt maladies qui sont des causes communes de S+. P<sub>4</sub> est vrai si, parmi les maladies graves, il existe au moins une cause commune de S+.

La définition de M+ peut alors s'écrire : (2)

$$M+ = \{ x / P_1(x) \wedge P_2(x) \wedge P_3 \wedge P_4 \vee P_1(x) \wedge \neg P_3 \vee P_1(x) \wedge P_3 \wedge \neg P_4 \}$$

Les deux formulations (1) et (2) sont totalement équivalentes d'un point de vue formel. Considérons maintenant que l'on veuille utiliser ces formulations pour un calcul effectif. La forme (2) peut devenir intéressante à deux conditions :

- on s'intéresse à un test d'hypothèse du type "est-ce que m<sub>i</sub> appartient à M+ ?"
- P<sub>3</sub>, P<sub>4</sub> peuvent être évalués sans calculer explicitement toutes les causes communes.

La première condition ne relève que d'une intention de l'utilisateur. Néanmoins, n'oublions pas que, pour la prescription d'examens complémentaires, nous avons besoin de toutes les maladies possibles. La deuxième condition peut être réalisée dans certains contextes. Par exemple, si l'on ne dispose que de quelques symptômes très communs, on peut savoir a priori que ces symptômes ont toujours plus de vingt causes communes. La connaissance d'une classe S<sub>communs</sub> de ce type permettrait de supprimer la détermination des causes communes dans l'évaluation de P<sub>3</sub>.

Nous pensons néanmoins que l'usage de quelques techniques purement informatiques permettrait d'assurer définitivement la supériorité opérationnelle de l'approche ensembliste. Même si les ensembles en présence sont très grands, les calculs d'intersection et de dénombrement peuvent se programmer efficacement. D'autre part, une analyse préalable des méthodes autorise une optimisation de l'écriture. Par exemple, dans l'exemple ci-dessus, l'expression "tmp1=tout\_intersection(S+, C)" peut être remplacée par une fonction plus complexe qui compterait, au fur et à mesure de leur apparition, les maladies graves et qui arrêterait le calcul s'il s'en présente plus qu'un nombre N passé en paramètre. La recherche

de fonctions de base adaptées, du point de vue des temps de réponse, pourrait se placer dans une phase terminale d'optimisation.

#### *4.5.4. Coefficient numériques ou ensembles symboliques ?*

D'un certain point de vue, on peut ressentir cette incapacité à manipuler des coefficients comme un avantage. Historiquement parlant, l'usage de coefficients numériques avait pour but de reproduire certaines préférences, marquées par les experts d'un domaine, sans pour autant se soumettre aux contraintes fortes associées à une approche purement probabiliste.

Nous soutenons, à titre d'hypothèse de travail, qu'il est souhaitable d'exprimer et de manipuler ces préférences conformément aux habitudes de langage des experts d'un domaine. Ainsi, les médecins classent plus volontiers les maladies en quelques groupes comme "les maladies très rares", "les maladies peu courantes", "les maladies fréquentes", que sur une échelle numérique continue.

Ces groupes ne constituent pas une discrétisation d'une échelle de fréquence objective. Par exemple, à fréquence objective égale, une maladie bénigne pourra être perçue (fréquence subjective) comme "plus rare" du fait que son omission temporaire aura moins de conséquences sur la qualité d'un diagnostic. Donc, il existe en réalité plusieurs critères (fréquence, gravité, coût, etc) qui sont intimement liés dans un contexte opératoire. Or, pour la conception d'une base de données, c'est à l'expert que nous allons demander la "fréquence" d'une maladie. Il importe alors d'accorder à cette fréquence subjective la même signification que celle accordée par l'expert, par exemple en considérant que, dans certains contextes, les maladies fréquentes sont à placer en priorité. Dans ce cas, la notion même de fréquence est déconnectée de toute considération relative au nombre d'apparition de telle ou telle maladie ; on ne s'intéresse pas à la manière dont l'expert a construit se classement. Par une sorte de définition à rebours, on peut alors considérer que la fréquence est une propriété d'une maladie et que cette propriété lui confère une certaine priorité dans un processus donné (le diagnostic). Nous ne pouvons simultanément obtenir une "fréquence" auprès d'un expert et nous servir de cette fréquence comme si elle représentait effectivement la fréquence objective d'apparition d'une maladie.

La fréquence subjective est inscrite dans un contexte opératoire spécifique d'un domaine alors que la fréquence objective est indépendante dudit domaine. En médecine, on retarde l'examen des hypothèses relatives aux maladies dites "rares" au profit des maladies dites "fréquentes". Il existe un recouvrement important de ces deux notions de fréquence objective et subjective, mais non une identité. Nous conserverons, pour notre part, les différentes classes relatives à une même grandeur (ici la fréquence) comme des entités distinctes. Par exemple, nous manipulerons indépendamment des classes comme "maladies très fréquentes", "maladies fréquentes" et "maladies rares". Cela nous préservera de la tendance à accorder une

valeur sémantique uniforme et graduée à ces différentes classes. Par exemple, les maladies fréquentes sont utilisées dans une heuristique de réduction des hypothèses, en première analyse, alors que les maladies rares peuvent prendre un caractère prioritaire dès lors que l'investigation clinique ne donne que des signes erratiques. Les maladies rares ne sont pas une variante à un coefficient près des maladies fréquentes, elles relèvent d'un contexte totalement différent.

En conclusion : soit on travaille "mathématiquement" à partir de données objectives, auquel cas il n'est nul besoin de disposer d'expert ; soit on travaille à partir de données subjectives<sup>1</sup> fournies par les experts et l'on doit alors respecter les méthodes de manipulation qu'ils utilisent relativement à ces données.

En ce qui concerne le diagnostic médical, nous avons donc le choix entre une approche statistique pure, excluant toute intervention des médecins, et une approche IA où les experts sont à la fois sources d'informations, références et clients. Des problèmes apparaissent dès lors que l'on veut "projeter" des données d'un domaine sur l'autre. Les médecins utilisent mal et interprètent difficilement des données statistiques [Fieschi 84], de même les données manipulées par les médecins sont quasiment impossibles à intégrer dans un système statistique.

Il existe certainement un gros travail à faire pour : a) objectiver les paramètres utilisés par les experts, b) relier ces paramètres à un contexte opératoire lui-même difficilement caractérisable. On pourrait poser le problème ainsi : étant donné une maladie dont le corps médical reconnaît qu'elle est "fréquente", quelle est la combinaison de paramètres objectifs mesurables qui caractérise cette "fréquence" ? Si les médecins ont construit cette combinaison, c'est probablement pour des raisons d'efficacité à l'intérieur d'un raisonnement (un processus productif) ; quel est ce raisonnement ?

Dans l'attente de résultats relatifs à ce domaine, nous avons préféré développer une méthode originale pour utiliser, telle quelle, la représentation des experts et pour reproduire leurs méthodes de manipulation. De nombreux travaux fondés sur l'usage de données statistiques objectives se développent également (voir, par exemple, [Robert 88, Rialle 87]). A terme, une hybridation devrait être possible, les observations objectives et volumineuses étant comparées à des bases de données statistiques, les observations peu quantifiables<sup>2</sup> étant manipulées par des "systèmes experts" (systèmes visant à reproduire le comportement d'un expert).

---

<sup>1</sup> une précision est nécessaire : une donnée peut être considérée comme "objective" dès lors qu'elle s'intègre dans une théorie forte, un modèle. Si nous disposions d'une théorie complète relative au comportement des experts, les données "subjectives" qu'ils expriment pourraient devenir des données "objectives" de la théorie. La notion de subjectivité est donc ici une facilité de langage qui indique essentiellement l'absence de théorie à laquelle rattacher ces données. Notre approche peut également être perçue comme l'établissement d'une "théorie faible".

<sup>2</sup> deux éléments interviennent pour déterminer la nature d'une information : son mode d'observation (mesure objective ou avis d'un tiers), la nature des connaissances permettant de la traiter (mesures statistiques ou savoir-faire).

Exemple d'observation objective et volumineuse : le dépouillement d'un électrocardiogramme de longue durée (certaines d'heures d'enregistrement).  
Exemple de donnée peu quantifiable : la pénibilité d'un examen pour un malade donné en fonction de son âge, de son sexe, du gain de confort qu'entraînerait une meilleure adaptation de la thérapeutique, etc.

#### 4.6. Complétude ; définition et évaluation

Avant toute évaluation de la complétude, il est nécessaire de la définir. Nous donnerons, dans le cadre de cette thèse, une définition volontairement axée sur le diagnostic. Tout lecteur intéressé pourra étendre cette définition au cadre général de la propagation de propriétés.

Etant donné un ensemble  $S$  de symptômes et un ensemble  $M$  de maladies, nous considérerons défini un diagnostic sur  $S/M$  si, à tout sous-ensemble  $S+$  de  $S$ , nous pouvons faire correspondre une *explication*  $M+$ . Autrement dit, nous connaissons (par exemple en extension) une fonction  $F$ , des parties de  $S$  dans parties de  $M$ , qui représente le "bon diagnostic".

Soit maintenant un système  $S_{BD}$  comprenant une propriété  $PP_{M+}$  définie par une formule bien formée. Cette formule s'appuie sur une unique propriété initiale  $IP_{S+}$ . Un tel système  $S_{BD}$  sera dit "complet" si la proposition suivante est vérifiée :

$$\forall S+, e_{IP_{S+}} = S+ \Rightarrow e_{PP_{M+}} = F(S+)$$

Autrement dit, le système de diagnostic est une implantation de la fonction  $F$ . Moyennant quelques libertés d'écriture, nous pouvons exprimer cela comme  $F_{PP_{M+}} = F$ .

Le système présenté dans le chapitre 4 est complet. En effet, il suffit de définir autant de propriétés unaires  $SP_i$  que d'instances possibles de  $S+$ , autant de propriétés unaires  $SP_i'$  que d'instances possibles de  $F(S+)$ , puis d'écrire  $F_{PP_{M+}}$  comme une fonction d'appariement du type :

$$\begin{aligned} & \text{Si } (e_{IP_{S+}} = e_{SP_1}) \text{ Alors } e_{SP_1'} \\ & \text{sinon} \\ & \text{Si } (e_{IP_{S+}} = e_{SP_2}) \text{ Alors } e_{SP_2'} \\ & \dots \end{aligned}$$

$$\text{Avec } \forall i, e_{SP_i'} = F(e_{SP_i}).$$

Tout ceci revient simplement à dire qu'il nous est possible, au pire, de représenter la fonction  $F$  en extension.

Cette solution est en fait assez contraire à la philosophie sous-tendant la représentation proposée. Les formules  $F_{pp_i}$  étaient censées décrire en compréhension les dépendances de propriétés. Or, nous venons de considérer une expression en extension comme un cas particulier d'expression en compréhension. Si cela est formellement faisable, ce n'est guère souhaitable, surtout dès que la fonction  $F$  décrit un "diagnostic réel". En effet, tout système réel sur lequel on peut pratiquer un diagnostic possède un comportement qui ne se réduit pas uniquement à un agrégat de cas particuliers. Une interprétation est donc à la fois possible et nécessaire. Dans le cas contraire, il faut mémoriser l'ensemble des configurations et l'on ne peut réellement parler "d'interprétation" pour qualifier une simple recherche en mémoire.

L'expression présentée ci-dessus n'exploite aucune propriété binaire  $BP_i$ . Le modèle de J. Reggia exploite une unique relation  $C$ . Est-il donc possible d'écrire une formule  $F_{PPM+}$  exploitant une unique propriété  $BP$  (équivalente à la relation  $C$ ) en vue de reproduire une fonction  $F$  donnée ? Et ceci en évitant d'établir une représentation en extension (agrégat de cas particuliers).

En d'autres termes : est-il possible de construire un système complet à partir d'une unique propriété binaire et d'une méthode générale ? Si la réponse était oui, cela signifierait que la connaissance des maladies évoquées par chaque symptôme, pris isolément, serait suffisante pour reconstruire la notion "d'évocation par un groupe de symptômes", quelle que soit cette dernière.

Dans le cas général la réponse est non. Il existe des fonctions  $F$  non reproductibles par un doublet  $(BP, F_{PPM+})$ . Cette démonstration est présentée en annexe B.

Elle est malheureusement assez peu constructive, elle ne nous permet pas de déterminer combien de relations sont nécessaires et suffisantes ni comment les construire.

D'autre part, dans le cadre d'une application réelle, nous ne disposerons jamais de la description complète de la fonction  $F$ . Outre la combinatoire que cela impliquerait, il existe des combinaisons de symptômes qui ne se produiront jamais. De plus, à supposer que, pour un  $F$  donné, une ou plusieurs représentations existent, il n'est pas évident que ces représentations soient pertinentes relativement à une interprétation humaine de  $BP$  et  $F_{PPM+}$ . Rappelons en effet que  $BP$  est censé représenter, peu ou prou, un pouvoir d'évocation d'un symptôme et que  $F_{PPM+}$  doit décrire une méthode de combinaison de ces évocations évoluant autour de la notion de "cause commune".

En conclusion, toute solution produite automatiquement a fort peu de chances d'être conforme à la philosophie globale de l'approche IA. C'est pourquoi nous avons considéré que le nombre d'applications  $BP_i$  nécessaires à la résolution d'un problème devait être laissé à l'appréciation du concepteur. En pratique, l'étude du langage utilisé par les experts est un guide utilisable en première analyse.

## 4.7. Méthodologie de réalisation d'un système

Etant donné un problème de diagnostic, nous pouvons décomposer la réalisation d'une application, conformément au formalisme proposé, selon les étapes énoncées ci-dessous. Dans le processus de réalisation complète d'un système diagnostic (d'un programme), ceci fait suite à l'analyse en cinq points que nous avons défini au chapitre 2. Il s'agit ici d'ordonner les choix relatifs aux différents objets du formalisme (objets, propriétés, etc). Cet ordre découle de la nature des objets et correspond à une difficulté de réalisation croissante. Nous avons défini les six phases suivantes :

- 1 : identification des objets du domaines (liste de noms) ;
- 2 : à partir du langage relatif au domaine, extraction des termes décrivant des propriétés unaires et binaires s'appliquant aux objets (liste de noms) ;
- 3 : classement des propriétés unaires en propriétés intrinsèques aux objets (gravité d'une maladie, coût d'un examen), en propriétés initiales (présence d'un symptôme) et en propriétés propagées (possibilité d'une maladie) ;
- 4 : explicitation, en langage courant, de la signification des propriétés binaires relativement aux propriétés unaires (exemple : si un symptôme appartenant au groupe "symptômes constatés" est lié par une propriété binaire "affirme" à une maladie donnée, cette maladie doit obligatoirement apparaître dans le groupe "maladies possibles") ;
- 5 : écriture d'un jeu de "bonnes" formules  $F_{PP_i}$  ;
- 6 : test et mise au point progressive.

Lors des phases 1, 2 et 3, le problème initial devra être nécessairement simplifié. Toutes les subtilités du domaine ne seront sans doute pas représentables dans le formalisme imposé (propriétés ternaires, relations temporelles, etc).

La phase critique est la phase 5. Tout d'abord, parce qu'il est difficile de prévoir toutes les interactions possibles entre les propriétés définies. Par exemple, il est évident que, parmi les maladies possibles, les maladies graves ont le pas sur les autres. De même, les maladies fréquentes sont prioritaires sur les maladies rares. Mais qu'en est-il des maladies graves et rares ? Ces différentes configurations devront être évaluées par les experts. D'autre part, il sera également difficile de transformer les relations unissant les valeurs des propriétés en des méthodes de calcul ensembliste qui garantissent que ces relations seront vérifiées.

En termes de complexité, on peut apparenter l'écriture d'un jeu de formules à l'écriture d'un réseau de règles dans un système à règles de production. Les

formules ne peuvent produire des contradictions, mais des interactions imprévues peuvent engendrer des résultats inacceptables. En ce sens, les formules peuvent être fausses mais pas incohérentes.

C'est essentiellement sur le problème de la rédaction, voire de l'acquisition, de ces formules qu'un important travail reste à faire afin de rendre ce formalisme plus facilement utilisable. Nous avons essayé diverses approches, telles que l'acquisition semi-automatique de méthodes à partir d'exemples [Raoult 87] ou la définition d'un "langage de programmation" compréhensible et manipulable directement par les experts. Nous n'avons obtenu que des résultats trop partiels pour être présentés ici. Seule la réalisation effective d'applications complètes, en collaboration avec des experts du domaine, pourra nous permettre d'avancer sur ce point.

Le lecteur intéressé par les problèmes de mise au point pourra également se référer à la seconde partie de cette thèse. Nous y examinerons, d'un point de vue "orienté problème", les principales caractéristiques pratiques de notre formalisme. Le chapitre suivant décrit quelques particularités de l'implantation que nous avons réalisée.

#### 4.8. Conclusion

Ce formalisme peut être vu comme un parallèle entièrement symbolique (non-numérique) de l'approche de J. Reggia. Il peut également être perçu comme une généralisation du mécanisme d'inférence d'un système de règles de production (ordre 0) ; au lieu de propager uniquement la véracité des faits (la propriété d'être vrai), nous propageons autant de propriétés que nécessaires. En tout état de cause, il est formellement équivalent à l'usage du CP1 sur des ensembles finis (cf. paragraphe 4.5.3) et il s'apparente (donc ?) fortement avec les approches "schémas + règles" des approches orientées objets.

Nous exploitons la différence qui existe entre un *objet*, la *connaissance d'un objet* (qui se réduit, dans notre approche, à l'énumération des propriétés que vérifie cet objet) et l'*usage de cette connaissance* dans un processus productif (la méthode qui conduit au diagnostic). Cette différence, intuitivement significative, permet un mode d'expression "naturel" de certains problèmes, élude les problèmes de cohérence (cf. 2<sup>o</sup> partie, paragraphe 2.4) et s'implante à un coût informatique négligeable.

#### 4.9. Récapitulatif

Le tableau ci-dessous résume succinctement les principales composantes conceptuelles de quatre approches. Ce tableau n'est nullement exhaustif

relativement à la variété des approches possibles ; il n'a que valeur d'exemple et vise surtout à montrer que les éléments d'analyse que nous avons développés sont pertinents, même pour des systèmes a priori fort dissemblables.

	<b>De Kleer</b>	<b>R. Reiter</b>	<b>J. Reggia</b>	<b>O. Raoul</b>
<b>Spécification de bon fonctionnement</b>	liste de contraintes	prédicats d'anormalités	liste de symptômes	liste de symptômes
<b>Capacités perceptives</b>	paramètres inter-contraintes	variables du système	symptômes constatés	propriétés initiales
<b>Connaissances a priori</b>	réseau de contraintes	système d'équations	liste de maladies + relation causale	ensemble d'objets + propriétés
<b>Qualification</b>	combinaison de contraintes violées	combinaison de prédicats vrais	combinaison de maladies	propriétés propagées
<b>Mécanisme d'interprétation</b>	ATMS	démonstrateur CP1 + HS-Trees	probabilités + critères	propagation de propriétés

## **Deuxième Partie**

**Etude d'un système d'aide à la  
prescription et à l'interprétation  
d'examens médicaux**



Cette seconde partie s'attache à préciser les particularismes du diagnostic dans le contexte médical, et à dégager quelques points cruciaux relatifs à la réalisation d'applications (choix du niveau d'abstraction des langages, identification de différentes natures de connaissances, etc). Ensuite, nous décrivons comment le formalisme de propagation de propriété peut être dérivé en une structure "données + programme". Nous expliciterons les phases successives de réalisation d'une application concrète, en insistant sur la modularité de notre approche. Enfin, nous présenterons une application complète : l'aide à la prescription d'examens dans le cas des fièvres au long cours.

Nous désignerons par GBDI (Gestion de Base de Données Intelligente) le programme générique, le "générateur d'application", et par GBDI-IPLSYS l'application finale engendrée.

---

# 1. Considérations relatives aux S.E. médicaux

Nous avons jusqu'à maintenant soigneusement évité de manipuler des mots comme "intelligence", "connaissance", "raisonnement". Pour être plus exact, l'usage de ces mots est tout à fait licite dès lors qu'ils se réfèrent à un être humain et non à un programme, si judicieux soit-il.

La présente thèse étant relative à l'Intelligence Artificielle, nous avons néanmoins jugé utile de proposer une interprétation cognitive des concepts dégagés précédemment. La plupart des idées présentées ci-dessous feront fréquemment appel au sens commun et ne constituent aucunement une trame rigoureuse. Le but de cette présentation est de souligner certains points, présents dans le formalisme, qui peuvent être difficile à "sentir".

Nous allons voir deux exemples de diagnostic, l'un relatif aux pannes automobiles, l'autre à la médecine. Ils permettent une présentation intuitive des notions de processus de diagnostic, connaissances factuelles et connaissances expertes. Il sera alors possible de ré-analyser GBDI selon ce vocabulaire, certes moins rigoureux qu'une description formelle, mais plus explicite pour le non-spécialiste.

## 1.1. Le diagnostic, un processus

Prenons un exemple que nous connaissons tous : une panne de voiture. Imaginons qu'après être monté dans votre voiture vous mettiez distraitement le contact et lanciez le démarreur. Rien ne se passe ! Vous reconnaîtrez immédiatement que cela est anormal et conclurez donc à la présence d'une panne, probablement une panne de batterie. Vous tenterez alors d'allumer les phares. Toujours rien ! La panne de batterie semble se confirmer. Par acquis de conscience, vous appuyez sur le klaxon ... qui fonctionne parfaitement. Le problème est plus délicat que prévu.

Laissons là notre automobiliste malchanceux et reprenons le processus qui a produit l'ensemble de ce comportement. Il s'agit bien d'un processus répétitif faisant intervenir une *interprétation* (construction d'hypothèses) et l'*observation* (prise d'information sur l'état du système).

Pourquoi la panne de batterie semble-t-elle si probable (bien qu'aucune probabilité numérique n'ait été évaluée) ? Bien d'autres hypothèses sont en effet possibles : mauvais contact au niveau de la clé, relais du démarreur grillé, démarreur déconnecté ou grillé, etc. On peut voir là l'usage de l'heuristique :

*s'intéresser d'abord aux pannes les plus courantes !* Nous considérerons que les pannes courantes sont connues a priori.

Admettons donc la panne de batterie comme hypothèse première et essayons de la confirmer. Notre automobiliste aurait pu ouvrir le capot et tester la batterie à l'aide d'un voltmètre. Cette opération aurait été plus fiable qu'une mesure indirecte comme l'essai des phares. Là encore, on peut voir une règle de comportement : *effectuer d'abord les actions les plus simples, même si leurs résultats sont peu fiables !* Nous considérerons que les actions simples sont exhaustivement connues.

Une nouvelle recherche d'hypothèse confirme la panne de batterie, c'est une cause commune aux deux symptômes constatés ! Est-ce suffisant ? On pourrait imaginer que le démarreur est grillé et que, simultanément, la manette des phares présente un mauvais contact, ou que toutes les ampoules sont mortes. Mais ces hypothèses ne sont pas également plausibles (au sens de [Peng 86]) et la règle pourrait être : *préférer, si possible, une panne unique expliquant tous les symptômes à une combinaison de pannes !*

Si le klaxon (mesure toujours aussi simple à effectuer) est en état de marche, il est probable que la règle précédente ne pourra être conservée ... à moins que l'on sache que démarreur et phares se partagent un connecteur commun, alors que le klaxon dispose d'un circuit direct, auquel cas nos soupçons se porteront sur ledit connecteur.

Ne détaillons pas davantage. Les règles de comportement que nous avons énoncées peuvent paraître triviales, mais appellent deux remarques : un système informatique n'a aucune raison de les appliquer s'il n'a pas été programmé en ce sens ; ce ne sont pas des règles qui nous garantissent que nous allons trouver immédiatement la bonne solution (la véritable panne) mais que, si nous la trouvons, alors nous aurons fait le minimum d'efforts. En ce sens on peut dire que ces règles décrivent un *comportement* face à un ensemble d'informations incertaines ou imprécises, étant donné un ensemble d'actions possibles.

## 1.2. Différentes natures de connaissances

Prenons deux propositions très communes dans les articles médicaux :

- la tuberculose est une maladie courante ;
- il faut s'occuper en priorité des maladies graves et curables.

On perçoit intuitivement qu'elles ne sont pas de même nature. L'une décrit un objet (une maladie) et le situe dans un groupe, qui n'est autre qu'un objet plus abstrait (ou la représentation ensembliste d'une propriété). L'autre décrit une méthode de travail qui permet de faire les bonnes hypothèses, celles qui garantissent

---

une résolution efficace du problème. Nous dirons que la première est une *connaissance factuelle*, la seconde une *méthode* ou une *connaissance experte*.

Pour nous convaincre de la nécessité de ces deux niveaux, supposons qu'un bibliothécaire particulièrement compétent connaisse un très grand nombre de maladies, ainsi que tous leurs symptômes associés. Sera-t-il pour autant un bon médecin ? En pratique non ! Etant donnée une liste de symptômes, il ne pourra que nous fournir la liste, a priori fort longue, des maladies ayant un rapport avec un ou plusieurs symptômes.

Un moyen fort simple, susceptible d'améliorer les performances de notre bibliothécaire, consisterait à lui enseigner une première méthodologie : la recherche de cause commune (expliquer tout ou partie des symptômes d'après le minimum de maladies).

Encore une fois cela n'est guère suffisant si les symptômes connus sont peu nombreux, par exemple en début du processus de diagnostic, ou s'ils sont contradictoires entre eux, ce qui est fort courant en médecine.

Enseignons alors à ce dernier des heuristiques permettant de réduire le nombre des hypothèses ainsi que des méthodes de gestions d'informations contradictoires : que faut-il négliger ? quels sont les critères permettant d'identifier les informations pertinentes ? On voit bien que l'on se rapproche progressivement des capacités d'un médecin. Pourtant les connaissances livresques du bibliothécaire n'ont pas été augmentées. Il ne connaît pas davantage d'objets mais il les manipule fort différemment.

Ce phénomène peut s'analyser de la manière suivante : dans un premier temps, même si notre bibliothécaire connaît la proposition "la tuberculose est une maladie courante", il ne connaît pas pour autant le contexte opératoire associé à cette notion de fréquence subjective. Ce contexte opératoire va apparaître, par exemple, parmi les heuristiques de réduction du nombre d'hypothèses (s'il existe plus de 30 maladies possibles, se limiter aux seules maladies courantes). On peut donc attribuer deux adjectifs distincts à une propriété comme "maladie courante" : cette propriété existe de part sa définition en extension ; cette propriété signifie de part l'usage qui en est fait.

Il va de soi que, dans un système de propagation de propriétés, toutes les propriétés doivent "signifier" quelque chose soit pour le système lui-même, soit pour l'opérateur qui active ce système. Une propriété (c.a.d. l'ensemble qui la représente) qui ne serait jamais ni utilisée dans une formule  $F_{pp_i}$ , ni examinée par l'opérateur n'a pas lieu d'être.

De tout cela, on peut conclure que la séparation de la connaissance en connaissances factuelles et connaissances expertes n'est pas arbitraire. Il s'agit bien d'informations, complémentaires certes, mais de types différents. Les connaissances factuelles attestent l'existence de propriétés ; les connaissances expertes définissent la signification des propriétés existantes.

### 1.3. Connaissances, méta-connaissances, contrôle

Nous avons utilisé ci-dessus le terme de "connaissances expertes" de préférence à celui de "méta-connaissance". Il existe principalement deux raisons à cela. Tout d'abord, "méta-connaissance" est censé décrire une notion très générale, or nous avons ici besoin d'un nom décrivant une quantité précise : les  $F_{ppi}$  associés à la description des propriétés  $PP_i$ . D'autre part, les multiples définitions de la méta-connaissance se réfèrent plutôt à la "connaissance sur la connaissance". Dans notre cas, les connaissances factuelles et les connaissances expertes sont relatives à deux domaines distincts, respectivement la description de propriétés d'objets et la description de méthodes.

D'autre part, la notion de contrôle est absente de notre formalisme ; la structure même de la grammaire ne permet de définir que des fonctions effectivement calculables. Par contre, si l'on s'autorise à faire apparaître, dans une formule  $F_{ppi}$ , un autre terme  $PP_j$ , alors on définit implicitement un ordre sur les évaluations des formules. L'ordonnement des évaluations pourrait être laissé à la charge d'un mécanisme simple que nous pourrions assimiler à un "contrôle". Ce mécanisme pourrait également se charger d'effectuer des évaluations paresseuses (*lazy-evaluation*) en partant d'une requête spécifique de l'utilisateur ("valeur d'une propriété  $PP_i$  ?"). En pratique, ces différentes sophistications techniques nous ont parues inutiles. Le diagnostic, exprimé en terme de propagation de propriétés, ne comporte en fait que peu de propriétés terminales et il est nettement plus simple de les calculer toutes, éventuellement selon un ordre pré-établi. Remarquons au passage que cet ordre est relativement évident une fois spécifié le domaine. Dans l'application "fièvres au long cours", il est nécessaire de connaître toutes les maladies dites "possibles" (les hypothèses de pannes) avant de pouvoir déterminer les examens dits "pertinents" (qui seront, par exemple, discriminants par rapport aux hypothèses). Tant que les formules sont peu nombreuses, un contrôle n'est pas nécessaire. Un tel contrôle serait, de plus, d'une nature essentiellement informatique et indépendant du domaine ; il peut faire l'objet d'optimisations ultérieures.

Les termes usuels (connaissances, méta-connaissances et contrôle) sont impropres à décrire notre approche. Selon les goûts de chacun, ils peuvent s'appliquer à divers éléments et, en tout état de cause, ils ne sont réellement significatifs pour aucun. C'est pourquoi, même dans le cadre de cette présentation intuitive, nous ne les utilisons pas, leur préférant des termes qui sont moins connotés.

### 1.4. Diagnostic et raisonnement incertain

Nous pouvons émettre quelques remarques concernant le caractère "incertain" du raisonnement diagnostic. En effet, il pourrait paraître étrange que nous n'ayons pas

---

abordé le sujet jusqu'ici. Dans tous ce qui précède, nous avons caractérisé un processus diagnostique comme une boucle d'acquisition d'informations en vue de produire une description d'un objet observé, selon un langage donné : le langage de qualification.

Ce processus même peut être vu comme un méthode "d'amélioration de la qualité de l'information" (ou extension des capacités perceptives). J.M. Lefevre [Lefevre 89] effectue une thèse sur ce sujet et nous présentons ci-dessous les conclusions qui nous concernent en priorité.

Un système informatique peut être considéré comme "gérant l'incertain" dès lors qu'il utilise des informations qu'il s'autorise à remettre en cause. Un tel système présente un comportement global antropomorphique ; il peut prendre des décisions sur des bases faibles, acceptant un certain risque d'erreur, et, si le risque est inacceptable, il devra tenter de confirmer ou d'infirmer ces données initiales en faisant appel à de nouvelles observations (perception active).

Dans un tel système, il n'est pas nécessaire que l'incertain dispose d'une représentation explicite et localisée comme, par exemple, des coefficients numériques.

Nous considérons que la gestion de l'incertain est un propriété induite d'un système de diagnostic. Le diagnostic est un cas particulier de raisonnement sur la qualité de l'information (on pourrait également voir le raisonnement incertain comme une forme de diagnostic sur l'état d'un corpus d'informations).

Voyons maintenant en détail quelques sources d'incertitudes, dans le cadre du diagnostic. Le but d'un processus diagnostique est de décrire un objet observé dans les termes du langage de qualification. Pour cela, le système va interpréter des résultats d'observations à l'aide d'un ensemble de connaissances.

Les diverses sources d'incertitudes apparaissent nettement : les observations peuvent être bruitées et fournir des informations fausses ; les connaissances peuvent être incomplètes ou erronées ; le processus d'utilisation de ces connaissances (le "raisonnement") peut également être basé sur des prémisses peu fiables ; enfin le langage de qualification de panne peut introduire des incertitudes du point de vue de l'utilisateur (terme ambigu, description incomplète).

Nous considérerons que c'est l'existence même des incertitudes relatives aux résultat des observations qui impose un raisonnement de type diagnostique. Les médecins connaissent depuis longtemps les phénomènes de "faux négatif" ou "faux positif" associés, par exemple, aux examens biologiques (un germe présent peut ne pas être détecté, un germe absent peut être détecté suite à une contamination de la culture). De même, il existe une très grande variabilité des manifestations symptomatiques entre les individus atteints d'une même maladie. Les connaissances qui associent, en général, tel symptôme à telle pathologie sont potentiellement erronées dès lors qu'on les applique à un individu précis. Enfin, les protocoles d'investigations tiennent compte, entre autres, de considérations économiques qui, si elles garantissent une efficacité à long terme, peuvent conduire à une conclusion

correcte d'un point de vue méthodologique mais fausse eut égard à un malade donné. Par exemple, le respect d'un protocole peut conduire à pratiquer, malgré tout, certains examens dont on n'attend aucun résultat pratique (cas des potentiels évoqués auditifs pour le diagnostic de la sclérose en plaques).

Les incertitudes sur les observations, sur les connaissances et sur le raisonnement sont gérées principalement de deux manières dualles : la redondance et l'oubli. Si un examen peut éventuellement fournir un résultat faux, il est peu probable que N examens, basés sur des processus d'observations indépendants, s'avèrent simultanément erronés. De même, l'usage d'un ensemble assez vaste de connaissances distinctes diminue l'incidence de connaissances inexactes. Dans tous les cas, on cherchera donc à établir un faisceau d'arguments (cf mécanisme de confirmation [Lefevre 88]).

Cette notion implique également que l'on ne s'impose pas d'expliquer tous les symptômes constatés. On est en droit de considérer qu'un symptôme inexplicable est probablement un faux positif (ou faux négatif selon les cas), ou que les connaissances qui nous ont amenées à exhiber telle pathologie ne s'appliquent pas pour le malade considéré.

Ces deux mécanismes ont leurs propres limites. La recherche d'une forte redondance nécessite une base de connaissance pléthorique et risque d'entraîner un excès d'investigations parfois coûteuses. Des critères économiques définissent la redondance admissible. A l'inverse, s'il est possible de considérer qu'un symptôme est inexplicable, il est néanmoins impératif d'en expliquer une forte proportion. La description, en termes de pathologie, d'un malade donné doit tenir compte du fait que les connaissances sont "généralement" vraies et que les examens sont "généralement" fiables<sup>1</sup>.

L'usage de la redondance a pour conséquence une multiplication des observations suivie d'une recherche des causes communes. La faculté d'oubli est intégrée dans l'usage d'heuristiques (par exemple, considérer en priorité les symptômes fiables) ou de méthodes de gestion des contradictions (la plupart de ces méthodes reviennent à rejeter, plus ou moins explicitement, certaines informations). Ces deux idées sont à l'origine de la recherche d'une séparation explicite des connaissances factuelles et des connaissances expertes.

Un exemple extrême de gestion de l'incertitude par la redondance concerne la prescription homéopathique [Broussalian 88]. Ce domaine fait bien appel à des bases de données importantes (répertoires de Bœnninghausen, Kent et Barthel) et impose un tri judicieux entre symptômes à retenir et symptômes à négliger. On se place ici dans le contexte spécifique de l'homéopathie : tous les symptômes sont accessibles par interrogation du malade, les critères économiques limitant les

---

<sup>1</sup> Si cela n'est pas vrai pour toutes les connaissances, ou tous les processus d'observations, ce fait doit être explicitement pris en compte lors du raisonnement. Par exemple, certains examens présentant un fort taux de "faux positifs" peuvent être négligés si besoin est, sans pour autant être comptés dans le processus qui limite le nombre d'informations pouvant rester inexplicables.

observations réalisables ne s'appliquent donc pas, d'où la possibilité d'une forte redondance.

On voit dans ce qui précède comment les incertitudes sur les observations et sur les connaissances peuvent être gérées par un mode de raisonnement approprié sans que soit explicitement associé à chaque fait, ou règle, un coefficient numérique. On peut voir cela comme une manière de gérer un ensemble de faits qui serait tous affectés a priori d'un même coefficient de certitude (par exemple 0,5). S'il existe deux classes d'examens, les uns assez fiables, les autres assez peu fiables, il nous paraît plus explicite de conserver ces deux classes en des termes ensemblistes plutôt que d'user de deux coefficients arbitraires (par exemple : 0,4 et 0,8). En d'autres termes, l'usage de coefficients numériques ne nous paraît justifiée que si la pleine précision de l'intervalle 0-1 est effectivement requise, ou si ces coefficients proviennent de mesures objectives.

### 1.5. Diagnostic assisté, principe et philosophie

Le but de GBDI-IPLSYS (application relative aux fièvres au long cours) est de fournir un outil d'aide au diagnostic et non un système de diagnostic automatique. Nous avons vu dans l'introduction que seul un système totalement autonome pourrait prétendre effectuer automatiquement un diagnostic complet. Tout système qui dépend d'un opérateur, ne serait-ce que pour entrer des données simples, se repose sur ce dernier pour limiter le processus d'interprétation.

Prenons comme exemple un système de diagnostic qui requerrait l'information suivante :

- le malade a-t-il de la fièvre ? (1)

L'opérateur chargé de répondre à cette question doit donc évaluer la situation, à partir de paramètres variés, afin de décider si "il est possible de dire que le patient est fébrile". Par exemple, il devra décider si 37°9 est une température anormale, sachant que le patient est un jeune enfant qui a déjà pris une forte dose d'aspirine. Si notre système initial désire effectuer lui-même cette évaluation il devra disposer d'un très grand nombre d'informations (l'âge du patient, son historique clinique) et d'un grand nombre de connaissances (l'influence de l'âge sur la température, l'effet des médicaments, etc). La question initiale pourrait alors être étendue en :

- température du patient (en degrés) ? (2)

- âge ?

- absorption d'aspirine ?

- ...

On pourrait éventuellement se reposer sur l'opérateur à un moindre niveau que dans la question 1. Par exemple :

- température du patient ? (3)
- révision à la baisse (oui-non) ?
- révision à la hausse (oui-non) ?

Dans ce troisième cas, on se contente de sensibiliser l'opérateur à cette idée que la température, exprimée en degré, peut être non-significative. Le système va, quant à lui, exploiter ces trois réponses avec une approche simple de type "logique floue". Il n'y a pas de différence réelle avec le type 1 ; on a simplement ajouté un mécanisme obligeant l'opérateur à se focaliser sur un aspect du problème, sans pour autant rendre nécessaire l'ajout de nouvelles connaissances dans le système.

Le concept qui apparaît ici est celui du *niveau d'abstraction* des données initiales. Nous qualifierons un système de type 1 comme "de plus haut niveau" qu'un système de type 2.

Ceci appelle deux remarques :

- plus le niveau d'abstraction des données est bas (format 2), plus nombreuses seront les connaissances nécessaires pour que le système puisse synthétiser ces données en informations de haut niveau. Comme les méthodes de représentation de connaissances dont nous disposons actuellement sont assez frustrées, il existe certainement un niveau en dessous duquel nous ne pouvons pas descendre. Ce niveau se situe probablement très loin au-dessus du plancher défini par le système totalement autonome.
- le niveau d'abstraction choisi doit correspondre au seuil entre ce qui est évident, pour l'opérateur, et ce qui ne l'est pas. Si un système questionne un médecin compétent sur mille points de détails uniquement pour déduire que 39° correspond à un état fébrile, il aura peut-être prouvé sa propre correction mais n'aura fourni aucune aide. A l'inverse, une question comme "l'état général du malade est-il bon ?" peut être une question sans réponse pour un interne débutant, car c'est justement ce qu'il se demande.

Un système trop précis va donc être informatiquement lourd et produire des conclusions évidentes pour l'opérateur. A l'inverse, un système trop abstrait va reporter tout le travail d'interprétation chez l'opérateur et se contenter de déductions quasi-tautologiques comme, par exemple, "si le malade est fortement fébrile et que son état général est mauvais, alors il est dans un état grave".

Nous avons opté pour une attitude moyenne. **Les données entrées par l'opérateur doivent lui demander le maximum de travail d'interprétation dont il est capable, sans toutefois aller jusqu'aux**

---

**questions qu'il se pose lui-même.** Il importe particulièrement que l'opérateur ait conscience de ce fait : aucune des réponses aux questions qui lui sont posées n'est a priori parfaitement évidente.

Il apparaît assez vite que, pour un système de ce type, le raisonnement diagnostic est fait par l'opérateur. Le système se contente pour sa part de mener un simulacre de ce raisonnement afin de poser les questions pertinentes. On espère que ces questions vont obliger l'opérateur à explorer des possibilités qu'il avait écartées sans réellement y réfléchir. Qu'il soit, par la suite, amené à y trouver la solution ou à les écarter de nouveau n'est pas un problème en soi. Dans les deux cas, il aura augmenté sa compréhension du cas qu'il observe, ce qui est le seul but du système d'aide au diagnostic. On trouve un avis de ce type dans [Ganascia 84] : "Une situation est décrite au système qui aide alors l'opérateur à cerner le problème en lui posant des questions pertinentes, le guidant ainsi vers un diagnostic".

Ce type de conception d'un système d'aide au diagnostic semble donc s'éloigner radicalement des ambitions initiales de l'intelligence artificielle (la machine autonome). Cela dépend en fait du point de vue que l'on adopte. Notre but est ici de réaliser un système qui effectue un simulacre de raisonnement afin d'unir deux niveaux de qualification (symptôme et pathologie). Le choix du symptôme comme donnée d'entrée impose le recours à l'opérateur. La méthode utilisée peut néanmoins être testée à n'importe quel niveau : du signal au symptôme, du symptôme au syndrome, etc. Des résultats de ces tests émergeront progressivement les différents sous-systèmes nécessaires à la réalisation d'une chaîne complète (du signal à la pathologie). Nous ne prétendons pas avoir fait autre chose qu'un pas dans une longue marche, mais il ne faudrait pas prendre prétexte de la petitesse du pas pour renoncer à suivre la direction fixée.

Le système de propagation de propriétés (voir chapitre 4, première partie) est particulièrement bien adapté à la réalisation d'un système "symptomato-pathologique" (c.a.d. effectuant une qualification pathologique d'un ensemble de symptômes). Paradoxalement, du moins en apparence, ce système ne pose aucune question lors de la propagation. Nous avons en effet choisi de considérer que les "informations qu'il serait intéressant de connaître" constituent une production du système, mais qu'il ne lui appartient pas de s'en servir directement. Dans le cas de GBDI-IPLSYS, il existe deux propriétés propagées (Ex-conseillés et Fait-à-demander) qui contiennent respectivement la liste des examens para-cliniques potentiellement intéressants, ainsi qu'une liste de questions possibles. L'opérateur peut alors choisir d'effectuer, ou non, les seuls examens proposés. Pour ce faire il dispose en propre d'informations complémentaires (tel examen ne fournit jamais aucun résultat, contrairement à ce que contient la base de données du système ; tel autre prendra trop de temps, car le technicien est en vacances ; etc). Il va donc modifier les données de sortie pour en extraire une véritable séquence d'actions. A l'extrême, la simple mention d'un examen donné peut amener l'opérateur à réassembler différemment de petits fragments d'informations, à réinterpréter

totale de la situation, ce qui lui permettra alors de poser le diagnostic final sans aucune aide du système.

Conservons en mémoire que l'opérateur dispose toujours de plus d'informations que le système, soit parce que ces informations sont non-représentables dans le formalisme fixé, soit parce qu'elles n'ont pas été introduites lors de la conception, soit enfin parce que l'opérateur a omis de fournir effectivement ces informations au système. A l'inverse le système utilise toutes les informations dont il dispose de manière complète et explicite (il ne peut rien oublier). Il est donc naturel de vouloir établir une synergie médecin-système qui exploite au mieux ces capacités respectives.

Tout ceci peut se résumer en trois principes :

- il faut exploiter les capacités d'interprétation de l'opérateur, afin de limiter la quantité de connaissances à représenter dans le système ;
- il faut exploiter également l'autonomie de l'opérateur en matière d'actions. Comme il existe des informations que ce dernier possède, mais qui ne sont pas représentables dans le système, il sera à même de trier les conseils proposés. On peut donc se limiter à une "sélection préalable", sans viser l'obtention d'une solution applicable ;
- le système doit être considéré par l'opérateur comme un "agitateur intellectuel", une sorte de check-list complexe et réactive. L'opérateur doit se servir du système et non adopter une attitude passive.

L'efficacité d'un tel système d'aide au diagnostic est beaucoup plus difficile à évaluer que le traditionnel "pourcentage de bon diagnostic" d'un système visant l'automatisme complet. GBDI-IPLSYS n'est pas conçu pour produire effectivement un diagnostic quelconque. Seul nous importe la plus-value du couple médecin-système par rapport au médecin seul. Là encore, le seul dénombrement des bons diagnostics est insuffisant. Il faudrait y inclure la réduction de la durée des investigations, l'augmentation du confort pour le malade, ainsi que d'autres critères impalpables comme la tranquillité d'esprit du médecin.

## 1.6. Communication homme-machine

Les systèmes experts médicaux ont souvent été présentés comme des substituts du médecin, ce dernier étant limité au rôle, prétendument passif, d'opérateur de saisie.

Nous avons brièvement montré que tout système actuel se repose sur les capacités d'observations du médecin (utilisation d'un langage d'observation fondé sur des symptômes). Nous aimerions argumenter le fait qu'une collaboration homme-machine, dans le cadre de l'aide au diagnostic, peut prendre d'autres formes

---

que celle de la soumission de l'opérateur à la machine. La plupart des idées présentées ci-dessous proviennent d'adaptation des idées de J. Quinqueton et J. Sallantin relatives aux systèmes d'apprentissage contrôlables par réfutation [Quinqueton 87].

Le principe central est que la communication entre deux agents rationnels peut passer par un processus de contestation mutuelle.

Imaginons un système informatique utilisable de la façon suivante : un expert entre les observations effectuées et y ajoute la conclusion à laquelle il est parvenu ; le système reprend toute ces données et cherche à établir des conclusions alternatives. Dans ce cas, la conclusion initiale de l'expert n'intéresse le système que d'un point de vue négatif : tout raisonnement qui tendrait à confirmer ce résultat est sans intérêt. Admettons que le système réussisse à établir des alternatives, l'expert doit être à même de réfuter ces solutions. Deux cas peuvent se présenter : l'expert dispose d'informations complémentaires que la machine est à même d'intégrer (informations représentables) et ces informations invalident la réfutation fournie ; l'expert dispose d'informations non-représentables et prend sur lui d'affirmer que la réfutation est irrecevable, auquel cas la machine doit établir d'autres réfutations.

Si, à un moment donné de ce processus, l'expert s'avère incapable de réfuter une alternative, il devra admettre que son raisonnement est incomplet. A l'inverse, si la machine s'avère incapable de proposer des alternatives, l'expert pourra considérer que son raisonnement possède une relative solidité.

Un tel système n'est pas inconcevable. Par exemple, un logiciel qui sélectionnerait, au hasard, un nom de maladie autre que celui fournit par l'expert présenterait le comportement décrit ci-dessus.

Bien entendu, un tel système serait parfaitement insupportable du fait de l'absence de pertinence des alternatives. Il est nécessaire que les solutions proposées soit difficiles à réfuter ; le système, malgré sa volonté de donner tort à l'expert, doit respecter certaines règles. Entre autres, ces règles doivent le mettre dans l'incapacité de poursuivre indéfiniment la réfutation.

La propagation de propriétés peut être orientée selon cette optique. Nous pouvons définir, parmi les diverses propriétés initiales, un ensemble du type "conclusion-expert". Il reste alors à définir des méthodes (des formules) permettant de calculer une propriété de type "conclusions-alternatives".

Nous avons bien conscience que l'argument est partiellement fallacieux. Ce formalisme est, par sa nature même, proche d'un langage de programmation. A ce titre, il peut être orienté dans de nombreuses directions selon l'habileté du concepteur des méthodes. En soi, la propagation de propriétés est plus adaptée à un raisonnement "en avant" consistant à modifier la description d'une situation, plutôt qu'à un raisonnement "en arrière" consistant à réfuter une conclusion pré-établie.

---

Néanmoins, selon certains points de vue, la différence entre ces deux fonctionnements est arbitraire. Dès lors qu'un système fournit un ensemble de conclusions et que l'expert possède "une idée sur la question", il lui est tout à fait possible de s'intéresser en priorité aux conclusions qui diffèrent de la sienne propre. Un problème subsiste si l'expert est dans l'incapacité de forcer le système à rejeter une alternative pour en explorer d'autres (ce qui se traite facilement par l'introduction d'une propriété initiale "maladies-écartées"). Quoiqu'il en soit, ces quelques considérations ouvrent des perspectives qui mériteraient d'être explorées plus avant.

---

## 2. GBDI, du formalisme au programme

Nous allons maintenant présenter en détail une implantation complète du système formel défini au chapitre 4 de la première partie. Dans un premier temps, nous justifierons les quelques transformations nécessaires pour associer au formalisme une structure de données simple et homogène. Ensuite, nous examinerons les mécanismes de développement afin d'évaluer la modularité de l'approche. Rappelons que le diagnostic, tel que nous l'avons défini, consiste à émettre des hypothèses de pannes (de maladies) mais surtout à proposer des examens complémentaires conformément à la pratique d'experts du domaine.

GBDI (Gestion de Base de Données Intelligent) est en fait une "boîte vide", un générateur d'application. Tout expert peut, sans aucune aide, créer un ensemble d'objets, les structurer selon des propriétés unaires et binaires et nommer les propriétés propagées. Dans l'état actuel de GBDI, une intervention d'un informaticien reste nécessaire pour l'écriture des formules  $F_{ppi}$ . Ces formules sont incluses, sous forme procédurale, dans un module complémentaire au programme. Une application complète est donc constituée par le noyau GBDI (moteur d'inférence ?), un module complémentaire contenant les méthodes  $F_{ppi}$  (méta-règles ?) et la base de données structurant les objets (règles ?).

Il est plus difficile ici de faire la traditionnelle distinction "générateur de système expert" et "système expert engendré". Pour passer de GBDI à une application complète (en l'occurrence GBDI-IPLSYS), il faut créer une base de données mais également modifier une partie du programme. D'un certain point de vue, il n'y a pas de différence réelle entre ces deux formes de modification d'un logiciel. Dans un cas, on considère le programme comme une sorte d'interpréteur d'un langage ad-hoc, dans le second cas, on se réfère au langage source (ici LISP ou PASCAL).

Nous avons bien tenté de mettre à disposition de l'utilisateur une sorte de pseudo-langage de haut niveau lui permettant de réaliser lui-même les procédures complémentaires. Outre que, même rendue quasi-déclarative, cette phase serait restée trop complexe pour être menée par le seul expert, cela ne s'est pas avéré possible à un coût logiciel raisonnable. La volonté de rester sur micro-ordinateur (Macintosh +) nous interdisait l'utilisation d'environnements de programmation qui auraient rendu cela possible.

### 2.1. Particularités d'implantation

Soit  $O$  la liste des objets connus du système ; ces objets sont "personnalisés" par des propriétés statiques unaires définies en extension ( $SP_i$ ), des propriétés statiques

binaires également définies en extension ( $BP_i$ ). Le caractère individuel de chaque objet réside dans le fait qu'il participe d'une configuration unique (il possède certaines propriétés unaires, il est lié à certains objets par telle ou telle propriété binaire). Chaque objet est susceptible de se voir attribuer des propriétés qui ne lui sont pas intrinsèques, mais qui varient selon les sessions. Ces propriétés unaires sont définies par l'utilisateur (propriétés initiales  $IP_i$ ), et elles constituent alors la perception qu'a le système de son environnement. Elles peuvent également être calculées par le système à partir de formules (propriétés propagées  $PP_i$ ) et constituent le langage de qualification .

Du fait de notre approche ensembliste, nous ne manipulerons pas les propriétés comme des applications  $P$  de  $O$  dans  $\{\text{vrai, faux}\}$ , mais plutôt comme des ensembles d'objets  $\{x / P(x)=\text{vrai}\}$ . Calculer une propriété revient alors à déterminer l'ensemble des objets antécédents de "vrai" selon la propriété en question.

Nous avons montré, au paragraphe 4.5.3, en quoi il s'agit d'une simple variante d'une formulation en logique du premier ordre. Nous y avons également précisé les conditions pour lesquelles cette formulation ensembliste est plus opératoire qu'une formulation purement logique.

A priori, nous devons donc gérer trois éléments : des objets (chaînes de caractères), des propriétés unaires (ou classes d'objets), des propriétés binaires (relations entre objets). Une méthode naturelle pour homogénéiser ceci est d'associer à chaque propriété unaire (statique, initiale ou propagée) un *objet fictif* qui sera relié aux objets possédant ladite propriété par un lien "système" dénommé *sorte-de* . Chaque propriété binaire  $BP_i$  sera représentée par un lien "utilisateur", de nom  $bpi$  , unissant les objets de  $O$  concernés.

On dispose alors d'une structure à deux composants (et non plus trois) : les objets (objets de  $O$  + objets fictifs) et les liens ( $bpi$  + *sorte-de* ).

La figure 2.1 illustre ce processus de simplification.

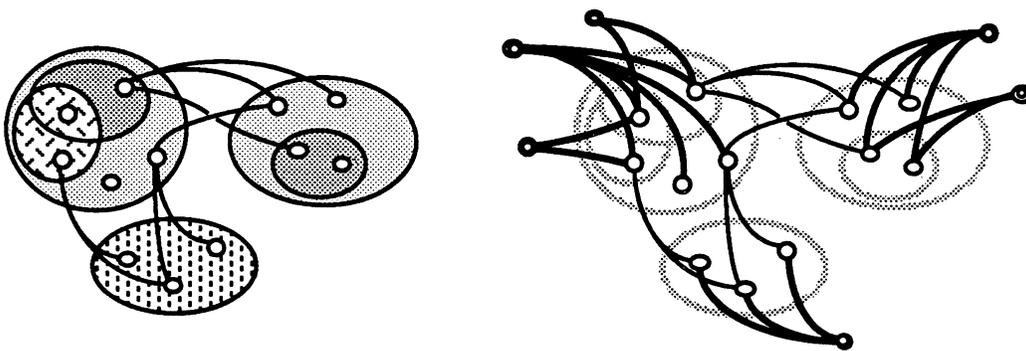


figure 2.1: simplification classes/relations

Parmi les objets fictifs se trouvent des représentants des propriétés unaires statiques, dont les liens *sorte-de* sont invariables, des représentants des propriétés unaires initiales, dont les liens *sorte-de* sont affectés par l'utilisateur et des représentants des propriétés unaires propagées, dont les liens *sorte-de* sont établis par le système (après évaluation de la formule  $F_{pp}$  associée à la propriété propagée en question).

Afin de faciliter la manipulation des objets ainsi définis, nous pouvons créer quelques objets supplémentaires ainsi que d'autres liens "systèmes". Par exemple, tous les objets représentant des classes (des propriétés) de maladies (M-Graves, M-rares, M-exotiques, etc) peuvent être regroupés sous un objet #Maladie qui serait alors l'équivalent d'une méta-classe.

Nous pouvons également créer un objet qui regroupe tous les IP-objets (méta-classe des classes initiales), un autre qui regroupe tous les PP-objets (méta-classe des classes propagées), etc. Il est laissé au concepteur de la base de données toute latitude pour structurer ses objets d'une manière plus riche que celle strictement exigée par le formalisme.

L'annexe D présente un exemple de la meilleure manière, selon nous, d'organiser une telle base de données. L'exemple y est volontairement éloigné du problème diagnostique. Un autre exemple de base de données sera présenté dans le chapitre 3 de cette seconde partie, dans le cadre de l'application GBDI-IPLSYS.

A cette structure de données, il faut adjoindre la spécification des formules  $F_{ppi}$ . Cette spécification est actuellement réalisée sous forme procédurale, directement dans le langage hôte.

D'autre part, à chaque  $IP_i$  est associé un protocole de saisie tenant compte de la signification, pour un utilisateur, de la propriété correspondante. Nous verrons sur un exemple que la représentation d'une action unique, du point de vue de l'opérateur, comme "faire un examen" peut correspondre à la modification simultanée de plusieurs propriétés  $IP_i$  (un nouvel examen devient "fait", de nouveaux symptômes deviennent "observés", parmi ceux-ci certains deviennent "effectivement constatés" et les autres sont "absents"). Ces protocoles de saisie ont pour but de simplifier l'interaction avec l'utilisateur et de garantir une certaine cohérence des données. Par exemple, si un examen a été effectivement réalisé, tous les symptômes qui lui sont associés doivent être connus comme présents ou absents. Il serait incohérent de considérer qu'un symptôme observé reste inconnu (c'est du moins ce qui a été décidé dans le cadre de l'application sur les fièvres au long cours).

L'application GBDI est donc principalement un système de gestion d'une base de données totalement relationnelle (les objets n'y ont pas de valeur, seules les relations entre objets sont porteuses d'informations) associé à un évaluateur de formules  $F_{ppi}$  et à un évaluateur de protocoles de saisies. Ces deux dernières entités sont définies sous forme procédurale, dans le langage hôte.

## 2.2. Développement et modularité

Du fait de la représentation procédurale des  $F_{ppi}$ , nous serons amenés à séparer deux types de modifications de portée nettement différente : celles qui n'imposent pas de réécriture des formules  $F_{ppi}$  ; celles qui imposent, ou sous-entendent, une réécriture de ces formules. Les premières peuvent être effectuées par simple modification de la base de données, alors que les secondes imposent une reprogrammation partielle (protocoles de saisies et méthodes de calcul).

### 2.2.1. Extensions ne modifiant pas les formules

Il est facile, en observant la grammaire que nous avons définie, de noter que les modifications suivantes ne remettent pas en cause l'expression des formules :

- 1 : ajout ou suppression d'un objet de  $O$  (objet non-fictif) ;
- 2 : modification des valeurs d'une application  $SP_i$  ;
- 3 : modification des valeurs d'une application  $BP_i$ .

Ces trois types de modifications correspondent en réalité à la même finalité. Supposons que l'on ajoute un nouvel objet  $o_i$  à la base de données. Il va falloir donner une signification à cet objet, or, dans notre formalisme, seul l'attribution de propriétés est susceptible de personnaliser cet objet. Ces propriétés peuvent être unaires (ce nouvel objet est-il un symptôme, une maladie ?), ou binaires (si cet objet est un symptôme, à quelles maladies est-il associé ?).

Quoi qu'il en soit, comme la détermination des  $PP_i$  est basée sur les propriétés et non sur les objets eux-mêmes, il n'est pas nécessaire de modifier les  $F_{ppi}$ . L'adjonction d'un nouvel objet va bien entendu modifier les résultats obtenus (à  $IP_i$  identiques). Ainsi, si l'objet est une maladie, il pourra maintenant apparaître parmi les maladies possibles, étant donné un jeu de symptômes constatés.

### 2.2.2. Extensions complexes

A l'inverse des modifications précédemment présentées, les actions décrites ci-dessous sous-entendent une réécriture des formules.

- 1 : ajout ou suppression d'une propriété statique unaire (un objet fictif) ;
- 2 : ajout ou suppression d'une propriété statique binaire (un lien) ;
- 3 : ajout ou suppression d'une propriété  $IP_i$  (un objet fictif modifiable) ;
- 4 : ajout ou suppression d'une propriété  $PP_i$  (un objet fictif calculable) ;
- 5 : modification explicite d'une formule  $F_{ppi}$ .

Un exemple du cas (1) est le suivant : dans un système de diagnostic, on désire introduire une certaine priorité entre les maladies fondées sur un critère C donné (la gravité, le coût des soins, la fréquence, etc). La première étape consiste bien entendu, pour toutes les maladies connues, à préciser celles qui satisfont le critère C (création d'un objet fictif  $o_C$  relié par un lien *sorte-de* aux maladies considérées). Si l'on s'arrête à ce point, nous pouvons dire que C "existe" mais n'est pas pris en compte. Selon la signification que l'on donne à C, il va falloir modifier les formules  $F_{ppi}$  afin d'introduire cette notion de priorité en confrontant ce critère à d'autres critères préexistants.

Les cas (2) et (3) se comportent d'une manière identique avec la complexité supplémentaire, pour le cas (3), d'établir un protocole de saisie adapté.

Le cas (4) peut s'avérer plus simple. En effet, on peut simplement ajouter un nouvel ensemble calculable, une nouvelle propriété propagée, sans pour autant modifier l'expression des ensembles précédemment connus. Il va de soi qu'il faut donc créer une nouvelle formule  $F_{pp}$  mais cela ne remet pas forcément en cause les anciennes expressions.

Le cas (5) pourrait consister, par exemple, en la modification de l'arbitrage entre deux critères concurrents  $C_1$  et  $C_2$ . Imaginons par exemple que l'on estime que le système privilégie trop la détection des maladies socialement coûteuses aux dépens des maladies graves. Il va alors être nécessaire de reprendre l'expression de certaines formules afin de rééquilibrer ces deux critères.

### 2.2.3. Extension de la grammaire

Notre formalisme contient d'emblée des fonctions de comptage (cardinal d'un ensemble) mais il ne contient pas nécessairement toutes les fonctions indispensables à la résolution d'un problème. Rappelons que la détermination des ensembles calculables reposait sur une liste de fonctions de base, de type ensembliste.

```
ensemble --> e_SPi | e_IPi |
              intersection( ensemble, ensemble ) |
              union( ensemble, ensemble ) |
              exclusion( ensemble, ensemble ) |
              tout_intersection( ensemble, ch_BPi ) |
              tout_union( ensemble, ch_BPi ) |
              Si val_booléenne Alors ensemble Sinon ensemble.
```

Les cinq fonctions de base ici présentées ne permettent pas de coder une assertion comme "si plus de 50% des maladies possédant la propriété X possèdent également la propriété Y...". L'évaluation d'un pourcentage nous fait défaut.

Nous pouvons proposer l'alternative suivante : soit il existe une liste, de taille raisonnable, de fonctions susceptibles de satisfaire toutes les applications concevables et cette liste peut être incluse dans une unique grammaire ; soit les différentes fonctions de base sont spécifiques de chaque application et il est alors nécessaire de les considérer comme des données du problème. Actuellement, nous ne pouvons fournir de réponse. Seule la réalisation d'un grand nombre d'applications pourrait nous renseigner sur ce point.

Remarquons toutefois que cette interrogation relative au nombre minimal de fonctions de base ne remet pas en cause les propriétés intrinsèque du formalisme. La recherche d'une grammaire minimale, d'un point de vue formel, n'est pas significative. Nous avons vu au paragraphe 4.6 que la seule forme "Si ..Alors.." suffit à assurer la complétude (mémorisation brute de toutes les situations descriptibles et de chaque diagnostic associé). Les autres fonctions de base ne sont donc, en aucun cas, réellement nécessaires. Au mieux, elles permettent une expression plus concise. La recherche d'une grammaire optimale obligerait à définir précisément ce critère de "concision".

Le caractère "incomplet" du formalisme permet en fait d'y introduire, comme éléments de base, des méthodes potentiellement complexes qui, chez un expert du domaine, sont perçues comme des opérations élémentaires. Par exemple la recherche des causes communes d'un ensemble de symptômes (ou le calcul des symptômes discriminants d'une liste de maladies possibles) est une opération qui n'a pas besoin d'être expliquée à un utilisateur ayant un minimum de compétence en médecine. C'est pour cette raison qu'il n'est pas nécessaire qu'elle soit explicitement représentée dans le système. A l'inverse les interactions entre des critères de coût, de gravité ou de fréquence sont des équilibres sujets à caution qui doivent donc être accessibles et modifiables lors de la mise au point.

### 2.3. Aperçu des structures de données

Nous avons vu précédemment comment ramener la représentation des classes (des propriétés unaires) et des relations (propriétés binaires) à une unique forme : objets / liens. Voyons maintenant en détail la représentation interne d'un objet, d'un lien et des triplets objet-lien-objet. Un objet est l'adresse mémoire (absolue) d'une cellule définie comme suit :

```
Objet =   record
          pname : string(20);
          commentaire : CharsHandle;
          lprop : ListPtr;
          end;
```

```
ObjPtr = ^Objet;
```

Le champ "pname" contient la chaîne de caractère qui constitue la représentation-utilisateur de l'objet. GBDI n'utilise pas de table des symboles au sens classique en cela qu'il n'accède jamais à un objet par l'intermédiaire de son nom (GBDI ne manipule que des ObjPtr). Par exemple, lorsque l'utilisateur sélectionne un objet dans une liste, en "cliquant" dessus (cf. interface Macintosh), l'indice N de l'objet dans la liste nous permet d'adresser directement le N<sup>ième</sup> élément d'une liste de pointeurs et d'obtenir ainsi l'adresse de l'objet.

Le champ "commentaire" adresse un texte qui permet à l'utilisateur de décrire complètement ce qu'il entend par le descriptif placé dans pname. Ainsi le simple mot "fièvre" pourra être précisé par un descriptif du contexte dans lequel on situe ce mot (ex : fièvre modérée de longue durée, inférieure à 39° pendant une dizaine de jours), à un ensemble de référence bibliographiques, voire à des notes purement locales (numéro de téléphone, etc). Ce commentaire n'est en aucune façon utilisable par le "système expert" qui manipulera la base.

Enfin, le champ "lprop" pointe sur une liste de listes. Chacune de ces listes comprend, comme premier élément, l'adresse d'un lien, suivi d'une séquence d'adresses d'objets qui sont relié à l'objet initial par le lien en question. Un lien est lui-même un objet particulier dont le champ lprop est vide (en fait, ce champ contient uniquement des pointeurs particuliers de type "système"). La représentation n'est donc pas du type "obj<sub>1</sub>-lien<sub>1</sub>-obj<sub>a</sub>, obj<sub>1</sub>-lien<sub>1</sub>-obj<sub>b</sub>, ..." mais plutôt du type "obj<sub>1</sub>-lien<sub>1</sub>-{obj<sub>a</sub>, obj<sub>b</sub>, ...}". Ce détail nous fera gagner un temps précieux lors des calculs ensemblistes.

La structure résultante est schématisée figure 2.2.

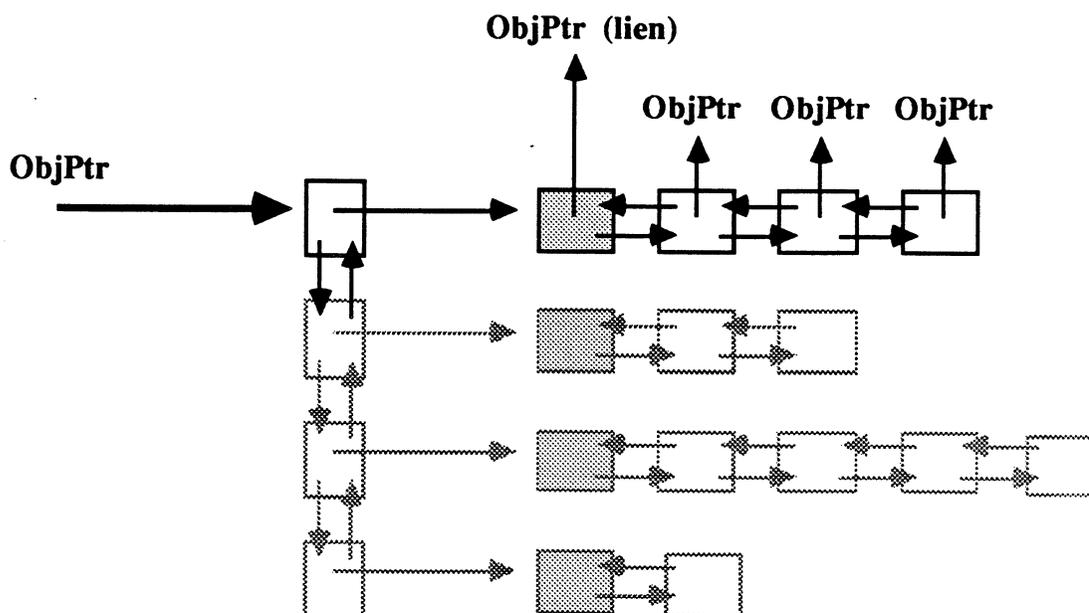


figure 2.2: implantation d'un objet

Contrairement à LISP, toutes les listes manipulées par GBDI sont des listes doublement chaînées, ce qui permet d'optimiser de nombreuses procédures.

Le système ne connaît, de manière absolue, que deux entités : #SYSTEM-OBJ#, le père de tous les objets selon un lien système (#system-prop#), et LISTPROP, la liste des liens (qui ne contient initialement que #system-prop#).

La création d'un nouveau type de propriété binaire consiste à étendre la liste des liens (une adresse est alors associée à la propriété en question). La création d'un objet consiste en l'allocation d'une cellule-objet et en l'ajout de l'adresse de cette cellule dans la liste commençant par #system-prop# du champ de #SYSTEM-OBJ#.

Afin, une fois de plus, d'optimiser les accès aux données, toutes les liaisons sont bi-directionnelles. Donc la relation "obj<sub>1</sub>-lien<sub>1</sub>-{..., obj<sub>x</sub>,...}" implique l'existence de la relation symétrique "obj<sub>x</sub>-lien<sub>1</sub>-{..., obj<sub>1</sub>,...}". Ce choix peut aisément être modifié, mais, outre qu'il est pertinent dans le cadre de notre formalisme à deux niveaux (objets de base, propriétés unaires), il permet une grande convivialité d'accès et de nombreuses optimisations.

La figure 2.3 montre une structure qui ne contient que deux objets. L'existence même de ces objets se traduit par les trois relations :

```
#SYSTEM-OBJ# - #system-prop# -> { obj1, obj2 }
obj1         - #system-prop# -> { #SYSTEM-OBJ# }
obj2         - #system-prop# -> { #SYSTEM-OBJ# }
```

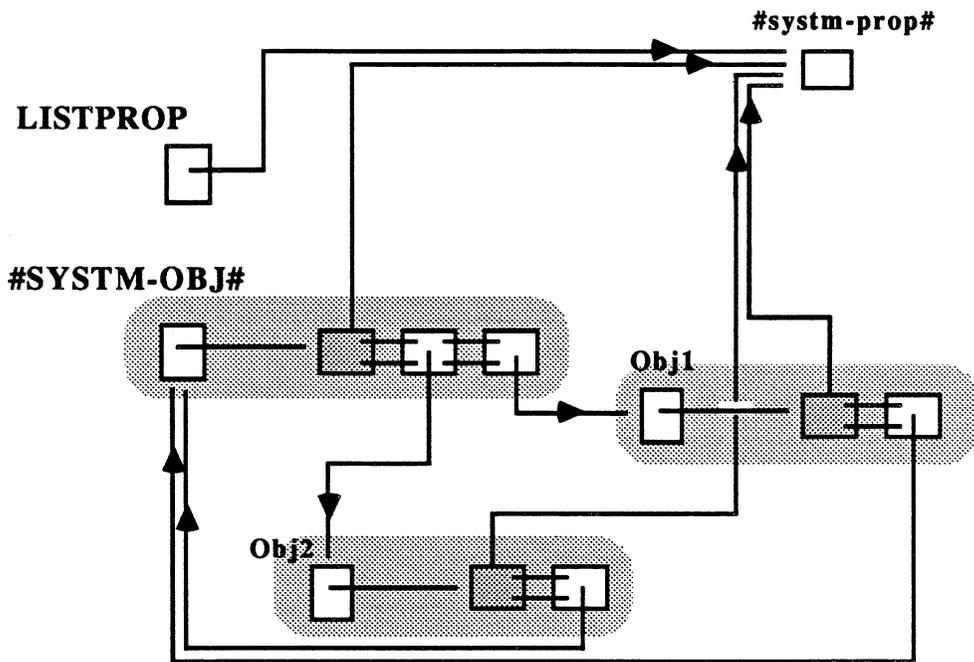


figure 2.3 : accès aux objets

### 3. GBDI-IPLSYS, une application médicale

Nous avons présenté un formalisme de "propagation de propriétés", puis une implantation de ce formalisme. Il est nécessaire de présenter maintenant une application complète, un exemple de fabrication d'un système opérationnel à partir de GBDI. Cette application, GBDI-IPLSYS, se place dans le cadre du projet S.A.P.I.E.M. (Système d'Aide à la Prescription et à l'Interprétation d'Examens Médicaux), mené conjointement par la société G.S.I. (Générale de Service Informatique), l'Institut Pasteur de Lyon ainsi que par le L.I.F.I.A.

#### 3.1. Buts poursuivis

Ce prototype, a été réalisé principalement avec le Dr. Cloppet, de l'Institut Pasteur de Lyon, et le Dr. F. Vandenesch. Il a pour but de fournir une aide à la prescription d'examens complémentaires dans le cadre restreint, mais néanmoins vaste, des fièvres au long cours.

Il est connu, dans les milieux de la Santé, que nombre d'examens sont prescrits à tort, soit parce qu'ils sont prématurés (demande de scanner sur des bases insuffisantes) soit, plus simplement, parce qu'il était possible de déduire *a priori* qu'ils ne pouvaient fournir aucun résultat significatif.

D'autre part, il a été mis en évidence un phénomène de "sédimentation". D'anciens examens sont toujours demandés, alors qu'ils pourraient avantageusement être remplacés par des examens plus modernes, plus fiables, voire moins coûteux. Les médecins intègrent mal, dans leur pratique quotidienne, l'apparition de nouveaux examens.

Enfin, une importante part des erreurs médicales sont imputables à de simples oublis. Ces oublis se situent souvent lors de l'interrogatoire (anamnèse) où, par exemple, un médecin peut omettre de se renseigner sur le régime alimentaire de son patient et perdre ainsi un renseignement crucial. De nombreux articles médicaux insistent d'ailleurs sur la nécessité de reprendre totalement et périodiquement cet interrogatoire dès lors que le diagnostic tarde à se préciser.

Les buts d'un système d'aide sont donc multiples. Il nous appartient de concevoir un outil qui puisse satisfaire le maximum de fonctionnalités. Les effets espérés de GBDI-IPLSYS, ou de tout outil similaire basé sur GBDI, sont les suivants :

- réduction du nombre d'examens para-cliniques demandés ;
- création d'un vecteur de communication facilitant la prescription d'examens modernes ;

- formation des internes sur un outil de type "simulateur de cas" ;
- amélioration de la qualité et de la rapidité du diagnostic par réduction des oublis.

Comme il était inenvisageable de concevoir un système couvrant toute la pathologie infectieuse (des millions de maladies, des milliers d'examen), nous avons réduit l'application à un sous-domaine néanmoins ambitieux : les fièvres au long cours.

Ce domaine est lui-même encore trop vaste pour qu'il soit possible ne serait-ce que de créer une base de données complète. On obtient une situation paradoxale où il est plus simple de préciser des méthodes (des formules  $F_{ppi}$ ) que d'énoncer exhaustivement les objets du domaine. Néanmoins, ce domaine était bien choisi car, plus que tout autre, il impose une clarification des protocoles d'investigations. Les fièvres au long cours regroupent la plupart des pathologies connues sous une forme atypique ; la seule protection contre une *errance* du diagnostic est donc une progression rigoureuse dans un protocole bien compris (ne pas prescrire de scanner avant une analyse de sang, s'occuper en priorité des maladies graves, etc).

Nous nous sommes donc imposés a priori deux contraintes fort constructives :

- 1 - ouverture vers de grosses bases de données ;
- 2 - explicitation des "méthodes" expertes.

### 3.2. Données de GBDI-IPLYS

Le mode de représentation des données est conforme au formalisme du chapitre 4 de la première partie, moyennant les quelques particularités d'implantation décrites au chapitre 2 de la présente section. Nous nous intéresserons ci-dessous au jeu de données tel qu'il existe.

Rappelons que GBDI-IPLSYS constitue un système de démonstration visant à valider la méthodologie de développement. Les données qu'il contient ne sont donc pas nécessairement "correctes" d'un point de vue strictement médical.

Une portion de la base de données est présentée figure 3.1.

On retrouve dans cette base un ensemble d'informations statiques comme :

- l'hépatite-A est une maladie ;
- la recherche de signes hématologiques est un examen ;
- l'adénopathie est un symptôme (une spécification) ;
- la découverte d'une adénopathie évoque l'hépatite-A ;
- la recherche de signes hématologiques nous renseignera sur une adénopathie.

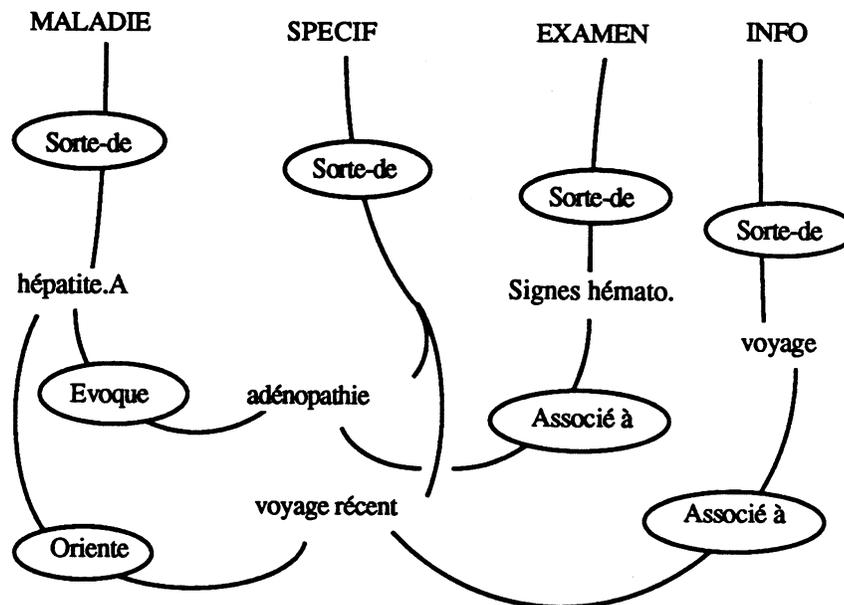


figure 3.1 : Base de données GBDI-IPLSYS

Comme nous l'avons déjà esquissé au chapitre précédent, parmi les multiples liens unissant les objets, il existe des liens statiques et des liens dynamiques (figure 3.2). Ces derniers permettent aussi bien de coder des informations d'entrée (informations fournies par l'opérateur) que des informations de sortie (conclusion du système à destination de l'opérateur).

Les informations initiales ne sont pas entrées par l'utilisateur directement comme l'établissement de multiples liens, mais par l'intermédiaire de protocoles de saisie dénommés "faire un examen", "poser une question". Une fois ceci fait et sur demande expresse de l'utilisateur, l'évaluation du jeu de formules  $F_{ppi}$  peut produire des informations comme :

- l'hépatite-A est une maladie évoquée par les symptômes constatés ;
- l'hépatite-A est retenue (ou non) comme maladie possible ;
- la recherche de signes hématologiques est un examen conseillé.

L'activation d'un protocole de saisie comme "faire un examen" déclenche un processus qui va alterner une requête à la base de données et une action de l'utilisateur. On peut représenter ce processus de la façon suivante :

action opérateur	requête B.D.
activation "faire un examen"	
	extraction des "EXAMEN<--sorte-de-->?"
choix d'un examen <i>e</i>	
	extraction des " <i>e</i> <--associé-à-->?"
classement des symptômes en vrai/faux	
validation globale	

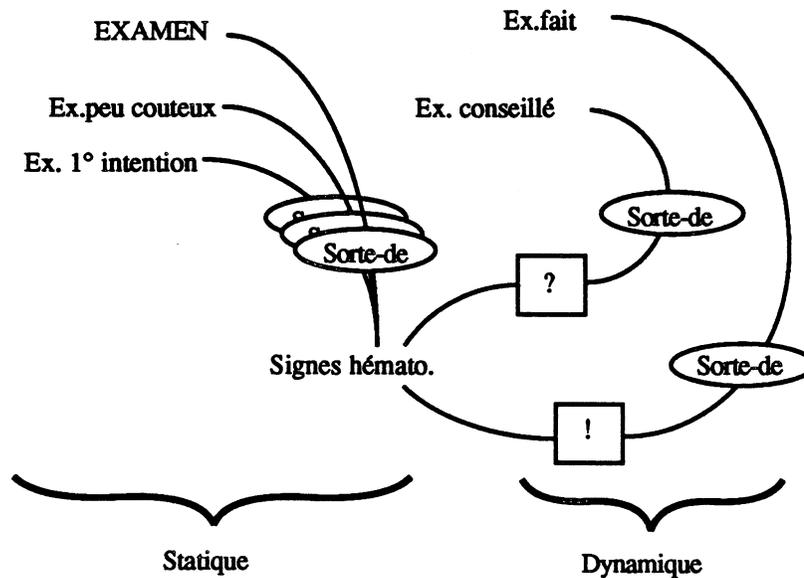


figure 3.2 : Entrées-sorties

Il y a alors création de multiples liens comme :

```
e <--sorte-de-->Ex. fait
si <--sorte-de-->spécif-vrai (ou fausse, selon le classement fourni)
...
sj <--sorte-de-->spécif-vrai ..
```

En fin de ce processus il y a donc eu modification des propriétés initiales "Ex.fait", "spécif-vrai" (symptômes observés et constatés présents) et "spécif-fausse" (symptômes observés et constatés absents).

Ce protocole exploite donc certaines propriétés unaires statiques et binaires statiques afin de faciliter la saisie des propriétés initiales et de garantir une certaine cohérence opérationnelle. Par exemple, on peut imposer que, si un examen a été pratiqué, les résultats associés doivent être connus (vrais ou faux). Ces contraintes doivent être établies en collaboration avec un expert du domaine ; elles capturent une partie de la sémantique globale du problème.

Dans le cas de GBDI-IPLSYS, il existe six protocoles de saisie nommés de la façon suivante :

"faire un examen"	voir exemple détaillé ci-dessus
"poser une question"	idem avec les "INFO <--sorte-de-->"
"récapitulatif"	modification des spécifications déjà connues
"oublier examen"	ôte un examen de Ex.fait
"oublier fait"	ôte une information de Info.connue
"tout oublier"	initialisation de toutes les propriétés initiales

Outre ces protocoles, il existe une unique commande d'activation du mécanisme de propagation : "orientation diagnostic". Cette commande déclenche l'évaluation de formules  $F_{ppi}$  qui déterminent les maladies évoquées, celles qui sont retenues comme possibles ainsi que les examens et questions conseillés.

En l'occurrence ces formules sont tout à fait schématiques et ne traitent que des configurations simples : recherche de causes communes ; heuristiques simples ; quelques méthodes de gestion de contradiction. Nous allons voir maintenant, en détail, des exemples de transformation de méthodes, initialement exprimées en langage naturel, en formules ensemblistes.

### 3.3. Méthodes et calcul ensembliste

#### 3.3.1. Présentation générale

Du point de vue d'un médecin, il est possible de situer les différentes méthodes d'exploitation de connaissances factuelles parmi trois grandes classes :

- 1 - les recherches de causes communes, de symptômes discriminants ;
- 2 - les heuristiques de réduction ;
- 3 - les méthodes de gestion de contradictions.

La première classe est le coeur même du processus diagnostic. Elle découle directement du principe de parcimonie (recherche d'une "explication" minimum). Ce premier processus peut produire deux types de configurations : soit il n'existe aucune explication commune du fait de contradictions entre symptômes, soit il existe un trop grand nombre d'explications communes pour qu'il soit possible de les considérer toutes. Dans le premier cas, il faudra "élargir" en négligeant certains symptômes soigneusement choisis (classe 3), dans le second cas, il faudra "réduire" en utilisant certaines considérations a priori comme "préférer les maladies courantes" (classe 2).

Globalement, la classe 1 relève d'une exploitation de la notion de "couverture logique". La classe 2 s'appuie essentiellement sur le dénombrement (on ne désire qu'un nombre limité de solutions concurrentes) et la classe 3 peut être vue comme une pré-condition de la recherche de causes communes (on désire avoir au moins une solution, quitte à dégrader nos informations initiales).

Il est important de bien comprendre que cette classification n'est qu'imaginée. Elle catégorise certaines caractéristiques du raisonnement des experts du domaine en

s'appuyant sur des critères qu'ils ont eux-mêmes fabriqués. Dans le formalisme que nous avons défini, il n'est ni possible, ni nécessaire, de classer formellement ces trois classes de méthodes.

D'autre part, on constate qu'une propriété comme "le fait d'être possible, pour une maladie" contient, étant donné la compréhension intuitive que nous en avons, une limitation quant au nombre d'objets qu'elle englobe. Une caractéristique des maladies possibles, telles qu'elles sont fournies par les experts, est d'être ni trop nombreuses, ni trop peu nombreuses. Les propriétés que possède un objet ne lui sont pas intrinsèques mais dépendent également de ses concurrents.

De telles expressions s'implémentent relativement facilement dans le cadre de la grammaire définie dans la première partie. Pour nous en convaincre, examinons rapidement comment les trois classes pré-citées s'intègrent dans une représentation ensembliste.

La figure 3.3 donne un exemple où l'on cherche une explication commune à trois symptômes donnés. Il est évident que, s'il existe des maladies communes à ces trois symptômes, ces dernières deviennent de meilleures explications que les maladies concurrentes. Ces causes communes s'obtiennent simplement par intersection de tous les ensembles de maladies associées à chaque symptôme constaté présent (fonction *tout-intersection* ).

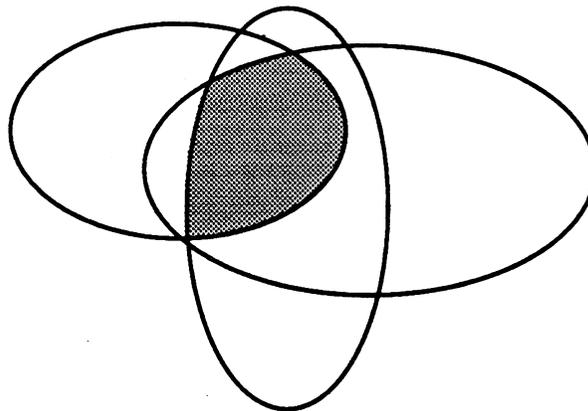


figure 3.3 : Recherche de causes communes

Voyons maintenant un exemple de représentation ensembliste d'une heuristique comme "s'occuper en priorité des maladies graves et curables" (figure 3.4). A partir d'un ensemble de maladies, causes communes des symptômes constatés, on effectue successivement l'intersection avec l'ensemble des maladies graves, connu a priori, puis l'exclusion des maladies incurables, également connues a priori. Au dire des experts, l'ensemble restant représente les "meilleures" hypothèses, non pas en termes de réelle possibilité, mais relativement à la détermination d'un protocole d'investigation. Il serait considéré comme une faute d'effectuer un examen visant à

confirmer la présence d'une maladie incurable, alors qu'il reste des hypothèses de maladies graves mais curables.

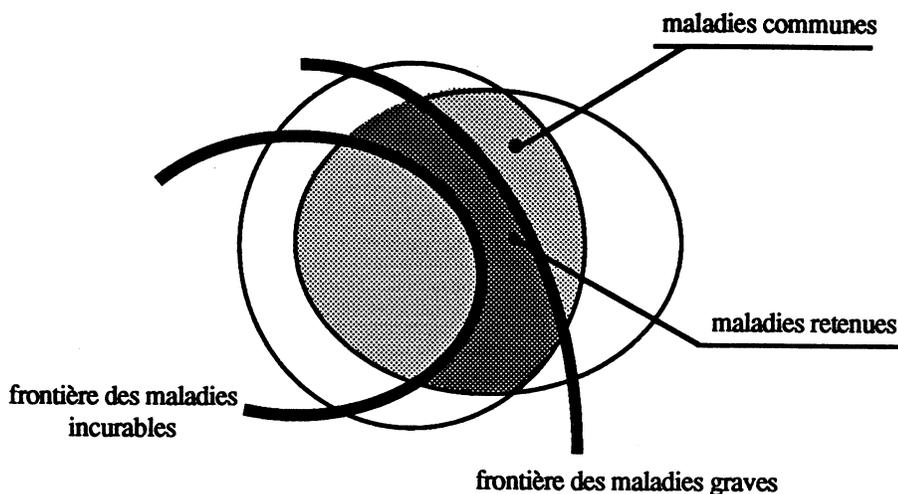


figure 3.4 : Heuristique

L'usage effectif de cette heuristique sera limité par une pré-condition (il faut que l'ensemble initial de maladies possibles soit assez grand) et par une post-condition (il faut qu'il reste un nombre minimum d'hypothèses concurrentes en sortie). Ces conditions, ici limitées à de simples dénombrements, peuvent aisément être exprimées selon la grammaire proposée.

Supposons maintenant que les symptômes constatés soient contradictoires. Ceci peut se détecter simplement par l'application de la recherche d'une cause commune. Si le résultat est vide (voir exemple figure 3.5) alors il ne sera pas possible d'expliquer simultanément tous les symptômes.

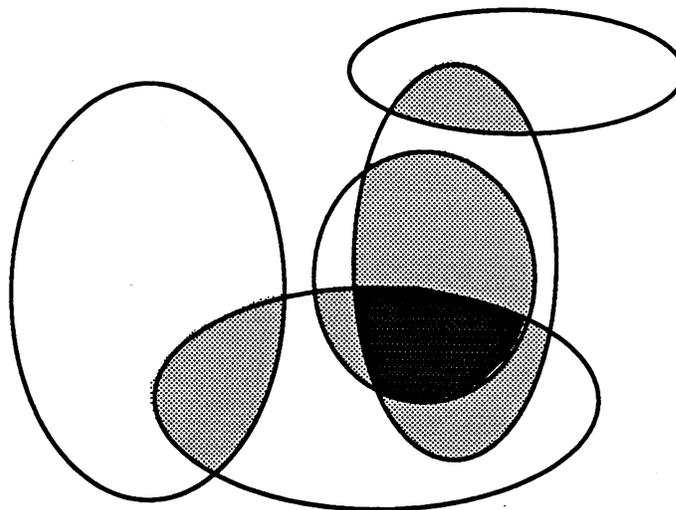


figure 3.5 : Détection d'une contradiction

En l'absence d'informations complémentaires, il est impossible, en toute rigueur, d'exhiber des hypothèses.

Dans une application réelle, il est néanmoins rare que toutes choses soient "égales par ailleurs". Il est toujours possible de se référer à d'autres propriétés. Par exemple les symptômes ne seront pas d'égale fiabilité (la fiabilité est considérée connue a priori), les maladies couvertes ne seront pas d'égale gravité, etc. On peut donc introduire des méthodes de gestion de contradictions sous-tendues par les idées du type suivant : tant qu'à négliger des symptômes, autant négliger les moins fiables en priorité ; tant qu'à négliger des symptômes et donc écarter les maladies correspondantes, autant négliger ceux qui se réfèrent à des maladies bénignes, etc.

Les trois figures qui suivent illustrent d'autres exemples de stratégies de gestion de contradictions qui, outre leurs expressions ensemblistes, disposent d'une représentation graphique didactique. Ces trois méthodes sont respectivement : le rejet des informations "isolées", le rejet des informations les moins spécifiques, l'acceptation d'une association morbide. Ces trois méthodes sont basées sur des notions de dénombrement et de couverture ; elles se programment facilement.

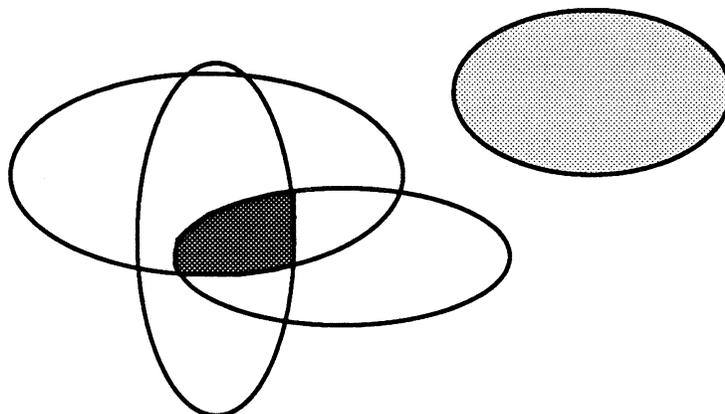


figure 3.6 : Rejet de l'information isolée

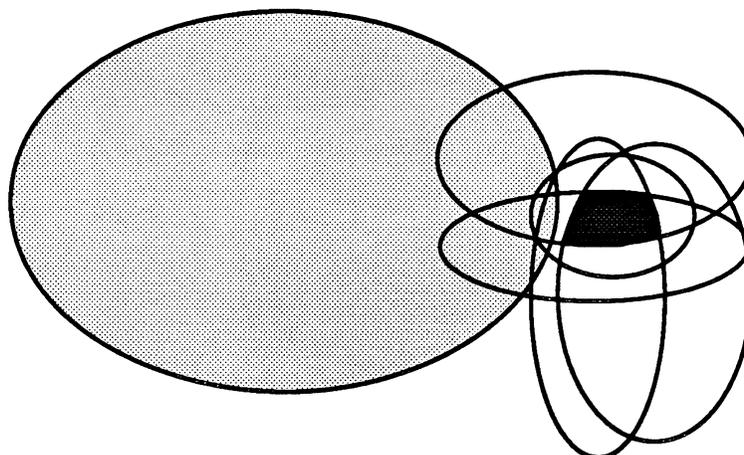


figure 3.7 : Rejet d'information peu spécifique

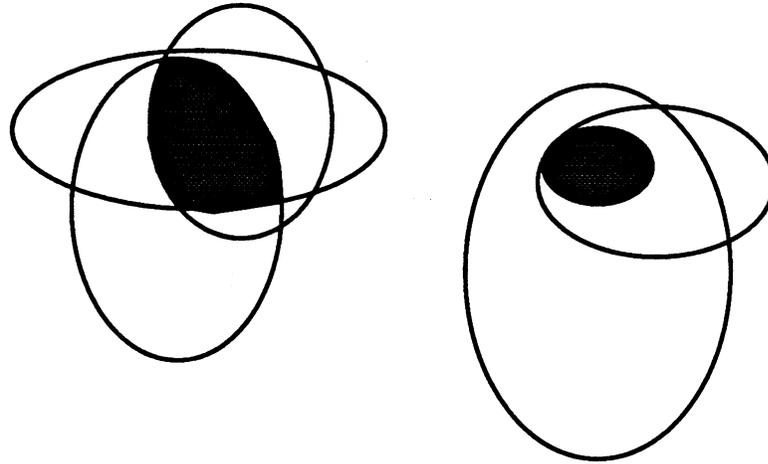


figure 3.8 : Acceptation d'une association morbide

La troisième méthode mérite une explication approfondie. Il peut paraître étrange de gérer une contradiction en modifiant le problème posé (remplacement de la recherche d'une maladie par la recherche d'un couple de maladies) ; c'est néanmoins une méthode effectivement utilisée en médecine. Certes, le cas est rare ; il faut une contradiction d'un type bien précis pour forcer le recours à cette méthode. Par exemple, il faut disposer de deux groupes de symptômes bien disjoints, chaque groupe ayant une forte cohérence interne. En outre, c'est une méthode "balai", toutes les autres méthodes étant censées avoir échoué.

Le formalisme que nous proposons permet de caractériser, par une expression ensembliste, le type de contradiction nécessaire pour qu'il soit possible d'envisager une association morbide. Néanmoins, une fois ceci fait, le système devient incapable de manipuler correctement le problème. En effet, il faudrait ensuite examiner l'ensemble des couples de maladies possibles, se restreindre aux associations morbides plausibles, etc. Or, notre formalisme ne manipule que des ensembles d'objets et non des ensembles de doublets.

Il serait possible d'envisager une reconception du formalisme afin d'y introduire cette notion de "maladies multiples". En tout état de cause, cela entraînerait un accroissement important de la complexité, qui serait d'un faible apport au niveau de l'opérateur. Généralement, la seule acception de la possibilité d'une association morbide conduit le médecin au diagnostic final. Le but du système est donc plutôt de rappeler cette possibilité à l'opérateur, dès lors que les contradictions s'accumulent, et ceci afin d'éviter une errance prolongée du diagnostic. Si cette possibilité est écartée par l'opérateur, cela signifie néanmoins que le système a atteint ses limites en matière de gestions de contradictions (ceci peut être dû à la limitation, voire à l'incorrection, des connaissances factuelles, soit à l'emploi de méthodes erronées). Le médecin devra alors continuer seul le processus diagnostique.

Nous avons déjà indiqué que les connaissances factuelles ne pouvaient acquérir de sémantique que par l'explicitation des méthodes de manipulation des ces connaissances. Par exemple, une propriété statique unaire comme "maladie grave" ne prend de sens, du point de vue du système, que s'il existe au moins une méthode

y faisant référence, par exemple une heuristique de réduction du nombre d'hypothèses. De même, une propriété binaire comme "évoque" doit être exploitée d'une manière distincte d'une autre propriété comme "affirme", si l'on veut pouvoir affirmer l'existence de deux liens distincts.

D'une manière analogue, les diverses méthodes sont dépendantes les unes des autres. Lors de la présentation de l'exemple d'heuristique (voir figure 2.4), nous avons insisté sur le fait que les maladies finalement retenues ne sont pas réellement plus "possibles" que les autres ; elles ne sont retenues prioritairement que relativement à la détermination d'un protocole d'investigation correct, au dire des experts. La recherche des maladies possibles intègre implicitement un but ultérieur : le choix d'un examen qui devra se référer à une ou plusieurs de ces maladies possibles.

En dernier recours, les propriétés terminales du système (celles qui sont calculées, mais jamais utilisées dans d'autres calculs) puisent leur signification dans la présence de l'opérateur. C'est celui-ci qui va donner un sens à la propriété "examen conseillé" ; en tout état de cause, pour le système, cette propriété n'a pas de sémantique.

On en arrive à la constatation étrange, mais non inattendue, qu'il est impossible de décrire les connaissances factuelles sans faire référence aux méthodes qui les emploient, que ces méthodes sont dépendantes les unes des autres et que les propriétés terminales dépendent de la présence de l'opérateur. En bref, il est impossible de décrire uniquement une partie du système sans se priver des moyens de lui associer une signification. A posteriori, cela paraît relativement normal, cela découle naturellement de la notion de "système". En revanche, cela s'oppose à l'idée usuelle de granularité des connaissances.

Note : nous avons bien conscience que, selon le point de vue où l'on se place, de telles déclarations peuvent paraître hérétiques ou tautologiques. Un système est bien sûr, par essence, une synergie d'informations ; à ce titre, il est non-décomposable. Inversement, pour décrire un système, on se doit de négliger cette synergie et de considérer, pour les besoins de l'exposé, les différentes parties comme peu dépendantes ; c'est classiquement la décomposition en niveaux. Lorsqu'un système repose sur un formalisme mathématique fort (par exemple la logique) le deuxième point de vue est très opératoire, la synergie étant totalement assumée par la théorie. A l'inverse, dès lors qu'un système repose sur un formalisme ad-hoc, toute décomposition risque de le dénaturer. Il est en effet impossible de renvoyer l'utilisateur à une interprétation globale avec un argument du type "le système est basé sur telle logique, voici nos particularismes ; voir la théorie complète pour une interprétation globale".

### *3.3.2. Méthodes effectivement implantées*

La base de données contient quatre types de liens : oriente, évoque, évoque fortement, affirme. Nous voulons donner à ces liens les signification intuitive suivante :

- orientation : "un voyage en Afrique oriente vers les maladies tropicales". Donc, des maladies qui seraient improbables sans ce fait deviennent "évoquées" (équivalent au "placement en mémoire à court terme" ?). Ces maladies évoquées ne sont pas nécessairement "possibles" (ce n'est pas parce que l'on voyage que l'on tombe malade) mais elles seront utilisées, si besoin, par des heuristiques de réduction du nombre d'hypothèses. En l'absence d'informations plus précises, on se restreindra à des examens simples (anamnèse).

- évocation : "une adénopathie évoque une hépatite". Si un symptôme évoque certaines maladies, on devra considérer au moins une de ces maladies comme "possibles" ou bien admettre que ce symptôme reste inexpliqué. Ce lien n'est toutefois pas assez fort pour entraîner l'acceptation d'une association morbide.

- évocation forte : "une modification d'un souffle (cardiaque) ancien évoque fortement une endocardite". Ce lien est semblable à l'évocation simple, sauf que l'apparition de deux groupes de symptômes relativement cohérents entre eux peut, sous certaines conditions, entraîner l'acceptation d'une association morbide.

- affirmation : "la présence de bacilles de Koch affirme la tuberculose". Si des maladies sont affirmées par un symptôme, l'une d'entre elles est, non pas "possible", mais "certaine". On admettra une association morbide dès lors que deux maladies sont affirmées, sans autres conditions. S'il existe des symptômes "affirmant", le système peut énoncer quelques conseils mais doit se déclarer incompétent pour la suite de l'investigation.

On voit, dans cette définition intuitive, que l'on se réfère en permanence au caractère opératoire de ces liens. La signification se confond avec l'usage.

Les données initiales sont constituées de deux groupes de symptômes dits "Spécif.vrai" et "Spécif.fausse" correspondant à un groupe d'examen "Ex.fait".

Les propriétés propagées sont les classe de maladies "M.évoquée", "M.possible" et "M.certaine", une classe d'examens "Ex.conseillé" et une classe de questions (objets de type Info) "Info.a.demander".

En premier lieu, on examine le nombre de symptômes connus (Spécif.vrai). S'il n'y en a aucun, le système affiche un message d'erreur et les propriétés propagées ont pour valeur l'ensemble vide.

S'il existe au moins un symptôme connu, on distingue trois cas :

- 1 - il existe au moins un symptôme qui "affirme" une maladie ;
- 2 - il n'existe que des symptômes qui "orientent" ;
- 3 - autre cas (pas de "affirme", mais des "évoque" et "évoque.fort").

Dans le cas 1, on est donc en présence d'un ou plusieurs symptômes très précis ou même pathognomoniques. Si ces symptômes admettent une cause commune le système affiche un message indiquant que ces informations doivent suffire pour poser le diagnostic, sans qu'il soit nécessaire de poursuivre les examens complémentaires. Les maladies possibles sont les (la) cause(s) commune(s) obtenues ; le groupe "Ex.conseillé" devrait logiquement rester vide, on y place néanmoins les examens en rapport avec les maladies possibles.

Dans le cas où ces symptômes "affirmant" sont contradictoires, on indique la certitude d'une association morbide.

Un autre cas de reconnaissance d'association morbide est le suivant : les symptômes "affirmant" sont cohérents, il existe d'autres symptômes qui "évoquent fortement", également cohérents entre eux, mais contradictoires avec ceux qui "affirment".

Dans le cas 2, on ne dispose que d'informations de faible qualité (conditions de vie, alimentation, etc) et d'aucun symptôme bien précis. Ces informations sont considérées comme trop faibles pour se renforcer mutuellement ; on renoncera donc à appliquer une recherche de cause commune. Les maladies évoquées seront toutes celles en rapport avec un quelconque des symptômes ; "M.possible" sera un sous-ensemble des maladies évoquées, obtenu par application de l'heuristique suivante : on se limite, si possible, aux seules maladies graves et curables (regroupées dans l'unique classe "Maladie de première intention"), à concurrence de cinq hypothèses. Les examens conseillés sont ensuite sélectionnés parmi les examens en rapport avec ces maladies, une préférence étant marquée pour les examens qui produiront éventuellement des symptômes qui "affirment" une des maladies possibles. Actuellement, il n'y a pas de critère de type "Ex.coûteux" qui pourrait venir contrecarrer cette préférence.

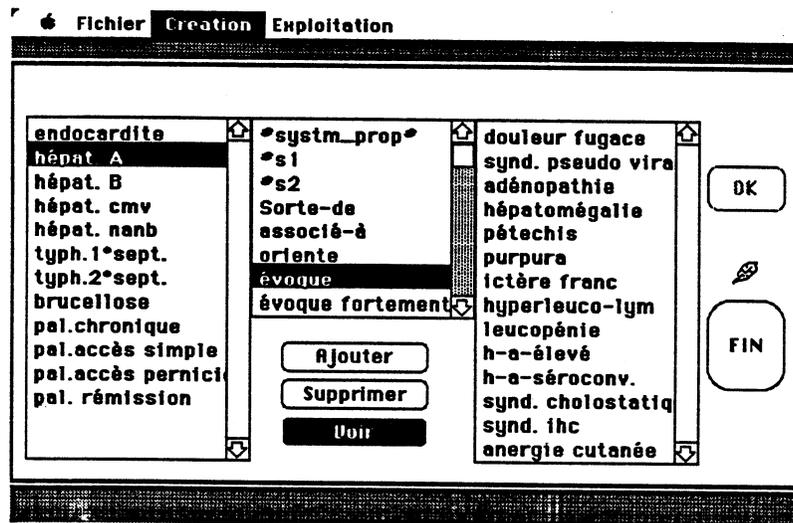
Le cas 3 est intermédiaire en cela que, si les symptômes admettent des causes communes, on se limite à celles-ci. Sinon, on essaye de trouver une "meilleure cause commune", en négligeant jusqu'à 5% des symptômes constatés. Si, finalement, on retient au moins une hypothèse, celle-ci est admise comme maladie possible et les examens conseillés sont choisis comme précédemment. Si la contradiction reste insoluble, un message en informe l'opérateur et l'on adopte une approche "maximaliste" du même type que le cas 2. On voit ici que l'on peut utiliser deux fois la même méthode dans des contextes différents.

### 3.4. Exemples de session

Il est difficile de présenter une "session" au sens classique. GBDI est un logiciel implanté sur Macintosh et il en utilise toutes les possibilités (menus, multi-fenêtres, raccourcis, etc).

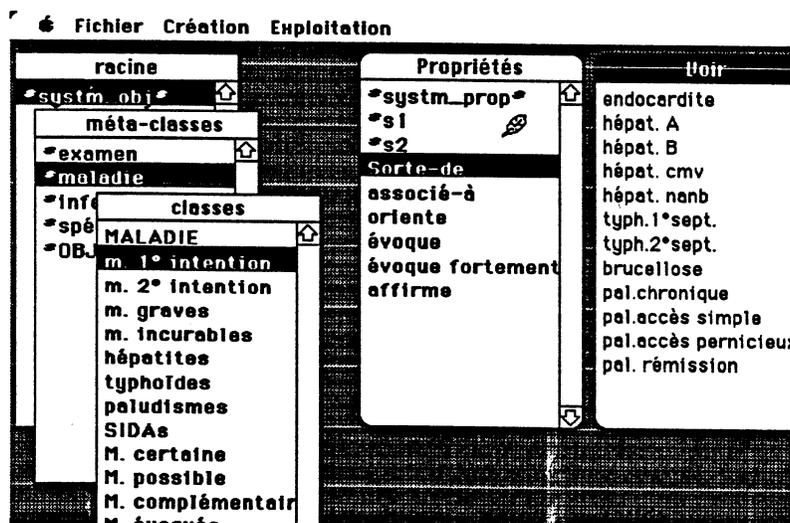
Nous présentons donc ci-dessous une série de copie d'écrans associés de quelques commentaires explicatifs. Il ne faut néanmoins pas oublier que les différentes modalités de GBDI (système expert, base de données) se superposent et s'interpénètrent. Par exemple, une "explication" relative à une activation des méthodes peut être acquise, moyennant un effort de l'utilisateur, par une requête (composée "graphiquement") à la base de données. La permutation entre ces deux modes est d'ailleurs fortement encouragée, ne serait-ce que pour son caractère didactique.

Ecran correspondant à la commande **Lier-Délier** (menu **Création**).



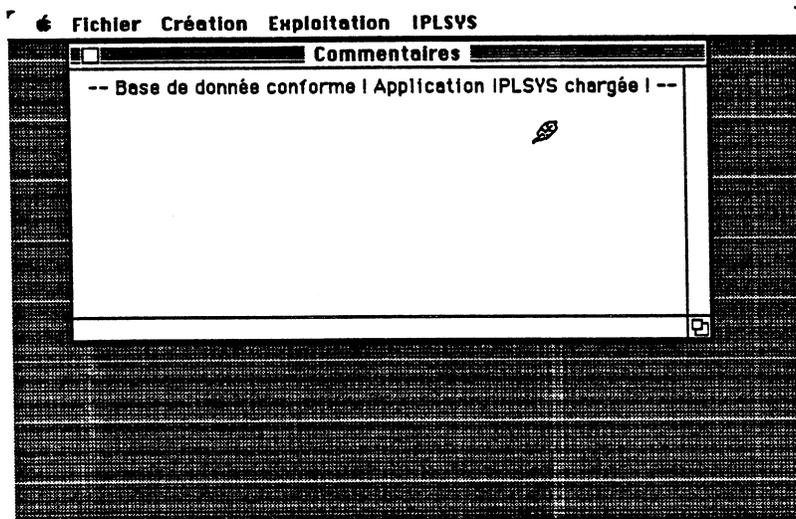
Cette commande permet de gérer l'ensemble des liens unissant les objets entre eux (objets de base, représentations des propriétés unaires, autres objets "commodes" pour l'accès aux données).

Ecran correspondant à **Début-exploitation** (menu **Exploitation**).



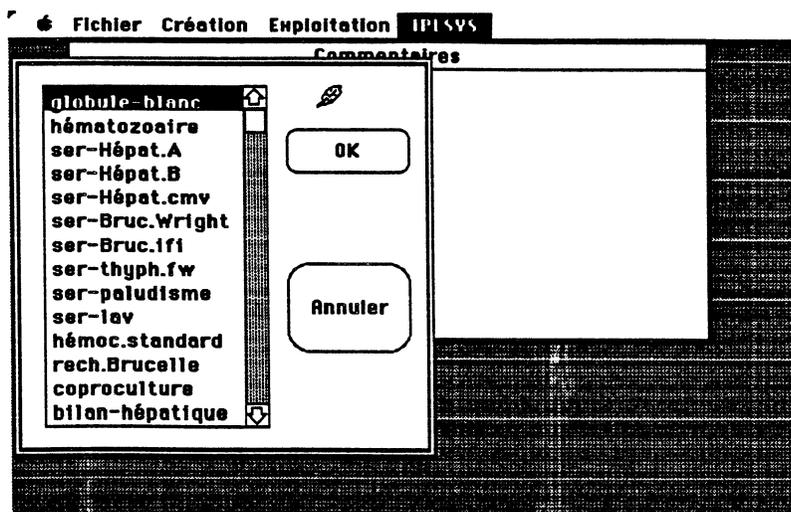
Ce mode permet la composition de "requêtes graphiques" à la base de données. Il est possible, par exemple, d'obtenir l'ensemble des symptômes qui "affirment" une maladie quelconque ainsi que l'ensemble des maladies en question, tout symptômes confondus. Il n'existe pas à proprement parler de "langage de requête" mais une série d'actions élémentaires, dont le résultat est un ensemble d'objets, de tels ensembles pouvant servir de base à une autre action.

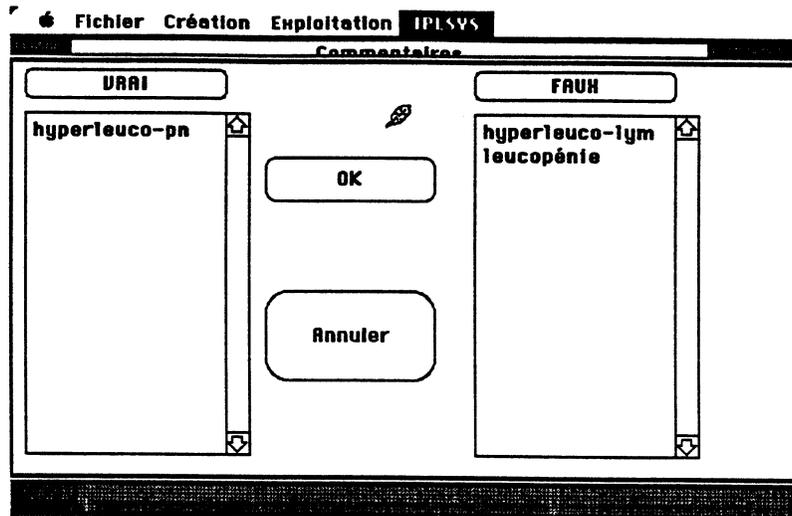
Ecran visible après activation du module IPLSYS.



Il est évident que les méthodes d'IPLSYS ne peuvent être activées que si la base de données contient bien les objets (propriétés unaires) ainsi que les liens (propriétés binaires) qui figurent dans l'expression procédurale des méthodes. Donc, l'activation du module IPLSYS débute par une vérification de la base de données disponible.

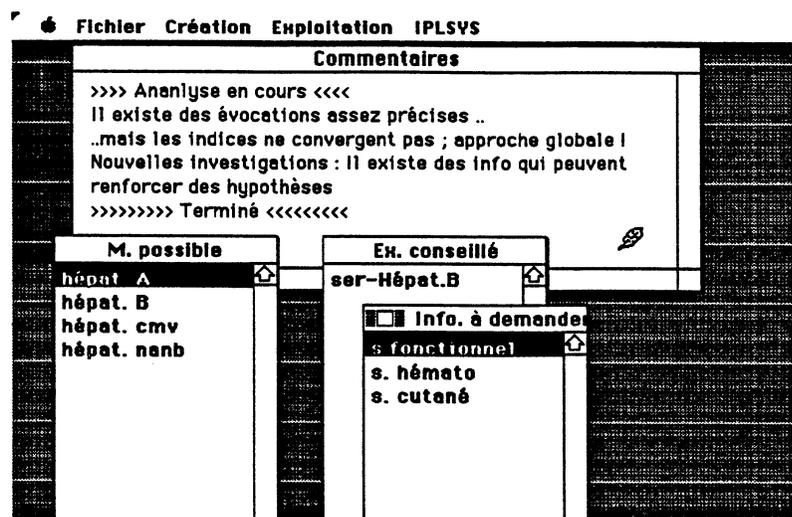
Protocole d'action **Faire un examen** (menu IPLSYS)





Deux étapes sont nécessaires : choisir l'examen ; affirmer la présence, ou l'absence, des symptômes "associés-à" l'examen choisit. Comme un même symptôme peut être perçu à partir de plusieurs examens différents, il est possible qu'après avoir choisit un examen, un des symptômes associés soit déjà dans la colonne de gauche (symptôme affirmé présent).

Activation des méthodes, **orientation diagnostic** (menu IPLSYS) suivi d'une demande de visualisation des résultats obtenus, **voir tous résultats** (menu IPLSYS).



L'ensemble de ce logiciel a été écrit en TML-PASCAL, à défaut de C ou LISP qui ne furent disponibles que six mois après le début de ce travail. L'usage de primitives évoluées (application event-driven, fenêtres, menus, etc) a profondément transformé la conception même que nous avions d'un outil d'aide au diagnostic. Une interface pauvre entraîne un dialogue homme-machine rigide d'où la nécessité de fonctions d'explications, de traces multiples, etc. A l'inverse, l'usage d'une interface conviviale permet de dépouiller l'application. le programme résultant est très sobre malgré un débit d'informations élevé.



## **Conclusion**

## Récapitulatif et conclusion

Le présent chapitre tente de récapituler les points importants caractérisant l'ensemble de ce travail tant du point de vue conceptuel que formel ou pratique.

### 1. Apport conceptuel

Le diagnostic est un processus visant à l'acquisition d'informations non directement perceptibles. Il se place dans le cadre d'une volonté de maintenance et peut être caractérisé par le paradigme suivant :

$\text{MAINTENANCE} = [ [\text{OBSERVATION} + \text{DIAGNOSTIC}]^* + \text{DEPANNAGE}]^*$
---

Un des pré-requis à l'existence d'un diagnostic est l'existence d'une référence de bon fonctionnement. Si cette référence est définie par un ensemble de contraintes entre paramètres mesurables, on parle usuellement de "diagnostic à partir de modèle profond". Si cette référence est définie en extension par une liste de symptômes pathologiques, on se référera plutôt aux travaux de J. Reggia sur la théorie de la couverture. Toutefois, l'existence de ces deux pôles techniques ne doit pas faire oublier l'unité conceptuelle.

De manière générale :

- une panne est la qualification d'une classe de divergences entre des observations et une référence.
- le diagnostic est le processus d'acquisition d'informations qui vise à nommer la panne.

Les trois entités nécessaires pour la spécification d'un problème diagnostic sont : un langage de description, une référence de bon fonctionnement, un langage de qualification. Du point de vue conceptuel, le choix de ces trois entités est totalement arbitraire ; il n'y a aucune nécessité particulière de se référer à des composants ou des maladies. L'imagination est reine.

Outre ceci, le diagnostic s'appuie sur un ensemble d'informations permettant l'analyse des observations, la production d'hypothèses de pannes et la production d'examens complémentaires possibles, ce que nous avons regroupé sous le terme de "processus de qualification". Nous avons désigné cet ensemble d'informations

---

par le nom de "connaissances a priori". Là encore, de multiples choix techniques sont possibles.

Nous avons montré que la notion de "modèle" n'était pas nécessaire pour la spécification du diagnostic, sinon sous un aspect pour le moins inhabituel. En effet, la somme du langage d'observation, de la référence, du langage de qualification et des connaissances complémentaires peut être vue comme un "modèle" selon le sens défini dans [Cadoz 87, Raoult 86]. La totalité de ces informations nous évite d'écrire la liste exhaustive des observations possibles et de leur faire correspondre les pannes associées ; il y a codage et compression de l'information. De plus, ce codage est productif, en cela que l'on se dote d'un mécanisme capable d'évaluer les pannes à partir des observations. Quoi qu'il en soit, cette façon d'envisager un modèle ne fait aucunement appel à cette notion intuitive du modèle décrivant une certaine réalité.

## 2. Apport technique

Nous avons défini un système formel dénommé "système de propagation de propriétés". Ce système peut être vu comme une extension du formalisme des règles de production. Au lieu de ne propager que deux propriétés (le fait, pour un prédicat, d'être vrai ou faux) nous avons la possibilité de propager autant de propriétés que nécessaires. L'usage du calcul ensembliste peut être vu comme une alternative au CP1, très efficace sur des domaines finis.

Parmi l'ensemble des interprétations possibles, nous lui avons associé plus particulièrement une interprétation relative au diagnostic. Les propriétés initiales forment le langage d'observation ; les propriétés propagées constituent le langage de qualification ; les connaissances complémentaires sont codées sous la forme de propriétés unaires ou binaires (prédicats et relations). Le type de diagnostic visé se place assez nettement dans la catégorie "diagnostic symptomato-pathologique". Un effort important a été fait relativement au choix de nouvelles observations.

Tout ceci nous a fourni un système de diagnostic que l'on pourrait qualifier "d'entièrement symbolique". C'est un système conceptuellement proche de la théorie de la couverture parcimonieuse, mais ne faisant appel à aucun calcul de score basé sur des informations numériques. Nous avons justifié cette attitude par la distinction qui existe entre des grandeurs objectives, comme la fréquence réelle d'une maladie, et les grandeurs subjectives manipulées par les experts d'un domaine. Par exemple, la notion de fréquence subjective n'est pas dissociable du contexte opératoire de l'ensemble du processus diagnostic (comment manipule-t-on les maladies dites "peu fréquentes", qu'elles soient effectivement rares ou non ?).

Le système proposé est formellement complet. D'autre part, il a l'avantage de bien séparer l'ensemble des informations sur lequel s'appuie l'interprétation et l'ensemble des méthodes d'exploitation de ces informations. On pourrait le définir

---

comme un système de diagnostic à trois niveaux : objets, propriétés, méthodes. Les propriétés ne concernent que les objets, les méthodes ne manipulent que des propriétés.

### 3. Ouvertures possibles

#### 3.1. Apprentissage des méthodes

La phase la plus critique, lors de la réalisation d'une application en vraie grandeur, est l'écriture des "bonnes" formules (des  $F_{ppi}$ ). S'il est possible d'étendre aisément la grammaire de base afin d'intégrer des processus de calcul complexes, il n'est pas aussi facile d'envisager l'ensemble des interactions entre les diverses propriétés. De même, la spécification des relations qui unissent les différentes propriétés n'est pas simplement dérivable en procédures de calcul.

Il est difficile de dire si la meilleure solution réside dans l'utilisation de techniques d'apprentissage ou dans la création d'une interface système-concepteur qui permette de simplifier l'écriture des méthodes (langage de type "objets", programmation graphique ou autre).

Rappelons que notre but n'était pas uniquement de réaliser un système efficace (du point de vue de la correction des diagnostics effectués) mais surtout de définir une architecture qui permette de représenter correctement et explicitement un certain niveau de raisonnement. De ce point de vue, nous estimons, non pas avoir atteint notre but, mais avoir pris la bonne direction.

#### 3.2 Hybridation symbolique-numérique

Un système de propagation de propriétés, dédié à une application médicale, se base sur des requêtes comme "trouver N maladies qui représentent les meilleures causes communes des symptômes constatés". Dans le système que nous avons proposé, de telles requêtes sont réalisées par l'intermédiaire de calculs ensemblistes exploitant une base de données purement relationnelle. Il est extrêmement simple, dans ce cas précis, de substituer à ces méthodes un sous-système de quelque nature que ce soit, pour peu qu'il réponde à la requête en question. Il paraît alors évident que certaines requêtes puissent s'appuyer sur des sous-systèmes de type statistiques prenant en compte les données épidémiologiques du moment. Ceci serait particulièrement pertinent dans tous les domaines où existe un fond important de données purement statistiques, à défaut d'une expertise bien établie.

Le mécanisme de propagation de propriété peut alors être vu comme un répartiteur de requêtes sur des sous-systèmes spécialisés. La séparation entre

---

connaissances factuelles et connaissances expertes s'accroît d'autant. Cette hybridation conserve le caractère explicite du raisonnement simulé.



**Annexes  
&  
Bibliographie**

## Annexe A - Exemple de modélisation

Soit le circuit présenté figure A.1. Il est composé d'un générateur de tension pure (de valeur  $U_0$ ) pourvu d'un indicateur de court-circuit, d'une charge  $R$  (de valeur nominale  $R_0$ ) ainsi que des connexions nécessaires.

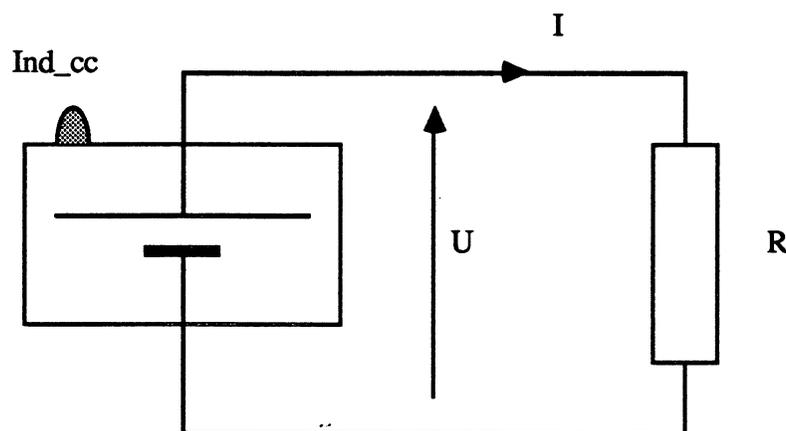


figure A.1 : un système analogique simple

Nous nous autorisons a priori quatre mesures : la tension aux bornes du générateur, le courant effectif, la valeur de la charge et la lecture du voyant de détection de court-circuit. Ces quatre mesures vont donc correspondre à quatre paramètres du modèle, respectivement :  $U$ ,  $I$ ,  $R$ ,  $Ind\_cc$ .

Notons rapidement que la mesure de  $R$  ne correspond pas à un relevé direct mais implique une manipulation du circuit (ouverture du circuit puis usage d'un ohmmètre). Il existe un coût potentiel qui n'est pas représenté ; à coût fixé, il existe donc des paramètres non-mesurables qui peuvent néanmoins apparaître dans le modèle sans que l'intérêt en soit évident.

A partir de ces quatre paramètres, nous pouvons exprimer un ensemble arbitraire de contraintes définissant le bon fonctionnement. Soit, par exemple, le jeu de contraintes suivant :

- C0 : [  $Ind\_cc = \text{faux}$  ]
- C1 : [  $(U = U_0 \wedge Ind\_cc = \text{faux}) \vee (U = 0 \wedge Ind\_cc = \text{vrai})$  ]
- C2 : [  $U - R \cdot I = 0$  ]
- C3 : [  $R - R_0 = 0$  ]

Ces contraintes expriment respectivement que le voyant doit être éteint, que si le voyant est allumé alors le générateur est en court-circuit, sinon il doit délivrer la

tension nominale, que la loi d'Ohm est respectée (nous reviendrons sur ce point) et que la résistance doit valoir  $R_0$ .

Nous pouvons alors doter le graphe associé (du fait des connexions implicites par variables) d'une représentation graphique telle que présentée figure A.2.

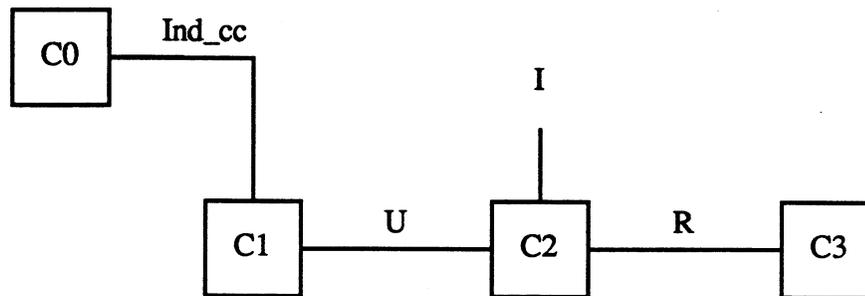


figure A.2: un modèle possible

La comparaison des deux figures A.1 et A.2 montre bien, à l'inverse de l'exemple classique des multiplieurs et additionneurs, qu'il n'existe pas de rapport direct entre un schéma de câblage et la représentation graphique du réseau de contraintes. Il va de soi que rien ne nous empêche de réécrire les contraintes d'une autre manière afin d'obtenir une autre représentation ; en groupant C2 et C3, on pourrait, par exemple, obtenir un modèle à trois "composants" C0, C1 et C2'.

$$C2': [ (U=R_0*I) \wedge (R=R_0) ]$$

Il n'existe pas nécessairement de rapport entre l'expression d'une ou de plusieurs contraintes et un quelconque élément physique (abstraction des composants). Le seul cas, extrêmement particulier, où apparaît une telle similitude se produit si chaque composant physique peut être décrit par une unique équation unissant ses entrées-sorties. Nous avons volontairement choisi un système analogique où la notion d'entrée-sortie est vide de sens et où un tel choix des contraintes est impossible.

Indépendamment de cela, chaque contrainte, si elle s'avère violée, peut être interprétée par un opérateur humain comme décrivant une panne spécifique. Prenons l'exemple de la contrainte C2 (la loi d'Ohm). Il peut paraître étrange d'exprimer une contrainte dont on sait qu'elle ne peut objectivement être violée. Or, dans le modèle  $\langle C0, C1, C2, C3 \rangle$ , si la contrainte C2 n'est pas respectée, on pourra en déduire que le branchement de la charge n'est pas correct (circuit ouvert, branchement parallèle d'une autre charge, etc), sinon la loi d'Ohm serait vérifiée. Si le modèle utilisé est  $\langle C0, C1, C2' \rangle$  il ne sera pas possible de distinguer entre une charge présentant une mauvaise valeur ( $R \neq R_0$ ) et une charge mal connectée. **Donc, si le passage d'un élément aux contraintes qui le caractérisent n'est pas simple (phase de représentation), il en est de même pour la mise en correspondance d'une contrainte violée avec une description de panne (phase d'interprétation des résultats).** Au pire, il est possible d'établir une liste

---

exhaustive de toutes les combinaisons de contraintes violées et de lui adjoindre une relation vers une liste de descriptifs de pannes exprimées dans un langage de qualification quelconque.

On remarquera également que la propagation, au travers des contraintes, des valeurs mesurées implique que les contraintes soient réversibles ; ainsi la contrainte C2' ne permet pas de déduire I à partir de U dès lors que R est différent de R<sub>0</sub>. Ceci est la conséquence logique du fait que C2' ne doit pas être interprété uniquement comme "la charge possède sa valeur nominale" mais comme "la charge possède sa valeur nominale ET est correctement connectée". Dans un système réellement complexe, ces difficultés de modélisation ainsi que les erreurs d'interprétation associées peuvent devenir critiques.

Nous espérons avoir ainsi argumenté notre opinion, à savoir que les modèles profonds ne peuvent pas être établis directement à partir d'un schéma de câblage ou d'un schéma fonctionnel. Ils sont le résultat d'un processus complexe de choix lié aux paramètres mesurables et aux pannes dont on envisage la détection. De fait, ces modèles ne sont guère plus "profonds" que d'autres du point de vue d'une réalité (physique ou fonctionnelle) de la description. Ce label de modèle profond est essentiellement, de notre point de vue, un terme consacré par l'usage. Il regroupe les systèmes de diagnostic dont les spécifications de bon fonctionnement sont définies par un ensemble de contraintes. Ces systèmes sont principalement axés sur la découverte de divergences entre observations et référence ; ils négligent le choix de nouvelles observations et le plein développement de la qualification de pannes (en terme de composants).

## Annexe B - Représentation minimale

Le problème posé est le suivant :

Soit une fonction  $F : P(S) \rightarrow P(M)$  ; existe-t-il toujours une fonction  $f : S \rightarrow P(M)$  et une méthode d'exploitation  $F_{PPM+}$  telles que l'on puisse vérifier  $F = F_{PPM+}(f)$  ?

Rappelons brièvement que la fonction  $F$  définit le "bon diagnostic" que l'on cherche à reproduire. La fonction  $f$  est une écriture fonctionnelle de l'unique propriété binaire  $BP_C$  que l'on s'autorise à exploiter. Enfin, la formule  $F_{PPM+}$  décrit un mécanisme de calcul ensembliste, exploitant  $BP_C$ , par lequel on tente de reproduire  $F$ .

La réponse est non. On ne peut se limiter à une seule propriété binaire et conserver la complétude. La démonstration repose sur une stratégie de dénombrement. Nous allons montrer que, étant donné le nombre de fonctions  $f$  possibles, il n'est pas possible de fabriquer un assez grand nombre de fonctions de  $P(S) \rightarrow P(M)$  pour couvrir toutes les fonctions  $F$  concevables.

Soit  $s$  et  $m$  les cardinalités respectives des ensembles  $S$  et  $M$ . Il existe  $2^{m2^s}$  fonctions  $F$  possibles. Il existe  $2^{ms}$  fonctions  $f$  possibles. On constate déjà une disproportion importante entre ces deux nombres. S'il était possible de borner le nombre de méthodes d'exploitation de la fonction  $f$  (de la propriété  $BP_C$ ), on obtiendrait le résultat voulu sans difficulté. Il est toutefois possible de concevoir une infinité de méthodes de manipulation, pourtant nous allons montrer que ces méthodes ne peuvent produire qu'un nombre limité de résultats distincts.

Soit  $S = \{s_1, s_2, \dots, s_s\}$ . A chaque  $s_i$ , on peut associer sa valeur  $f(s_i)$ . Toute méthode de manipulation ensembliste ne peut donc exploiter, en vue de la construction du résultat, que les ensembles suivants :  $f(s_1), f(s_2), \dots, f(s_s)$  et l'ensemble  $M$ . On dispose donc de  $(s+1)$  sous-ensembles de  $M$  qu'il est possible d'intersecter ou d'unir de toutes les manières concevables.

A partir de  $n$  sous-ensembles quelconques, on ne peut produire qu'au plus  $(2^n - 1)$  ensembles disjoints deux à deux. A partir de  $p$  sous-ensembles disjoints on ne peut produire que  $2^p$  ensembles distincts. Donc, à partir des  $(s+1)$  ensembles initiaux dont nous disposons, nous ne pouvons produire qu'au plus  $2^{(2^{(s+1)} - 1)}$  ensembles distincts. Nous pouvons donc réduire l'ensemble des méthodes de manipulation possibles à un ensemble fini de méthodes canoniques. Tout se passe comme si le nombre effectif de méthodes était borné. Associons à chaque instance

de S+ une méthode donnée. Le nombre de résultats distincts constructibles, pour une unique fonction f, est alors majoré par  $(2^s * 2(2^{(s+1)} - 1))$ .

Le nombre de fonctions de  $P(S) \rightarrow P(M)$  qui sont constructibles est borné par :

$$2^{ms} * (2^s * 2(2^{(s+1)} - 1))$$

L'inégalité suivante étant vraie dès que  $m > 2$  et  $s > 4$ , on peut en conclure qu'il existe des fonctions F non-reproductibles.

$$2m2^s > 2(ms+s-1+2^{s+1})$$

note : cette démonstration s'étend sans difficulté pour un nombre k donné de propriétés binaires. Il suffit de remplacer (s+1) par  $k*(s+1)$ . Donc, si l'on fixe a priori le nombre de propriétés binaires, il est possible, si m et s sont assez grands, qu'une fonction F donnée a priori soit non-reproductible.

Comme nous l'avons déjà dit, cette démonstration est assez peu constructive. Il est peu probable que l'on dispose jamais d'une fonction F, exprimée en extension, comme caractérisation d'un problème diagnostic. Au mieux on disposera d'une description partielle de F (un ensemble d'instances typiques et un ensemble de cas particuliers). De plus, on peut toujours repousser les limites de représentation en créant autant de propriétés binaires que nécessaires, étant donné les cardinalités m et s. Mais on ne saura pas déterminer pour autant quelles valeurs doivent prendre ces relations, sauf si l'on se rapproche, une fois de plus, d'une représentation en extension.

Une approche de type IA nous permettra d'associer une signification intuitive à chaque relation causale définie, ainsi qu'aux méthodes d'exploitation employées. Par exemple, le calcul d'intersection peut représenter la recherche de causes communes ; l'union peut être vue comme l'implantation du critère de pertinence ([Peng 86], voir étude bibliographique) ; etc. La spécification, en langage courant, de la signification d'une relation causale va alors restreindre les modes de calcul susceptibles d'exploiter cette relation. C'est principalement cette caractéristique du formalisme qui nous évitera de faire face à la combinatoire présentée ci-dessus.

## Annexe C - Exemple d'étude préliminaire

Le but de cette annexe est d'illustrer le caractère opératoire de la méthodologie que nous avons défini eau chapitre 2. Rappelons-le brièvement, cette méthodologie demande d'identifier a) les capacités perceptives, b) une référence de bon fonctionnement, c) un langage de qualification des pannes et d) un ensemble de "connaissances a priori". Une fois tout ceci fait, la réalisation effective d'un système de diagnostic consiste à implanter ces informations et à réaliser un algorithme qui, à partir des observations, élabore des hypothèses de pannes et, éventuellement, des propositions d'observations complémentaires.

Nous allons présenter ci-dessous un domaine de problème : le dépannage et la mise au point de circuits électriques domestiques. Nous illustrerons ce domaine par un exemple réel, accompagné du raisonnement ayant permis de le résoudre. Nous montrerons que ce problème peut être vu comme un problème diagnostic.

Le problème est le suivant : un individu X désire installer un radiateur électrique dans un garage. Ce garage comprend déjà une lampe (L), commandée par un interrupteur mural (I). L'ensemble des fils, c'est-à-dire une phase, un neutre, les deux fils de l'interrupteur et les deux fils de la lampe, sont regroupés dans un unique boîtier (voir figure F.1). Ces six fils ne sont pas repérés. Monsieur X positionne les deux fils du radiateur sur deux des plots du boîtier. Le radiateur est équipé d'un interrupteur propre (R). Le fonctionnement désiré est le suivant : l'interrupteur mural commande la lampe ; l'interrupteur du radiateur commande ce dernier . La figure F2 décrit un schéma de câblage qui réalise cette fonction.

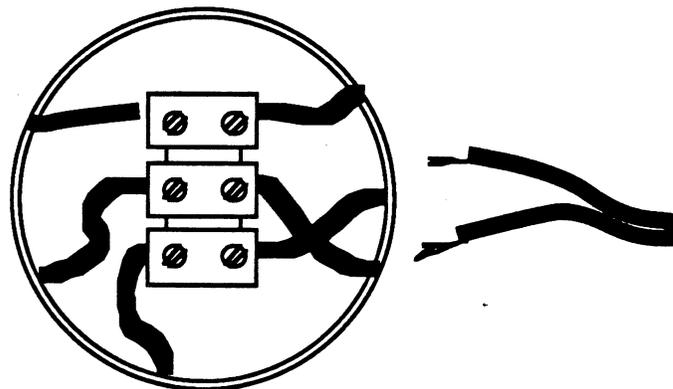


figure F.1 : le problème de connexion

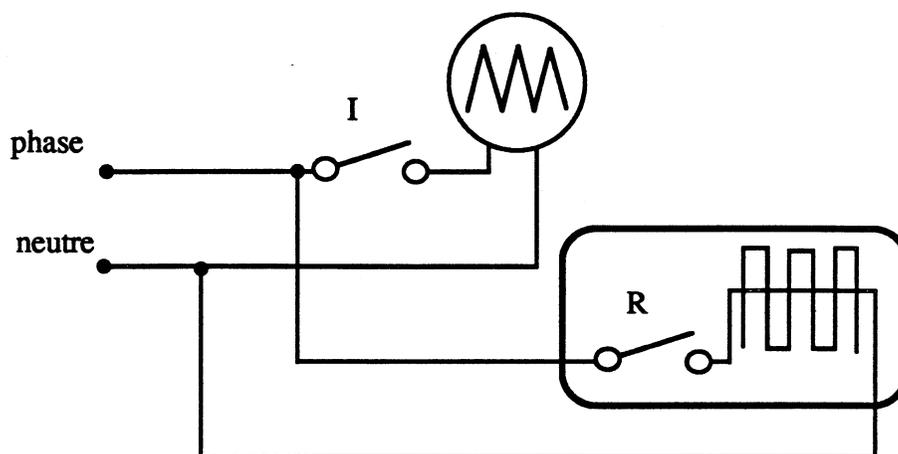


figure F.2 : configuration nominale

Une fois le branchement effectué, monsieur X teste le fonctionnement réel. Il constate ceci : pour une position de l'interrupteur R, l'interrupteur I commande bien l'allumage et l'extinction de la lampe ; pour l'autre position de R, et selon la position de I, on constate soit un faible allumage de la lampe et un échauffement du radiateur, soit un allumage total de la lampe et une extinction du radiateur. Décrétons que la perception que l'on a du montage électrique se limite à la description d'état<sup>1</sup> (état de la lampe, état du radiateur) et de transitions (une action sur un interrupteur entraîne un changement d'état). Décrétons également que la référence de bon fonctionnement est elle aussi définie par un graphe d'état donné a priori (ou par un graphe d'état déduit du schéma F2). La figure F3 contient, côte à côte, le graphe d'état nominal et le graphe observé. Ce dernier est complet, le montage ayant été complètement exploré ("complètement" au sens des capacités perceptives).

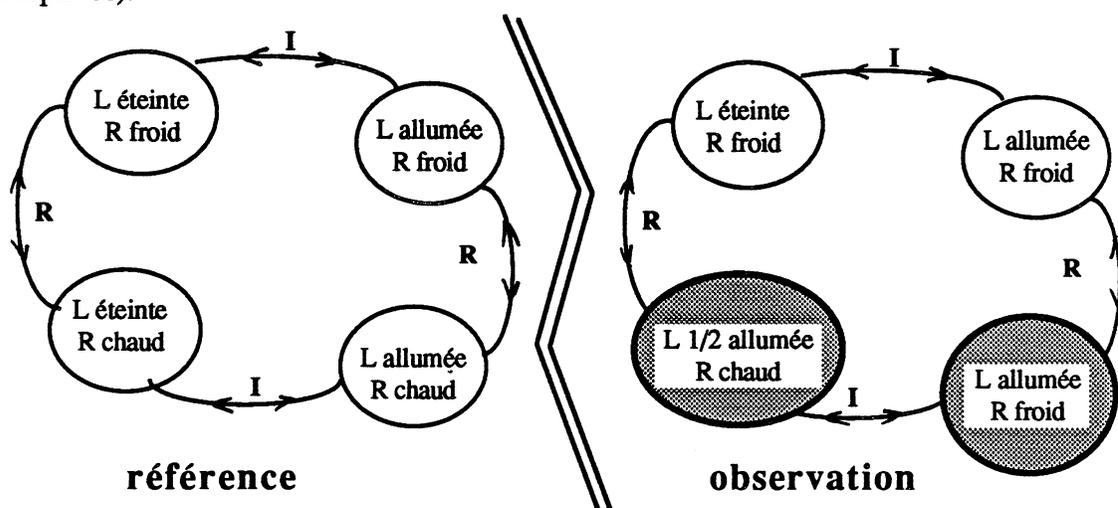


figure F.3 : graphes d'états

<sup>1</sup> Les positions on-off des interrupteurs I et R ne sont pas indiquées et n'apparaissent donc pas dans la description des états.

---

La différence entre les deux graphes affirme l'existence d'une panne selon la définition même que l'on a donnée à la notion de "référence de bon fonctionnement".

Il se trouve ici que cette panne n'est pas due à une perturbation inconnue, mais à une erreur imputable à monsieur X. D'un certain point de vue, cette distinction est sans importance. En effet, fixons-nous un langage de qualification de pannes qui exprime des "erreurs de branchements" (une panne sera, par exemple, décrite comme "branchement en parallèle des éléments X et Y alors qu'ils devraient être en série !"). Comme nous définissons une panne comme une qualification d'un écart entre une référence et une observation, l'erreur de monsieur X peut être considérée comme une panne et non comme une erreur de réalisation.

On voit ici que la différence entre le véritable dépannage et la mise au point est assez ténue. Dans les deux cas, il existe une fonction de maintenance (une référence à atteindre). Pour le dépannage, on suppose implicitement que l'objet observé a fonctionné antérieurement avant de tomber en panne, ce qui n'est pas le cas de la mise au point. Mais, une fois donnée la référence de bon fonctionnement, cette distinction s'efface.

La différence intuitive que l'on ressent entre ces deux activités est, en fait, effacée par le biais du langage de qualification de pannes. Pour le dépannage, ce langage fera appel à la notion de dégradation d'un composant (le terme composant est à prendre au sens large) avec la supposition que la structure est bonne. Pour la mise au point, ce langage sera relatif à des erreurs de réalisation, les composants étant supposés corrects.

Revenons à notre radiateur récalcitrant. Nous nous sommes fixé des capacités perceptives, une référence de bon fonctionnement et un langage de qualification de pannes. Bien entendu, la résolution effective de ce problème impose la connaissance de nombreuses informations a priori que nous ne détaillerons pas pour l'instant. Evoquons-les simplement par quelques propositions comme "un interrupteur permet de mettre en contact deux points d'un réseau", "si la phase est mise en contact avec le neutre, il y a court-circuit", "deux appareils branchés en série se partagent la puissance", etc.

Le raisonnement, mené par monsieur X, fut le suivant : il existe une configuration où la lampe et le radiateur sont en série, une première hypothèse est donc représentée par la figure F.4.

Si la configuration réelle était ainsi, une action sur I (ou sur R) éteindrait simultanément le plafonnier et le radiateur, or cette action allume totalement le plafonnier et éteint le radiateur. Donc cette hypothèse doit être rejetée.

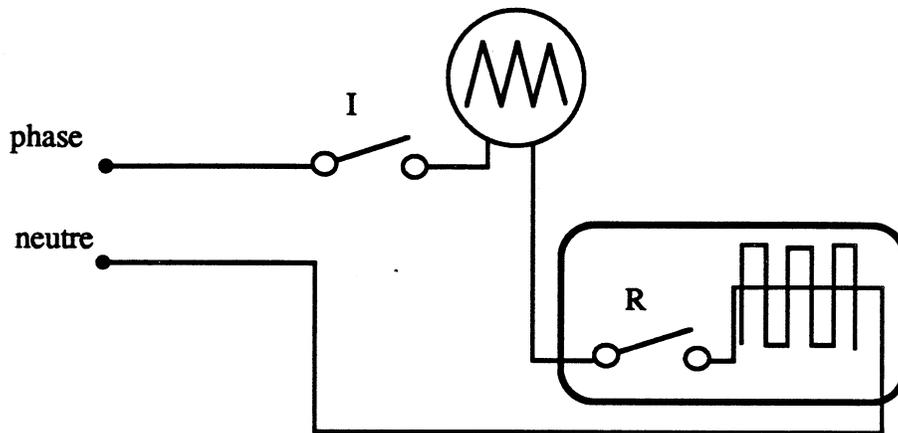


figure F.4 : première hypothèse

Prenons l'extinction du radiateur par action sur I ; il existe en fait deux méthodes pour éteindre un appareil électrique (figure F.5) : le montage série ou parallèle.

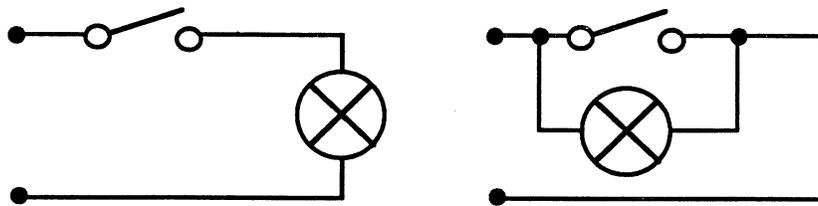


figure F.5 : deux méthodes "d'extinction"

La seconde configuration fournit la solution. En effet, le court-circuit des bornes du radiateur n'entraîne pas de court-circuit global, puisque la lampe est en série. La configuration déduite est celle de la figure F.6. Cette configuration correspond au graphe d'états effectivement observé.

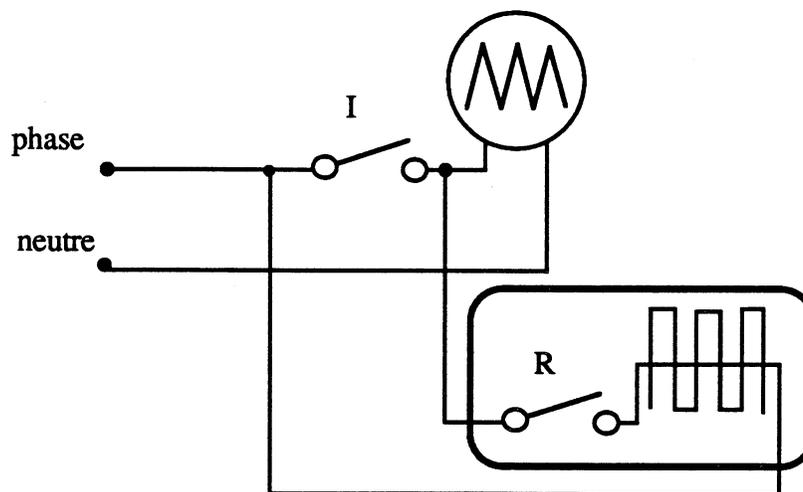


figure F.6 : solution finale

---

L'erreur de branchement, la panne, peut donc être décrite en ces termes : "le radiateur est branché en parallèle sur les bornes de I et non en parallèle sur la phase et le neutre".

Que peut-on penser de cet exemple ? Tout d'abord, ce type de problème pourrait donner lieu à l'étude d'un robot de maintenance totalement autonome. En effet, les primitives de perception sont relativement simples (détection, par une caméra, de l'état d'une ampoule ou de la position d'un interrupteur), de même que les actions de mesure (basculement d'un interrupteur). Enfin, le problème peut être complexifié à discrétion par ajout d'interrupteurs va-et-vient, de relais, etc.

Pour la réalisation du système de diagnostic, nous avons déjà explicité les capacités perceptives, la référence et le langage de qualification. Il nous reste à préciser les différentes connaissances a priori qui permettent de résoudre ce problème.

En première analyse, nous pouvons dire que ces connaissances permettent d'établir une correspondance entre une structure physique et une partie du graphe d'état. Elles permettent de relier des états et des transitions à des structures électriques et, réciproquement, de fabriquer le graphe correspondant à une structure donnée. Attention, nous ne disons pas que nous savons comment représenter ces informations. Néanmoins, tout système capable de résoudre le problème ci-dessus disposera nécessairement de mécanismes unissant les observations (graphes ou partie de graphes d'états) et les pannes (description d'erreurs de branchements). La philosophie globale de l'IA voudrait que ces informations soient modulaires et déclaratives ; c'est là un problème supplémentaire.

Récapitulons. Nous avons donc : des descriptions, éventuellement partielles, de graphes d'état comme capacités perceptives ; un graphe d'état comme référence de bon fonctionnement ; des relations complexes entre des éléments physiques et des propriétés du graphe d'état comme connaissances a priori ; un langage de qualification décrivant des erreurs de connexions.

Sans chercher à réaliser effectivement un tel système, nous avons examiné rapidement quelques exemples de connaissances a priori sur lesquelles pourraient reposer le diagnostic. Ces connaissances font correspondre une structure électrique à des propriétés des états ou des transitions entre états. Nous avons donné deux exemples ci-dessous :

Exemple de propriété sur les transitions : si un interrupteur est en série avec une lampe, tous les états du graphe où la lampe est "allumée" admettent une transition, par basculement de l'interrupteur, avec un état où la lampe est éteinte.

---

**Exemple de propriétés sur les états : si deux lampes sont branchées en parallèle, elles sont toujours simultanément éteintes ou allumées.**

Lors du diagnostic, les capacités perceptives nous fourniront une description partielle du graphe d'état effectif. Les états seront éventuellement incomplètement décrits ; on ne connaîtra pas nécessairement toutes les transitions possibles à partir d'un état.

De ces données, on devra tirer des hypothèses de structure. A chaque hypothèse, correspond un graphe complet incluant le sous-graphe exploré. Il reste à découvrir des mesures qui permettront de réduire, par enrichissement du sous-graphe, le nombre d'hypothèses. Si le système est totalement observé (le graphe effectif est complètement connu), alors les hypothèses concurrentes sont indiscernables.

On voit, dans l'exposé qui précède, que la méthodologie d'analyse du processus diagnostic permet de spécifier correctement un problème bien avant d'être capable d'en donner une solution opératoire. Il est possible d'identifier les capacités perceptives nécessaires, la référence de bon fonctionnement, le mécanisme de qualification et de donner une idée du mécanisme devant les mettre en rapport. Nous insistons sur le fait que cette décomposition conceptuelle n'est pas arbitraire : tout système de diagnostic peut nécessairement se décomposer en ces termes.

---

## Annexe D - Réalisation d'une base de données

Voyons un exemple de création d'une base de données avec GBDI. Afin de trancher quelque peu avec le diagnostic, nous avons choisi une application radicalement différente : nous désirons créer une base relative au choix d'une peinture ou d'un revêtement, pour un mur donné. Nous disposons donc des objets de base "peinture jaune", "peinture rouge", ..., "chaux", "crépi". Parmi tous ces objets, nous distinguerons la classe des peintures proprement dites et celles des revêtements. La réunion des objets de ces deux classes forme la classe des protections. Protection, peinture et revêtement sont donc trois classes (non-disjointes) que l'on peut considérer comme des instances d'une méta-classe. Appelons #Protection# cette méta-classe.

De manière générale, il paraît commode de nommer une méta-classe selon la *classe englobante*. Par convention, les méta-classes sont nommées #xxx# et la classe englobante porte le même nom XXX écrit en majuscule.

Les rapports entre méta-classe et classes peuvent être considérés comme différents des rapport entre classes et objets. Nous utiliserons donc des liens distincts. Une classe peut également être vue comme une propriété des objets qu'elle contient; alors qu'une méta-classe peut être perçue comme le type des objets associés.

Afin d'accentuer le coté hiérarchique de cette représentation, nous utiliserons l'objet racine (#system-obj#) pour accéder à la liste des méta-classes (en quelque sorte, la racine est considérée comme la seule méta-méta-classe connue du système).

Nous avons pris les conventions suivantes : le lien #s1 unira la racine à toutes les méta-classes ; le lien #s2 unira une méta-classe à toutes les classes qui en dépendent (dont une classe englobante écrite en majuscule) ; le lien *Sorte-de* unira les classes aux objets de base.

Ces conventions nous ont paru commodes, mais ne sont nullement indispensables. Le formalisme de propagation de propriétés se contente de la distinction objets-classes, sans nul besoin de classes englobantes ou de méta-classes. Ces derniers éléments visent plus à faciliter la gestion de la base et à permettre des accès rapides à des ensembles d'objets. Il faut utiliser le logiciel GBDI pour être sensible aux commodités que cela apporte. En tout état de cause, les méta-classes ne peuvent en aucun cas apparaître dans les formules F<sub>ppi</sub> et ne sont donc pas pourvues d'une sémantique opérationnelle.

Dans le cadre de notre exemple, nous aurons à établir les relations suivantes :

P.rouge <---Sorte-de---> Peinture (regroupement des objets en classes)

P.rouge <---Sorte-de---> PROTECTION

...

crépi <---Sorte-de---> Revêtement

crépi <---Sorte-de---> PROTECTION

PROTECTION <---#s2---> #Protection# (regroupement des classes)

Peinture <---#s2---> #Protection#

Revêtement <---#s2---> #Protection#

#Protection# <---#s1---> #system-obj# (liaison avec la racine)

On obtient donc globalement une quasi-arborescence:

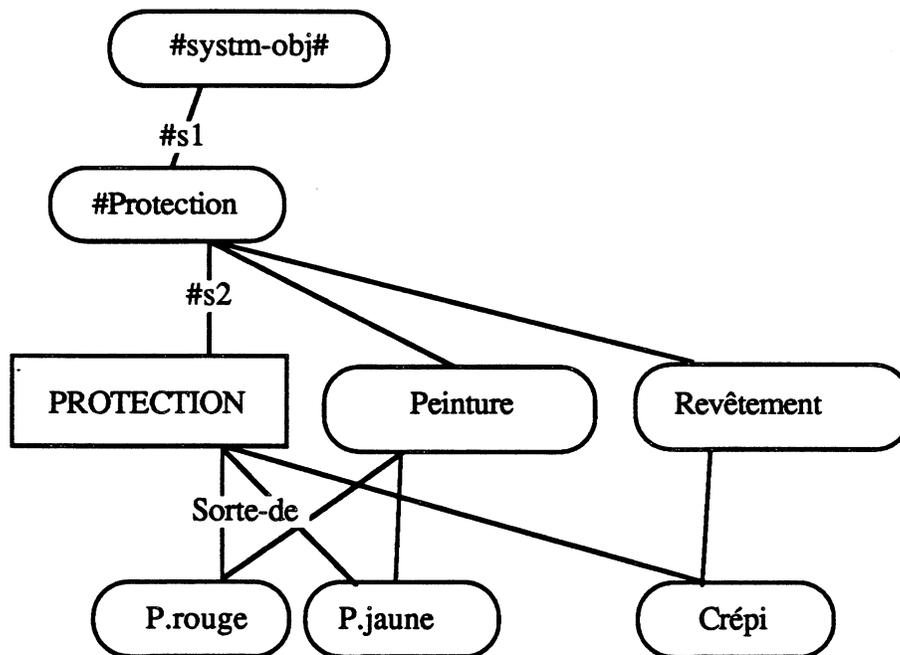


figure G.1 : taxonomie

Il est également possible de créer d'autres classes codant, non plus la nature des objets de base, mais d'autres propriétés. Nous pouvons créer, par exemple, une classe `<revêtement-peu-couteux>`, une classe `<revêtement-blanc>` qui contiendra la peinture blanche et le plâtre, etc. Ces classes sont des représentations des propriétés unaires  $SP_i$  susceptibles d'être utilisées dans les formules  $F_{ppi}$ .

Ayant constitué, de cette façon, toutes les taxonomies internes au problème, on doit se pencher sur les relations entre les objets de base. Ces relations vont représenter les propriétés statiques binaires  $BP_i$  également utilisées dans les formules  $F_{ppi}$ .

Nous avons défini un formalisme (de type grammaire hors-contexte) décrivant un mécanisme dit de "propagation de propriétés" qui ne tient compte que de relations inter-objets. Il ne contient pas, entre autres, de relations entre classes, voire entre méta-classes.

Reprenons l'exemple ci-dessus et supposons que nous disposions également d'une taxonomie décrivant la nature du mur (plâtre, brique, ciment, etc). Définissons deux liens "adapté-à" et "déconseillé-sur". Nous pourrions alors définir les relations :

P.rouge <---adapté-à---> plâtre  
 P.jaune <---adapté-à---> plâtre  
 crépi <---adapté-à---> brique  
 ...  
 crépi <---déconseillé-sur---> pierre-meulière

A la fin de ce processus nous disposerons donc d'un réseau assez fortement structuré qui peut se visualiser par le schéma suivant :

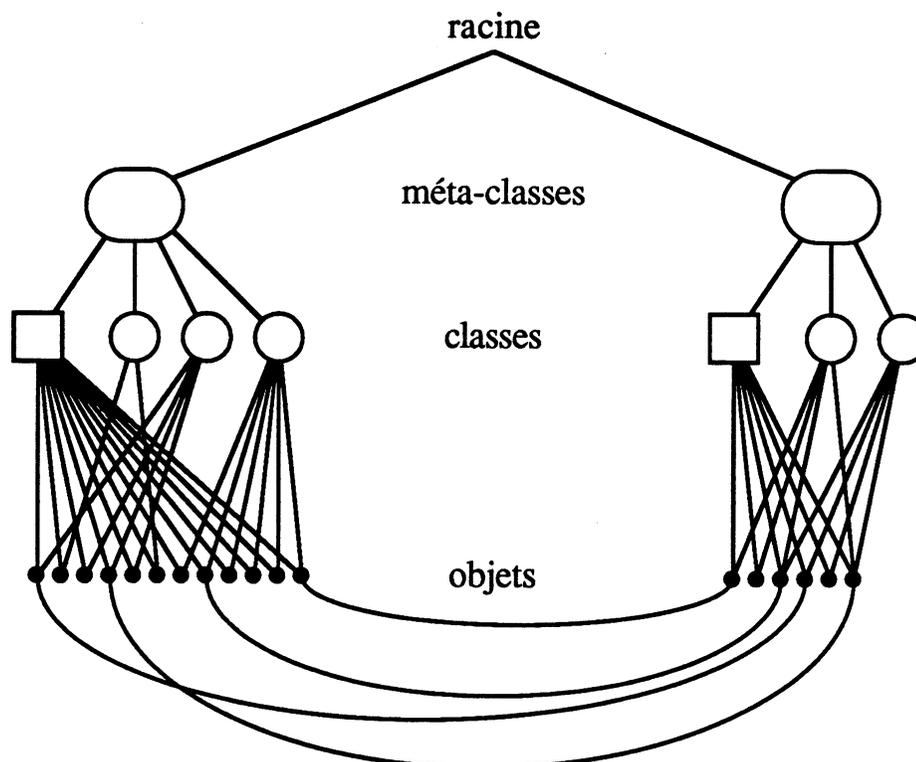


figure G.2 : liaisons entre taxonomies

En principe, tout objet appartient à une ou plusieurs classes, dont une classe englobante. Toutes ces classes doivent appartenir à la même méta-classe (le type de l'objet). Les relations "horizontales" ne concernent que des objets élémentaires.

Il nous reste à placer dans ce réseau les classes qui vont représenter les propriétés initiales et propagées. Ces classes sont identiques aux classes  $e\_SP_i$ , à cela près que l'on ne connaît pas, a priori, les objets qu'elles contiennent. On va donc se contenter de créer les classes correspondantes et de les rattacher à une méta-classe. De cette façon on définit implicitement les seuls objets qui peuvent entrer dans ces classes (les petits-fils de la méta-classe correspondante).

Ainsi, la classe <en-stock> sera rattachée à la méta-classe #Protection# et ne pourra donc contenir que des instances de la classe PROTECTION. Le contenu effectif de cette classe sera du ressort de l'utilisateur (caractéristiques des propriétés initiales). De même, la classe <usage-possible> sera rattachée à la même méta-classe mais ses instances seront déterminées par le système, au travers de l'évaluation des formules  $F_{ppi}$  (caractéristiques des propriétés propagées). Quoi qu'il en soit, lors de la création de la base, ces classes formeront des branches interrompues dans les diverses taxonomies.

Afin de faciliter l'accès à ces classes dynamiques, elle peuvent être regroupées, à titre exceptionnel, dans une seconde méta-classe. Là encore, cette entorse à la simplicité du formalisme ne vise qu'à offrir des facilités d'accès pour la lecture des résultats (comparaison, par l'utilisateur, des classe initiales aux classes propagées correspondantes). A partir de la racine, on pourra accéder à toutes les classes dynamiques sans être obligé de parcourir successivement toutes les méta-classes. De même, on pourrait créer une méta-classe des classes englobantes, etc. On obtient donc finalement le réseau suivant :

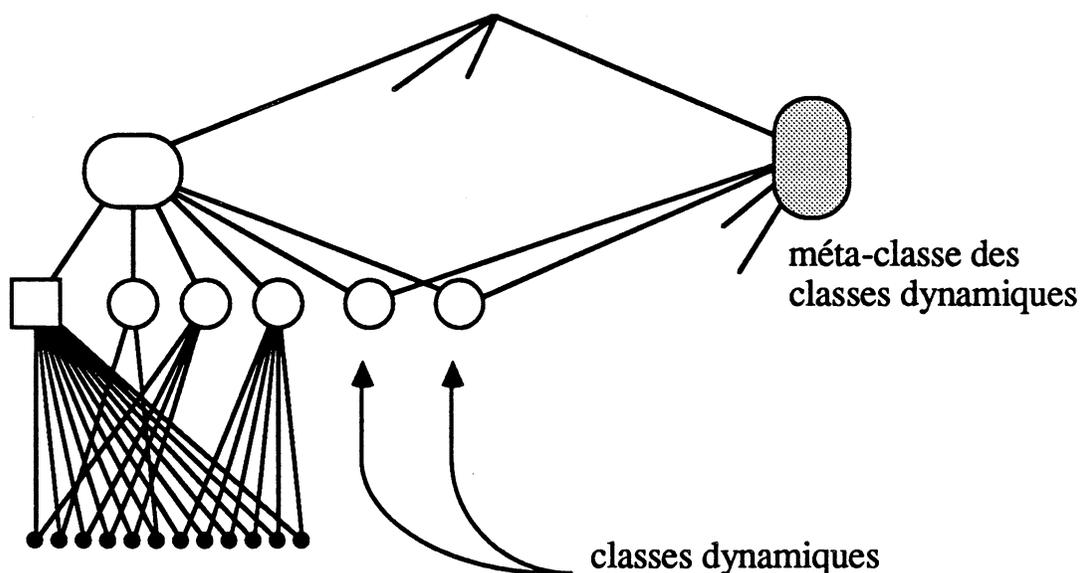


figure G.3 : classes dynamiques

---

## Glossaire des termes médicaux

De nombreuses définitions sont extraites du dictionnaire des termes techniques de médecine (Garnier & Delamare, Ed Maloine 1969).

**Anamnèse** : ensemble des informations initiales relatives à un malade. Cela comprend les habitudes de vie, le régime alimentaire, le sexe, l'âge, l'historique médical, etc. Le résultat de cet interrogatoire est fondamental pour réduire le champ d'investigation. Par exemple, un voyage récent en pays tropical orientera vers les maladies exotiques ; l'absence de toute sortie hors de France fera exclure *a priori* ces dernières. Cette exclusion n'est toutefois jamais catégorique.

**Examens cliniques, para-cliniques** : les examens cliniques regroupent l'interrogatoire et toute l'observation immédiate "au lit" du malade (pouls, respiration, palpations diverses, etc). Les examens para-cliniques nécessitent un matériel de mesure (radiographie, scanner, RMN, etc) et demandent un délai. Il est donc préférable d'exploiter au maximum les informations cliniques avant toute prescription d'examen complémentaire (voir à ce propos [Rialle 87] pour la classification des examens en cliniques immédiats, cliniques approfondis, para-cliniques simples et para-cliniques complexes).

**Faisceau d'arguments** : ensemble d'indices allant dans le même sens (i.e. évoquant un tableau symptomatologique cohérent). La découverte d'un symptôme pathognomonique (spécifique d'une unique maladie) permet d'éviter le recours au faisceau d'arguments. Ils sont malheureusement très rares.

**Invasivité** : un examen est dit non-invasif s'il n'impose aucune pénétration à l'intérieur de l'organisme observé. L'échographie ultra-sonore est un exemple d'examen non-invasif. Tous les examens nécessitant des prélèvements (biopsies, etc) sont dits invasifs. Tous les examens invasifs font courir un risque supplémentaire d'infection au malade et sont donc peu recommandés chez un immunodéprimé.

**Nosologie** : Etudes des caractères distinctifs qui permettent de définir les maladies. **Groupe nosologique** : Ensemble des maladies, ou des formes d'une même maladie, provoquées par le même agent pathogène ou par

des agent pathogènes très voisins les uns des autres. Exemple : groupe malarien, rickettsioses, fièvres récurrentes, etc.

**Symptôme** : Phénomène particulier que provoque dans l'organisme l'état de maladie.

**Syndrome** : réunion d'un groupe de symptômes qui se reproduisent en même temps dans un certain nombre de maladies. Par exemple, l'ictère (la jaunisse) est un syndrome qui peut être dû à une hépatite virale, une obstruction des canaux du foie, une dysfonction de la rate, etc.

**Traitement d'épreuve** : en cas d'urgence, une thérapeutique peut être prescrite sans que soit établie formellement la nature de la maladie. Si l'état général du patient s'améliore, on considérera que c'est un indice supplémentaire en faveur de l'hypothèse initiale. De telles prescriptions sont également pratiquées si aucun examen complémentaire n'a fourni de résultat probant.

**Traitement symptomatique, traitement syndromique** : actions thérapeutiques visant à faire disparaître le symptôme (resp. le syndrome) sans pour autant guérir la maladie. De tels traitements visent à rétablir le confort du malade en évitant toutefois de trop perturber l'organisme. Ces traitements ne doivent pas invalider les observations ultérieures (par exemple, pas de traitement antibiotique avant une recherche de germes).

---

## Bibliographie:

[Bernard 65] C. Bernard : Introduction à l'étude de la médecine expérimentale  
... 1865

[Bonnet 84] A. Bonnet : l'IA promesses et réalités  
Editions inter-éditions 1984

[Buisson 85] J.C Buisson & H. Farreny & H. Prade : un système expert en  
diabétologie accessible par minitel  
5° Journées internationales sur les S.E., Avignon 1985

[Broussalian 88] E. Broussalian : de l'intérêt de l'informatique dans la  
prescription homéopathique  
Thèse de doctorat en médecine, Université Joseph Fourier 1988

[Bylander 85] T. Bylander & S. Mittal & B. Chandrasekaran : CSRL : a  
language for expert system for diagnosis  
International Joint Conference on Artificial Intelligence, Los-angeles 1985

[Cadoz 87] C. Cadoz : Timbre et causalité  
Rapport interne LIFIA-ACROE, Grenoble 1987

[CSMA 88] C. de Sainte Marie : Apprentissage et autonomie ; fragments  
théoriques  
Thèse de docteur-ingénieur LIFIA-INPG, Grenoble 1988

[David 84] J.M. David & Y. Descotte : cost-efficient troubleshooting  
European Conference on Artificial Intelligence, 1984

[David 86] J.M. David & J.P. Krivine : reasoning from structure and behavior,  
four relevance criteria  
European Conference on Artificial Intelligence, 1986

[David 87] J.M. David & J.P. Krivine : three artificial intelligence issues in fault  
diagnosis  
2° European Workshop on Fault diagnosis, Manchester 1987

[Davis 82] R. Davis : diagnosis based on description of structure and function  
AAAI, Pittsburg 1982

- 
- [Davis 84] R. Davis : diagnostic reasoning based on structure and behaviour  
AI journal n° 24 1984
- [Dempster 67] A.P. Dempster : upper and lower probabilities induced by a multivalued mapping  
Annals of mathematical statistics n° 38 1967
- [De Kleer 76] J. De Kleer : local methods of localizing faults in electronic circuit  
MIT Artificial Intelligence laboratory 1976
- [De Kleer 79] J. De Kleer : causal and teleological reasoning in circuit recognition  
MIT Artificial Intelligence laboratory 1979
- [De Kleer 86] J. De Kleer : an assumption-based truth maintenance system  
AI journal n° 28 1986
- [De Kleer 86] J. De Kleer : problem solving with the ATMS  
AI journal n° 28 1986
- [De Kleer 86] J. De Kleer : reasoning about multiple faults  
AAAI, Philadelphia 1986
- [De Kleer 87] J. De Kleer & B.C. Williams : diagnosing multiple faults  
AI journal n°32 1987
- [Doyle 84] J. Doyle : a truth maintenance system  
AI journal n° 24 1984
- [Dreyfus 72] H. Dreyfus : Intelligence artificielle mythes et limites  
Editions Flammarion 1972
- [Fieschi 84] M. Fieschi : intelligence artificielle en médecine  
Collection Masson, "Méthode + Programme", 1984
- [Gallanti 86] M. Gallanti, L. Gilardoni, G. Guida & A. Stefanini : exploiting physical and design knowledge in the diagnosis of complex industrial systems  
European Conference on Artificial Intelligence, 1986
- [Ganascia 84] J.G Ganascia : reasoning and results in expert systems : main difference between diagnostic systems and problem solvers  
European Conference on Artificial Intelligence, 1984
- [Garvey 76] T.D. Garvey : perceptual strategies for purposive vision  
SRI AI center, Technical note n° 117 Menlo Park Ca. 1976

---

[Gascuel 84] O. Gascuel, J. Charlet, S. Genetet & B. Mari : le système LEZARD, un travail sur le raisonnement incertain en médecine  
4° Journées internationales sur les S.E., Avignon 1984

[Geffner 87] H. Geffner & J. Pearl : An improve constraint-propagation algorithm for diagnosis.  
International Joint Conference on Artificial Intelligence, Milan 1987

[Genes 82] M. Genesereth : diagnosis using hierarchical design models  
AAAI, Pittsburgh 1982

[Genes 84] M. Genesereth : the use of design description in automated diagnosis  
AI journal n° 24 1984

[Hamscher 83] W. Hamscher : using structural and functional information in diagnostic design  
AAAI, Washington 1983

[Hayes 83] F. Hayes-Roth, D.A. Waterman & D.B. Lenat : building experts systems  
Ed. Addison Wesley 1983

[Kuipers 84] B. Kuipers : commonsense reasoning about causality : deriving behavior from structure  
AI journal n° 24 1984

[Largeault 85] A.F. Largeault : la simulation du raisonnement médical  
la recherche n° 170 1985

[Lefevre 88] J.M. Lefevre : raisonnement sur la qualité de l'information  
Thèse de docteur-ingénieur LIFIA-INPG, Grenoble 1988

[Lux 85] A. Lux : algorithmique et contrôle en vision par ordinateur  
Thèse d'état INPG-USMG, Grenoble 1985

[Marz 88] M. Marzouki & B. Courtois : PESTICIDE a prolog-written expert system as a tool for integrated circuit debugging.  
Rapport IMAG RR 691-I- Janvier 1988

[Milne 85] R.W Milne : the theory of responsibilities  
Sigart newsletter n° 93 1985

---

[Milne 86] R.W. Milne & B. Chandrasekaran : fault diagnosis and expert systems

6° Journées internationales sur les S.E., Avignon 1986

[Peng 86] Y. Peng & J.A. Reggia : plausibility of diagnosis hypotheses : the nature of simplicity

AAAI, Philadelphia 1986

[Piette 86] D. Piette, C. Roche & J.C. Ianeselli : ALPA diagnosis expert system for supervision of nuclear reactor

European Conference on Artificial Intelligence, 1986

[Pople 75] H.A. Pople, J.D. Myers & R.A. Miller : DIALOG : a model of diagnosis logic for internal medicine

International Joint Conference on Artificial Intelligence, Tbilissi 1975

[Quinlan 83] J.R. Quinlan : learning efficient classification procedures and their application to chess end games

Machine learning 1983

[Quinlan 86] J.R. Quinlan : Induction of decision trees

Machine learning 1.1 1986

[Quinlan 87] J.R. Quinlan : generating production rules from decision trees

International Joint Conference on Artificial Intelligence, 1987

[Qinqueton 87] J. Quinqueton & J. Sallantin : système d'apprentissage contrôlé par réfutation

Journées Française de l'Apprentissage, Chamrousse 1987.

[Raoult 85] O. Raoult : diagnostic de pannes des systèmes complexes

rapport de DEA LIFIA-INPG, Grenoble 1985

[Raoult 87] O. Raoult : apprentissage formel, application au diagnostic

Journées Française de l'Apprentissage, Chamrousse 1987.

[Raoult 88] O. Raoult : diagnostic de panne des systèmes complexes, modélisation des stratégies expertes, IPLSYS

8° Journées internationales sur les S.E., Avignon 1988

[Rapp. 84] A. Rappaport & J.M. Chauvet : symbolic knowledge processing for acquisition of expert behavior : a study in medicine

CMU-RIT-TR-84-8 1984

---

[Reggia 83] J.A. Reggia, J. Nau, D. Wang : diagnostic expert system based on a set covering model  
Int. J. Man-machine Studies 1983

[Reggia 85] J.A. Reggia : abductive inference  
IEEE, 1985

[Reiter 85] R. Reiter : a theory of diagnosis from first principles  
University of Toronto 1985, réédition AI journal n°32, 1987

[Rialle 87] V. Rialle : aide au diagnostic et à l'apprentissage dans un domaine médical incertain, incomplet et évolutif  
Thèse de l'USMG, Grenoble 1987

[Robert 88] C. Robert : Observable Network, a Markov field approach for probabilistic reasoning  
Soumis à IEEE 1988

[Short. 76] E.H. Shortliffe : computer-based medical consultations : MYCIN  
American Elsevier New-York 1976

[Szolovits 82] P. Szolovits : artificial intelligence in medicine  
Westview Press 1982

[Utgoff 86] E. Utgoff : shift of bias for inductive concept learning  
Machine Learning vol II, 1986

[Velde 85] W. van de Velde : naive causal reasoning for diagnosis  
5° Journées internationales sur les S.E., Avignon 1985

[Velde 88] W. van de Velde : learning from experience  
Vrije universiteit Brussel, technical report 88-1

[Weiss 78] C. Weiss, C.A. Kulikowski & A. Safir : glaucoma consultation by computer  
Comput. Biol. Med. Vol n° 8 1978

[Wright 83] J.M. Wright & M.S. Fox : SRL 1.5 user manual  
I.S.laboratory, CMU, 1° december 1983

[Yen 86] J. Yen : a reasoning model based on an extended Dempster-Shafer theory  
AAAI 86



## Résumé

Nous analysons le noyau d'un système de diagnostic. Notre analyse se base sur l'étude d'applications concrètes, et nous permet de dégager trois entités caractéristiques du problème de diagnostic : le langage d'observation, la référence de bon fonctionnement et le langage de qualification. Le langage d'observation détermine l'information perceptible ; la référence de bon fonctionnement introduit le concept de panne et le langage de qualification définit l'ensemble des réponses possibles du système.

L'explicitation de ces trois entités nous amène à contester certaines affirmations courantes sur le pouvoir de représentation des modèles dits "profonds" ou "associatifs". Elle nous permet aussi d'en affirmer la généralité et l'intérêt, sur des bases bien comprises.

Dans la deuxième partie, nous étudions plus particulièrement l'aide au diagnostic médical. Notre approche vise à séparer la modélisation du raisonnement utilisé par les experts de la modélisation des connaissances factuelles employées. Pour ce faire, nous avons défini un formalisme de "propagation de propriétés" dans lequel (1) les objets, (2) les propriétés que possèdent ces objets, (3) les usages qui sont fait de ces propriétés sont représentés séparément. Ce formalisme a fait l'objet d'une réalisation concernant l'aide à la prescription et à l'interprétation d'exams complémentaires.

## Abstract

We analyse the core of a diagnosis system. Our analysis is based on studies of current applications, and enables us to define three entities which characterize the diagnosis problem: The observation language, the correct function reference, and the qualification language. The observation language defines the observable data; the correct function reference defines the concept of fault and the qualification language defines the set of the possible system outputs.

Pointing out these three entities enables us to contest some affirmations about the representation power of "deep" or "shallow" models. We can clarify the generality and interest of these approaches.

In the second part, we focus on the medical diagnosis assistance. Our purpose is to separate the modelling of reasoning from the modelling of knowledge which is used during this reasoning. Therefore, we define a formalism called "property propagation" which explicitly separates (1) objects, (2) object properties, and (3) uses of these properties. This formalism has been implemented within the aid of prescription and interpretation of complementary exams.

**Mots-clés : diagnostic, système expert, aide à la décision, diagnostic médical, modèles profonds, modèles associatifs.**

