



HAL
open science

Structuration des connaissances et raisonnement a l'aide d'objets

Christine Pierret-Golbreich

► **To cite this version:**

Christine Pierret-Golbreich. Structuration des connaissances et raisonnement a l'aide d'objets. RR-0847, INRIA. 1988. inria-00075706

HAL Id: inria-00075706

<https://inria.hal.science/inria-00075706>

Submitted on 24 May 2006

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.

INRIA

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt

BP105
78153 Le Chesnay Cedex
France

Tél. (1) 39 63 55 11

Rapports de Recherche

N° 847

STRUCTURATION DES CONNAISSANCES ET RAISONNEMENT A L'AIDE D'OBJETS

Christine PIERRET

MAI 1988



* RR - 8847 *



STRUCTURATION DES CONNAISSANCES ET RAISONNEMENT A L'AIDE D'OBJETS

ORGANIZING KNOWLEDGE AND REASONING WITH OBJECTS

Christine PIERRET
INRIA
Domaine de Voluceau
BP 105 - Rocquencourt
78153 Le Chesnay Cedex - France

Résumé

Le style "orienté objets" est une lame de fond qui déborde des applications usuelles de l'Intelligence Artificielle. Cet article concerne la structuration d'une base de connaissances dans un formalisme purement objet, pour un système d'aide à la modélisation en biologie (EDORA).

La base a été construite à l'aide d'unités interdépendantes, les schémas, sans faire appel à aucun autre type de représentation car les mécanismes propres au système de gestion de bases de connaissances suffisent (SHIRKA).

Cet article décrit comment la structuration de la base en treillis distincts contrôlés par un treillis de plus haut niveau est le support d'un mécanisme de raisonnement original, conduit par un algorithme de "multiclassification". Il illustre combien la conceptualisation d'un univers en objets est un art difficile mais permet d'apporter des solutions intéressantes.

Abstract

This paper deals with the organization of a knowledge base for a computer aided system in biological modeling, based on a pure object formalism. The base was designed using interdependent units called "schemes" without any other type of description unit. In particular, no rules were needed because the knowledge bases management system has its own inference mechanisms. The base has been organized around several objects families supervised by a higher level network. This structure is the support of a novel reasoning mechanism managed by a "multiclassification" algorithm. Thus, interesting solutions for the model formulation has been carried out.

This article shows that the universe conceptualization with objects is a difficult art. However, using a pure objects "model", brings benefits both to the expert and the computer scientist.

STRUCTURATION DES CONNAISSANCES ET CONDUITE DU RAISONNEMENT A L'AIDE D'OBJETS

L'aide à la formulation d'un modèle en dynamique des populations.

Christine PIERRET
INRIA-BP105
78153 LE CHESNAY CEDEX FRANCE

Ce travail s'inscrit dans le cadre de la structuration d'une base de connaissances pour un système d'aide à la modélisation en Biologie et en Ecologie (projet EDORA). Le point plus particulièrement abordé ici concerne l'étape de formulation du modèle. L'objectif est d'organiser la base de connaissance dans un formalisme de représentation centrée-objet en vue de fournir une aide à la formulation du modèle (choix ou construction). On se limite pour l'instant au domaine de la dynamique des populations.

La première partie est consacrée à l'analyse des connaissances et du raisonnement utilisés par le biologiste pour modéliser le système biologique étudié. Différentes questions sont abordées: nature des connaissances, nature du raisonnement. Ces questions sont à la frontière de différentes disciplines: psychologie cognitive, intelligence artificielle et bien sûr modélisation. On pourra donc se référer à des publications spécialisées mentionnées dans le texte pour chacune d'elles.

Le premier point de cette première partie concerne les questions de *niveau* et de *nature des connaissances*, la notion de *point de vue* de connaissance. L'importance de différents concepts, est mise en évidence: connaissances profondes et superficielles, connaissances quantitatives, qualitatives. On distingue ainsi deux *niveaux* de connaissances. Les connaissances profondes qui concernent les mécanismes internes du système. Elles se rapportent à l'aspect explicatif de son comportement. De manière générale on peut concevoir différentes alternatives pour formaliser ces connaissances: modèles mathématiques et de simulation d'un processus complexe, lois physiques gouvernant une situation, modèles fonctionnels [KLEER et BROWN, 1981]. Deux formalismes sont retenus ici l'un de type mathématique (modèles mathématiques) et l'autre de type biologique (formalisme dit "pseudo-chimique" permettant de représenter les processus biologiques impliqués dans le système). Par contre, les connaissances superficielles se réfèrent aux observations sur le système et concernent donc l'aspect phénoménologique. Ces notions de connaissances profondes et superficielles concernent un domaine actif récemment abordé en Intelligence Artificielle. On se rapportera par exemple à [HART, 1981] [STEELS, 1987], [CHANDRESEKERAN, 1983]. D'autre part, dans les deux cas, la *nature* des connaissances permet de discerner par ailleurs deux types de connaissance: connaissances qualitatives et quantitatives. Cet aspect peut être relié d'une certaine manière à un autre domaine celui de la Physique Qualitative, domaine récemment ouvert en Intelligence Artificielle, par J. DE KLEER et J.S. BROWN, K. FORBUS, ou B. KUIPERS. Enfin, on soulignera le fait que la description d'un objet selon différents points de vue, mathématique et biologique, est un élément important pour le raisonnement. C'est le type de démarche qui est développé par Zeigler [ZEIGLER, 1986]. Celui-ci propose une théorie de modélisation basée sur les différents aspects d'un objet. Ainsi dans "multifaceted modelling", il distingue par exemple, l'aspect électrique, thermique, etc... pour les composants électriques.

Le dernier point traité dans cette première partie concerne l'analyse du raisonnement. Le caractère abductif du raisonnement le rapproche du raisonnement de diagnostic. Il est montré comment le choix d'un modèle correspond à une démarche de classification au sein d'un catalogue de schémas. On retrouve là des concepts plus particulièrement étudiés en psychologie. La classification est étudiée par Clancey [CLANCEY, 85], la notion de schémas introduite par Minsky [MINSKY, 75] qui formulait déjà l'idée suivante: "... When one encounters a new situation (or makes a substantial change in one's view of the present problem), one selects from memory a structure called a *frame*". Enfin, le problème de la conception d'un modèle, dans le cas de situations complexes est abordé.



17.05.88

La deuxième partie est consacrée à la représentation à l'aide d'objets de ces différentes connaissances. La base de connaissance est structurée en familles d'objets. On définit ainsi quatre familles d'objets: modèle, processus, courbe, situation. Une organisation hiérarchique est proposée pour chacun d'eux. Il est montré comment un algorithme de classification permet de mener le raisonnement conduisant au choix d'un objet dans chacune de ces hiérarchies.

Enfin la dernière partie montre comment la représentation centrée-objet apporte une solution au problème de la cohérence des différentes connaissances. Celle-ci est obtenue par la définition d'un objet de "plus haut niveau", l'objet système. Chaque sous-classe de *système* est un "monde de connaissances cohérentes". Un exemple particulier celui du système logistique éclairera ce point.

On conclura en montrant comment cette structuration de la base de connaissances en treillis distincts contrôlés par un treillis de "plus haut niveau" permet d'apporter une aide à la formulation (choix ou construction) d'un modèle en permettant d'inférer à partir des connaissances superficielles (qualitatives et quantitatives) les connaissances profondes recherchées, c'est-à-dire un modèle ou un processus satisfaisant pour le système étudié.

1 - ANALYSE DES CONNAISSANCES

L'analyse des connaissances met en évidence que la modélisation en biologie fait intervenir deux niveaux de connaissances, niveaux profond et superficiel.

1.1. *Connaissances profondes, superficielles*

Ces questions sont exposées par exemple dans [HART, 81], [STEELS, 87], [DAVIS, 83]. Hart distingue en fonction de leur complexité conceptuelle deux types de systèmes: " By surface systems I mean those having no underlying representation of such fundamental concepts as causality, intent, or basic physical principles; deep systems, by contrast, attempt to represent concepts at this level...".

Hart illustre ces notions par la présentation d'un exemple du domaine pétrolier. Il présente ainsi trois niveaux de connaissances concernant la modélisation d'un réservoir de pétrole. Du plus superficiel au plus profond, on trouve:

Niveau1:

"IF:
The reservoir is volumetric
AND
There is no gas cap
AND
Gas is observed coming out of solution

THEN:
Pressure is less than the bubble point"

Bien que ce type de règles soit importante pour l'utilisateur, elle ne représente aucune notion concernant la causalité physique.

Niveau2:

"The second model can be thought as a form of lumped parameter model. It represents the reservoir as a sealed tank, with all movements of fluids occurring only via an oil well. Algebraically, the model balances the volume of fluid removed with an expansion of the rock and fluid remaining. As the remaining rock and fluid remaining expands, the pressure in the reservoir must drop".

Ce modèle capture en un certain sens une partie des notions de causalité physique (si un fluide sort, la matière restante se dilate et donc la pression baisse).

Niveau3:

C'est la représentation du système par un ensemble d'équations aux dérivées partielles et la simulation de ce modèle.

"Simulators embody the deepest "physical" representation of the complex phenomenon of a petroleum reservoir".

Ce problème des relations entre connaissances profondes et superficielles est un domaine de recherche récent et très actif en Intelligence Artificielle [STEELS, 87]. Les recherches dans ce domaine ont des motivations diverses. La nécessité de faire appel aux connaissances profondes est apparue aussi bien pour l'enseignement [CLANCEY, 81], l'explication [SWARTOUT, 81], l'acquisition des connaissances [WIELINGA & BREUKER, 84], le diagnostic [CHANDRASEKARAN, 83], la déduction qualitative du comportement à partir d'un modèle structurel des mécanismes [KUIPERS, 84]. Comme le fait d'ailleurs remarquer B. Kuipers à propos des systèmes experts dans le domaine médical [KUIPERS, 87], on peut distinguer deux générations de systèmes: "the first generation of AI systems in medicine- MYCIN, INTERNIST-1, CASNET- all represent knowledge in fundamentally the same way: as numerical associations between observable manifestations and disease hypothesis....research on a second generation of AI in medicine systems has been attempting to incorporate knowledge of physiological mechanisms and their behavior in health and disease (KUIPERS & KASSIRER 1984; PATIL 1981; PATIL, SZOLOVITS & SCHWARTZ 1981; POPLE 1982)".

Un système biologique est une structure constituée d'entités ou espèces biologiques (populations, organismes, etc..) interconnectées. Ces interactions sont régies par des mécanismes biologiques généralement inconnus ou incomplètement connus.

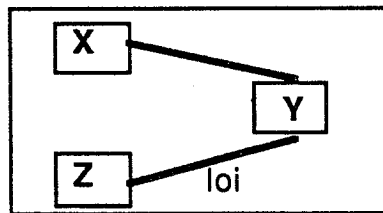


Fig1: Système biologique

□ Espèce biologique
— interaction

Un modèle est l'idéalisation dans un formalisme donné de la structure du système et des mécanismes biologiques qui le régissent. L'objectif de la modélisation en biologie est de construire des modèles pour accéder à la connaissance de ces mécanismes, dans un but de connaissance pure (aspect cognitif), dans un but prédictif (prédire l'évolution du système) ou de contrôle.

Le premier type de connaissances concerne l'aspect phénoménologique. Les observations du biologiste concernent d'une part la courbe de réponse, d'autre part la situation observée. Le biologiste sait décrire la forme de la courbe observée pour ses données. Il sait par exemple qu'elle présente une asymptote, un point d'inflexion situé au milieu de la courbe, etc... Il sait également décrire la situation biologique étudiée. Par exemple il sait que le phénomène étudié concerne la croissance des rats musqués et qu'une telle croissance dépend d'une hormone de croissance, etc... Dans la plupart des cas, il a par ailleurs en mémoire une base de "patterns", catalogues de courbes et de situations stéréotypiques. Il s'agit là de connaissances superficielles. Par contre, la connaissance qu'il n'a pas toujours, et à laquelle il souhaiterait accéder concerne le détail des processus biologiques qui régissent un tel phénomène. La structure et les mécanismes de fonctionnement du système sont des connaissances profondes que le biologiste cherche à identifier à partir de ces connaissances superficielles.

La modélisation d'un système biologique fait donc nécessairement intervenir à la fois les deux niveaux de connaissances mentionnés plus haut: les connaissances

profondes sont à construire et à valider de façon à expliquer les connaissances de surface.

Un exemple du domaine de la dynamique des populations, celui de la modélisation du système prédateur-proie, permettra de présenter plus en détail ces différents niveaux de connaissance. Pour représenter ce système, différents modèles sont proposés. La présentation de ces différents modèles permet d'explicitier trois niveaux principaux de connaissances dont les caractéristiques sont schématiquement les suivantes:

- le niveau 1 est le niveau phénoménologique. Ce type connaissances est souvent décrit sous forme de règles (ou d'un formalisme équivalent).

- le niveau 2 est le niveau fonctionnel. C'est un niveau de connaissance intermédiaire. Pour décrire ces connaissances, il existe des formalismes de description de type intermédiaire. En effet, les modèles utilisent fréquemment des représentations schématiques, "formalismes intermédiaires entre les hypothèses discursives, concernant la structure et/ou le fonctionnement du système, et la représentation mathématique" [PAVE, 88]. L'utilisation de représentations schématiques pour la construction de modèles mathématiques est connue et utilisée dans de nombreux domaines: diagrammes boîtes-flèches, bond-graphs, etc... Ce sont de véritables langages de description. Un algorithme de traduction permet de générer à partir de la représentation graphique un ensemble d'équations ou "contraintes".

On donnera donc pour ce niveau, trois descriptions des modèles :

une description à l'aide de règles.

une description sous forme de schéma graphique.

une description par un ensemble de contraintes

- le niveau 3 est le niveau de représentation des principes biologiques gouvernant le système. Les connaissances seront alors exprimées sous forme d'équations différentielles.

On notera sur l'exemple donné, l'enrichissement sémantique des règles lorsqu'on progresse du niveau 1 au niveau 3. De même, on notera que l'enrichissement progressif des langages de description et de la syntaxe des schémas fonctionnels permet d'obtenir une description qualitative du fonctionnement de plus en plus précise. Enfin, on notera l'évolution de la nature des contraintes, du qualitatif (contraintes ou équations qualitatives) vers le quantitatif (équations différentielles).

Du plus superficiel au plus profond, on trouve:

Niveau 1:

Il correspond à des règles du type:

"SI

il y a prédation d'une espèce P sur une espèce Q

ALORS

la biomasse de P augmente et celle de Q diminue"

Cette règle se situe à un niveau strictement phénoménologique. Elle décrit l'association entre une situation biologique et les manifestations observables. Elle n'exprime aucune connaissance sur la structure et le fonctionnement du système dans la mesure où elle ne représente aucune notion concernant la causalité biologique du phénomène. Un système informatique basé sur ce type de règles ne serait qu'une sorte de base de données de paires "pattern-decision" avec une structure de contrôle permettant d'évoluer dans la base.

Niveau 2:

Pour ce niveau intermédiaire, on présente différents modèles correspondants à des degrés de profondeur de plus en plus grands:

2.1 Loop models [PUCCIA & LEVINS, 85]

La représentation graphique du système prédateur-proie dans ce formalisme est la suivante:

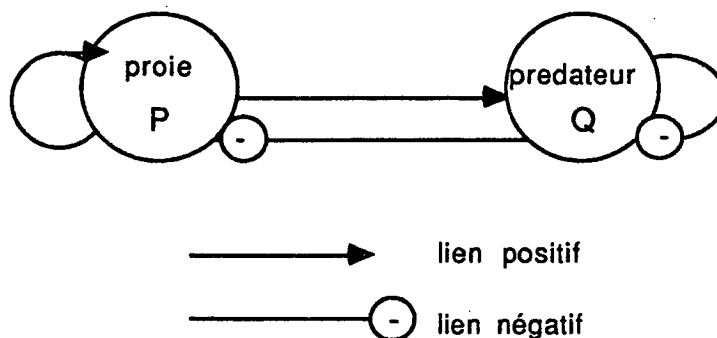


Fig2: Loop model du système prédateur-proie

Ce modèle décrit les variables en présence (P et Q) et les relations d'influence qu'elles entretiennent: effet positif de la proie sur le prédateur, effet négatif du prédateur sur la proie, auto-limitation du prédateur (mortalité naturelle), auto-accroissement de la proie (natalité naturelle).

Ce modèle capture en un certain sens plus de notions sur la causalité biologique puisqu'il indique les causes qui influencent l'évolution des variables: "influence négative du prédateur sur la proie". Ces notions de causalité sont minimales; le fonctionnement du système n'est décrit qu'à travers la description qualitative des dépendances entre variables mais aucun détail n'est donné quant aux mécanismes qui régissent ce fonctionnement.

Cette connaissance serait exprimée sous forme de règles de la manière suivante:

Règle 1:

"Si
il n'y a pas de prédateur

ALORS
la croissance des proies est **auto-favorisée**"

Règle 2:

"Si
il n'y a pas de proies

ALORS
la croissance des prédateurs est **auto-limitée**"

Règle 3:

"Si
la situation observée est une situation de prédation

ALORS
la population prédatrice affecte défavorablement la population des proies
ET
la croissance de la population prédatrice est favorisée par la population des proies"

2.2 System dynamics model.

On prendra pour exemple la représentation graphique utilisée dans STELLA [RICHMOND B., 85]

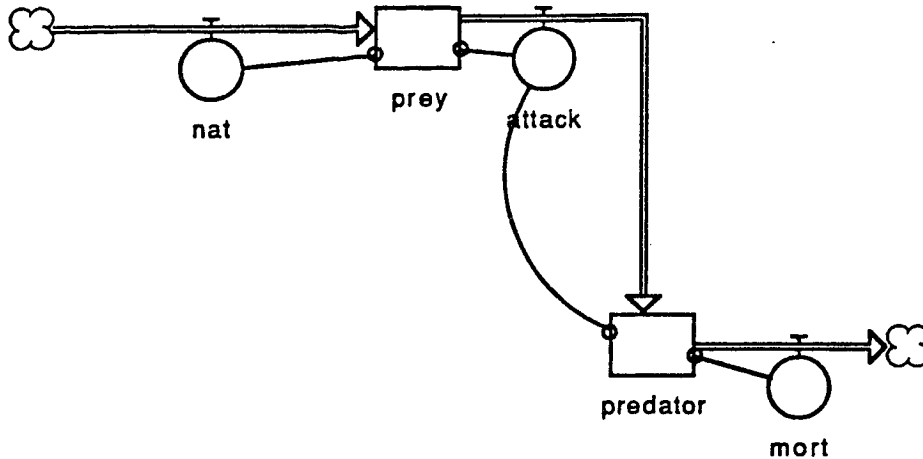


Fig3: modèle du système prédateur-proie dans le formalisme des systèmes dynamiques

Cette représentation graphique est associée de manière isomorphe à l'ensemble d'équations suivant:

- (1) $\text{predateur} = \text{predateur} + \text{attaque} - \text{mort}$
- (2) $\text{proie} = \text{proie} + \text{naissance} - \text{attaque}$
- (3) $\text{attaque} = f_1(\text{prédateur}, \text{proie})$
- (4) $\text{mort} = f_2(\text{predateur})$
- (5) $\text{naissance} = f_3(\text{proie})$

Cette approche véhicule certaines connaissances sur la structure et le fonctionnement du système puisque les deux premières équations expriment des connaissances structurelles: la densité de prédateurs (respectivement de proies) est le résultat d'un bilan arithmétique faisant intervenir différents mécanismes: naissance, mort, attaque. Les trois autres relations (3), (4), (5) indiquent les liens de dépendances fonctionnelles entre les variables puisqu'elle précise pour chaque mécanisme les variables qu'il fait intervenir. Aucune connaissance sur la nature de ces mécanismes n'est donnée. L'équation (4) par exemple, indique que le nombre de naissances des proies est fonction du nombre de proies sans préciser la nature de cette dépendance, (respectivement (4) signifie que le nombre de morts de prédateurs dépend de leur densité et l'équation (5) signifie que nombre d'attaques dépend du nombre de proies et de prédateurs).

L'expression sous forme de règles montre bien l'apparition de plus de précision sur la causalité biologique puisqu'elle précise en partie les mécanismes sous-jacents (attaque, mort, naissance). Les règles correspondantes seraient dans ce cas:

Règle 1:

Si

il n'y a pas de prédateur

ALORS

la croissance des proies est fonction du nombre de proies

16.05.88

Règle 2:

Si

il n'y a pas de proies

ALORS

la croissance des prédateurs est fonction du nombre de prédateurs

Règle3:

"Si

la situation observée est une situation de prédation

ALORS

la population prédatrice affecte défavorablement la population des proies **par attaque directe des proies**

ET

la croissance de la population prédatrice est favorisée par la population des proies **par consommation des proies capturées"**

2.3 Modèle de Physique qualitative [KUIPERS, 86]

Dans le formalisme de la physique qualitative, les relations entre les variables sont représentées à l'aide d'un ensemble de contraintes qualitatives. Trois grands types de contraintes sont disponibles pour décrire la structure et le fonctionnement d'un système: contraintes arithmétiques, fonctions qualitatives, équations différentielles qualitatives.

La structure du système prédateur-proie dans cette approche serait exprimée graphiquement par:

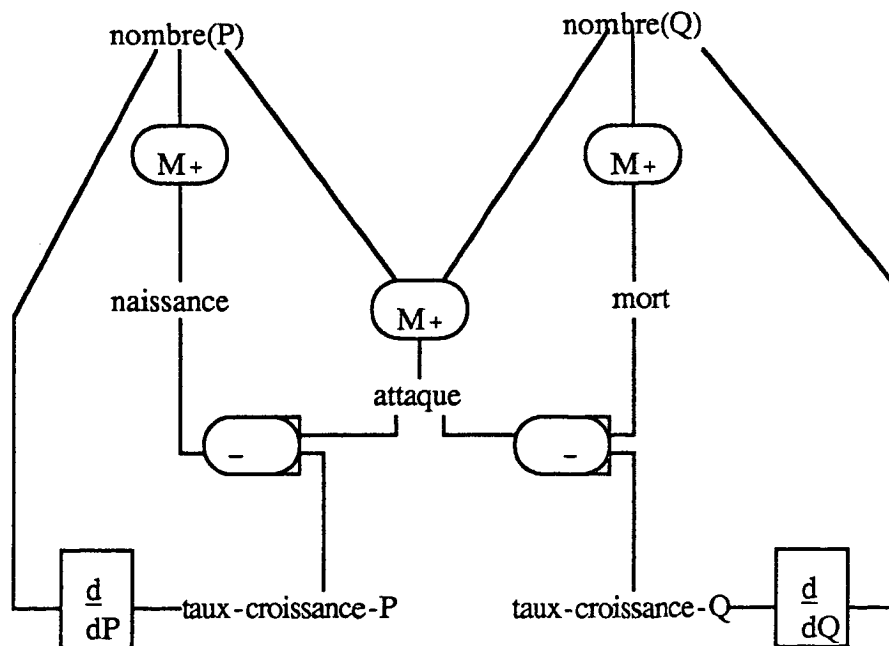


Fig4: modèle du système prédateur-proie dans le formalisme de KUIPERS

Son expression à l'aide de contraintes qualitatives est la suivante:

16.05.88

- (1) Deriv (P, f1)
- (2) Deriv (Q, f2)
- (3) M+(P, naissance)
- (4) M+(Q, mort)
- (5) M+(P, Q, attaque)
- (6) $f_1 = \text{naissance} - \text{attaque}$
- (7) $f_2 = \text{attaque} - \text{mort}$

Cette approche véhicule encore plus de connaissances que précédemment sur la structure et le fonctionnement du système. En effet, on retrouve comme dans l'exemple précédent les notions de causalité biologique: les relations (6) et (7) expriment que que la densité de prédateurs (respectivement de proies est le résultat d'un bilan arithmétique. D'autre part les relations (3), (4), (5) représentent des relations fonctionnelles mais fournissent en plus une connaissance qualitative sur les mécanismes: la contrainte (3) par exemple, indique que le nombre de naissances des proies est une fonction **croissante** de leur densité, (respectivement (4) signifie que le nombre de mrts de prédateurs est une fonction **croissante** de leur densité et (5) signifie que nombre d'attaques est une fonction **croissante** des densité de proies et de prédateurs).

Ces connaissances exprimées sous forme de règles s'énonceraient de la manière suivante:

Règle 1:

Si

il n'y a pas de prédateur

ALORS

la croissance des proies est une fonction **croissante** du nombre de proies

Règle 2:

Si

il n'y a pas de proies

ALORS

la croissance des prédateurs est une fonction **décroissante** du nombre de prédateurs

Règle3:

"Si

la situation observée est une situation de prédation

ALORS

la population prédatrice affecte défavorablement la population des proies **par attaque directe des proies**

ET

la croissance de la population prédatrice est favorisée par la population des proies **par consommation des proies capturées"**

Les connaissances fournies par ce modèle sur la structure et les mécanismes régissant le système sont les plus complètes. Se trouvent réunies ici à la fois des connaissances de type 2.1 sur la nature de l'influence réciproque des variables et des connaissances de type 2.2 sur les mécanismes qui en sont sous-jacents.

En conclusion, ce niveau concerne la description de la structure et des aspects fonctionnels du système. Cette description est toutefois incomplète puisqu'elle se limite à la description des variables concernées et d'un ensemble de "contraintes" décrivant partiellement les relations qu'entretiennent ces variables sans toutefois expliciter complètement ces contraintes. Dans le modèle 2.3 par exemple, la contrainte (3) n'explicite pas complètement la dépendance entre le nombre de naissances et le nombre de proies; elle ne fournit qu'une connaissance qualitative de son évolution.

Niveau 3:

Ce niveau correspond à l'explicitation précise des principes biologiques et donc des contraintes. La représentation la plus explicite de ces connaissances peut être fournie par un système d'équations différentielles. Ainsi dans les années trente, Lotka et Volterra ont proposé un modèle mathématique décrivant la dynamique d'un système prédateur-proie:

$$\text{(proie)} \quad \frac{dx}{dt} = r*x - a*x*y$$

$$\text{(prédateur)} \quad \frac{dy}{dt} = -m*y + b*x*y$$

Ces équations reposent sur les règles suivantes:

Règle 1:

"SI

il y a absence de prédateur

ALORS

le taux de croissance des proies est **constant**, égal à r (différence entre le taux de mortalité et le taux de mortalité des proies) et la croissance est donc **exponentielle**."

Règle2:

"SI

il y a absence de proie

ALORS

la population prédatrice disparaît **exponentiellement** avec un taux de mortalité m ."

Règle 3:

Si

il y a prédation

ALORS

Le nombre de proies tuées par unité de temps est **proportionnel** au produit des densités des proies et des prédateurs, et l'accroissement de la population de prédateurs dû à la consommation des proies est proportionnel au nombre de proies tuées.

Bien que facile à saisir intuitivement, une telle représentation d'un système véhicule plus de connaissances qu'il ne paraît. **Un système d'équations différentielles a une très grande puissance d'expression et d'inférence**. Les connaissances qui peuvent être déduites d'un tel modèle sont très importantes. Comme l'indique KUIPERS dans la conclusion de son article "Qualitative Simulation" [KUIPERS, 86]:

"... the dynamical approach to qualitative analysis of differential equations has greater expressive and inferential power than local qualitative simulation methods.... the theory of dynamical systems also includes global classification theorems delimiting the possible qualitatively distinct behaviors....".

On fera donc deux remarques à propos de ce niveau de connaissances, l'une concernant la puissance du formalisme différentiel, l'autre ses limitations.

Ainsi, comme on vient de le voir sur le modèle de Lotka-Volterra, l'interprétation biologique de ce modèle et les hypothèses sur lesquelles les équations reposent sont faciles à comprendre (cf règles 1, 2, 3 du niveau 3).

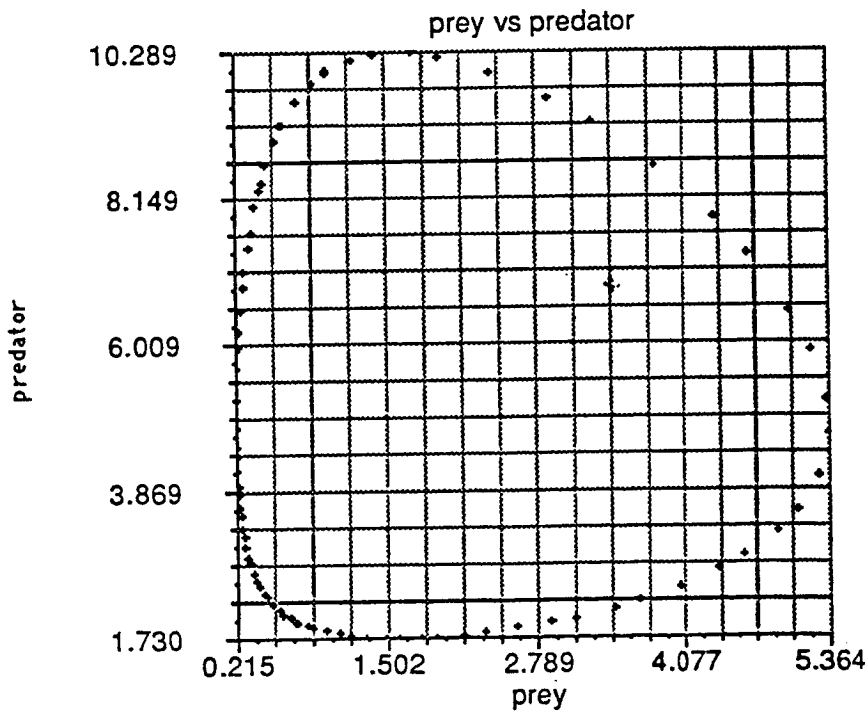
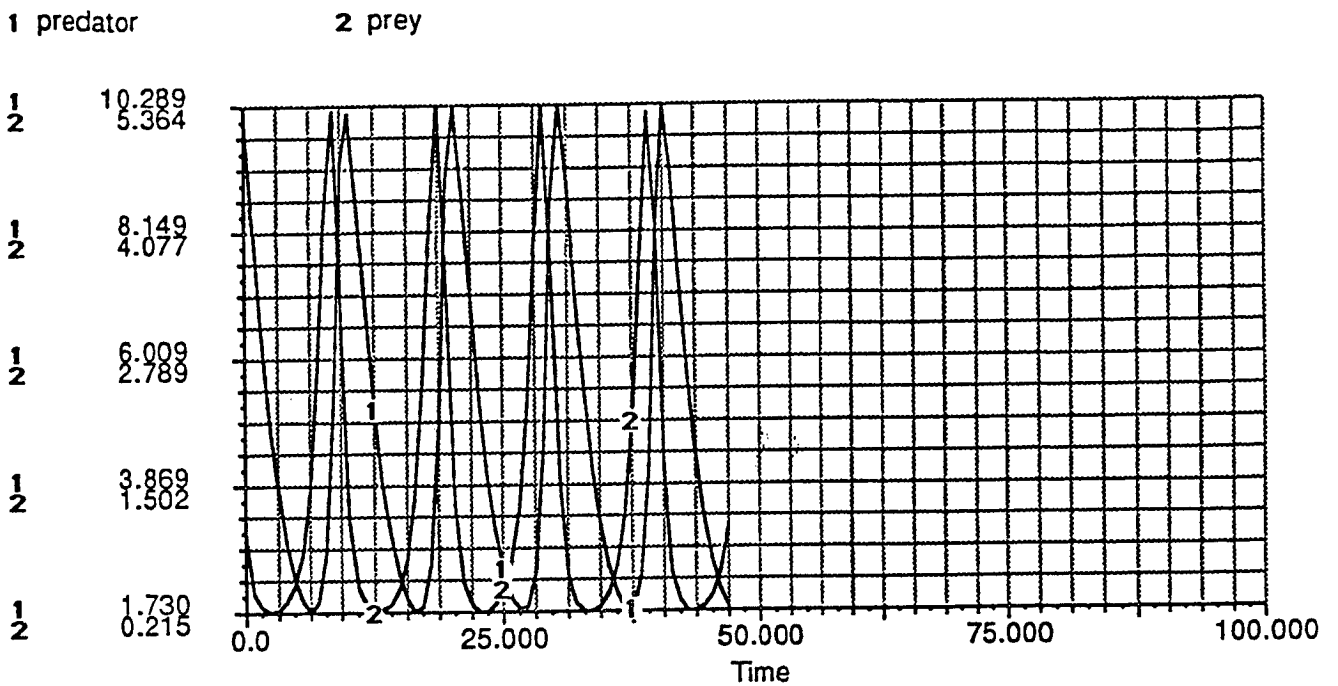
D'autres connaissances, moins évidentes, sont inhérentes au modèle et donc induites à partir de ce modèle. L'étude mathématique de ce modèle montre en effet, par exemple, que ce système est de type conservatif. La raison en est qu'il possède une intégrale première. La fonction:

$H(x,y) = (x-x^* - x^* \log(\frac{x}{x^*})) + \frac{a}{b} (y-y^* - y^* \log(\frac{y}{y^*}))$ est constante sur les trajectoires:

$$\frac{dH}{dt(x(t),y(t))} = 0.$$

Le comportement qualitatif des solutions en découle: oscillations entretenues. Cette conséquence est d'ailleurs une des raisons du succès de ce modèle chez les biologistes car ce type de comportement correspond bien à leurs observations de cycles entretenus dans la réalité. Par contre une conséquence moins acceptable est que l'amplitude de ces oscillations dépend non seulement des paramètres du modèle mais également des conditions initiales et comme le montre l'allure du portrait de phase, (figure 5) plus l'effectif est petit au départ, plus il est grand au bout d'une demi-période. Par ailleurs, ces amplitudes sont très sensibles aux perturbations. De nombreux travaux mathématiques ont été réalisés sur les propriétés des systèmes prédateur-proie, mettant en relief les relations entre le comportement qualitatif et les hypothèses structurelles assorties de leur signification biologique pour de tels systèmes.

Fig5: Oscillations entretenues du modèle de Lotka-Volterra



- predator = predator + attack - mort
INIT(predator) = 10
- prey = prey + nat - attack
INIT(pre) = 1
- attack = 0.25*prey*predator
- mort = 0.4*predator
- nat = 1.2*prey

On notera d'autre part que certaines limitations sont inhérentes au formalisme différentiel lui-même: les modèles différentiels contiennent déjà un certain nombre d'hypothèses qui constituent autant de limites à leur domaine d'applications.

Par exemple, comme le fait remarquer J.C. Hervé [HERVE, 87] la classe de modèles ci-dessous, correspondant aux modèles prédateur-proie:

$$\frac{dx}{dt} = x * F(x,y)$$

$$\frac{dy}{dt} = y * G(x,y)$$

sous-entend que les hypothèses suivantes sont satisfaites:

- " 1. chaque population peut être décrite par une variable. Les individus sont considérés comme identiques: leurs différences n'ont pas d'influence notable sur la dynamique du système.
2. Les variables mesurant la taille des deux populations sont continues. Cela suppose que les effectifs de chaque population sont suffisamment grands.
3. Le temps est une variable continue. (Ceci exclue un certain nombre de phénomènes saisonniers.)
4. Les fonctions F et G dépendent uniquement des variables x et y. L'influence des facteurs du milieu extérieur peut être prise en compte par l'intermédiaire des paramètres mais elle est ainsi globalisée de manière statique. Les variations du milieu extérieur sont donc négligées.
5. Les effets de l'interaction entre les deux populations sont supposées instantanées et la dynamique du système est indépendante de son histoire passée. En particulier on ne tient pas compte des différences d'âges dans les populations et il n'y a pas d'accumulation, au cours de l'évolution du système du produit influençant sa dynamique.
6. Les fluctuations stochastiques du milieu extérieur sont négligées. "

Ces limitations sont inhérentes au principe même de la modélisation: un modèle n'est qu'une "idéalisée de la réalité". Si le niveau de connaissances véhiculé par le modèle s'avère insuffisant dans la compréhension de certains comportements, il peut être nécessaire de recourir à un modèle mathématique de complexité supérieure, permettant d'accéder à un grain plus fin de connaissances.

Dans ce travail, on se limite pour l'instant, aux connaissances superficielles et aux connaissances profondes. On étudie le raisonnement qui permet d'inférer des hypothèses sur la structure du système à partir des observations sur son comportement.

1.1.1. Connaissances profondes

Pour formaliser ces connaissances deux points de vue ont été adoptés, celui de la biologie et celui des mathématiques.

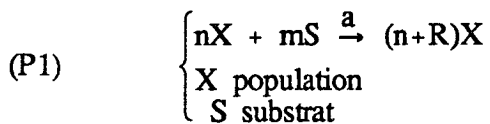
1.1.1.1. Point de vue du biologiste : processus

Le point de vue du biologiste correspond à la notion de processus biologique. Un processus biologique est la description fine dans une sémantique biologique des espèces en présence et de leurs interactions. Cette description véhicule deux types de connaissances. Le premier est de type qualitatif. Par exemple, la description d'un processus en tant que "processus de croissance d'une population par consommation d'un substrat", est purement qualitative. En effet, seule la nature des interactions est précisée, aucune indication quantitative n'est donnée à ce niveau: proportions de chacune des espèces, cinétique du processus, rendement etc....

Si l'on désire accéder à un grain plus fin de connaissance et préciser quantitativement ces informations, il faut faire appel à une formalisme précis. Etant donnée la parenté observée entre les

modèles de la cinétique chimique et ceux de la dynamique des populations, le formalisme choisi est celui des réactions "pseudo-chimiques". L'autre fondement de ce choix, en est que ce dernier se distingue des autres formalismes de type schémas fonctionnels, par le fait qu'il se situe au niveau connaissances profondes et non pas au niveau intermédiaire de connaissances (voir plus haut).

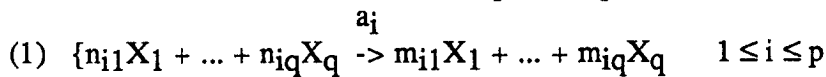
Par exemple, soit le processus biologique :



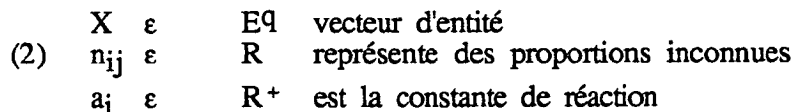
P1 symbolise un processus de croissance par consommation de substrat. Il s'agit d'une croissance de X avec un rendement R dans la mesure où X apparait dans les membres de gauche et de droite respectivement dans les proportions n et n+R. Il s'agit d'une consommation de substrat dans la mesure où S n'apparait que dans le membre de gauche de la réaction a est la constante de réaction.

Plus généralement:

La représentation dans le formalisme pseudo-chimique d'un système biologique sera appelée "processus". Sa définition est l'homologue, dans le formalisme pseudo-chimique, de celle de modèle dans le formalisme mathématique. Un processus est défini par la donnée de p réactions:



où



Ce formalisme permet de représenter les processus de base mis en oeuvre en dynamique des populations (consommation de substrat, vieillissement, mortalité, renouvellement, compétition, etc...). Il offre une possibilité de construction de nouveaux processus par combinaison de ces processus de base. Il peut donc servir de formalisme intermédiaire à la formulation ou à la construction de nouveaux modèles.

Un processus biologique (défini de cette manière) véhicule donc à la fois connaissances qualitative et quantitative.

Un processus biologique est une idéalisation du système, c'est la modélisation du système (structure et fonctionnement) dans le formalisme biologique.

On notera qu'un processus biologique peut être éclaté en plusieurs sous-processus, en particulier en processus "élémentaires". Inversement, on peut synthétiser de nouveaux processus par combinaison de sous-processus. Ce point est explicité en 2.1.2.

1.1.1.2. Point de vue du mathématicien : modèle

Un système dynamique peut être conçu comme une transformation dynamique de grandeurs particulières (les entrées) en d'autres grandeurs (les sorties) distinguées des premières parce qu'elles donnent les caractéristiques de ce qui est produit par le système. [FAURRE ET AL, 1974]. Il est courant de représenter graphiquement un système dynamique par un diagramme fonctionnel plus connu sous le nom de "boîte noire" (Fig6).

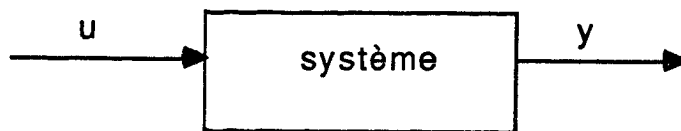


Fig6: Diagramme fonctionnel

La description des systèmes dynamiques utilise le formalisme mathématique. La représentation mathématique de cette transformation entre l'entrée u et la sortie y est appelée **modèle**. En automatique, cette représentation mathématique présente l'avantage d'avoir une définition rigoureuse. La définition d'un modèle est très précise. Il s'agit soit de description externe soit de description interne:

La description externe d'un système est une relation d'entrée-sortie formalisée en terme mathématique par une fonction f

$$y(t) = f(u(\cdot), t) \quad \text{où} \quad \begin{array}{l} t \text{ désigne le temps} \\ u \text{ désigne l'entrée} \\ y \text{ désigne la sortie} \end{array}$$

La description interne par vecteur d'état est formalisée en terme mathématique par deux fonctions :

$$\begin{array}{ll} x(t) = f(t, t_0, x_0, u) & \text{fonction de transition d'état} \\ s(t) = g(x(t), u(t), t) & \text{fonction de sortie} \end{array}$$

Dans le cas des systèmes différentiels, la fonction d'état est définie indirectement par une équation différentielle, équation d'état:

$$\dot{x} = h(x, u(t), t).$$

Un modèle est une idéalisation du système dans le formalisme mathématique.

Par exemple, on peut modéliser l'évolution de la taille du rat musqué par le système d'équations différentielles

$$\begin{cases} \frac{dx}{dt} = - aRs \\ \frac{ds}{dt} = - bxs \end{cases}$$

où x représente la taille et s représente le facteur hormonal de croissance.

Un modèle véhicule également deux types d'informations : qualitatives et quantitatives

- des informations qualitatives

par exemple la solution de l'équation $\frac{dx}{dt} = ax(1 - \frac{x}{K})$ est croissante, présente un point d'inflexion, a une asymptote.

- des informations quantitatives: la donnée exacte de l'équation fournit la correspondance entre x et t point à point. Cette correspondance est bien sûr une information plus forte que l'information précédente.

Les connaissances qualitatives sont "plus faibles" que les secondes. Ces notions sont développées dans une tendance relativement récente en modélisation, celle de la modélisation qualitative

[KUIPERS, 86]. Par exemple, dans le formalisme de Kuipers, la contrainte $x \stackrel{M}{\leq} y$ (signifiant que tout accroissement de x implique un accroissement de y) est évidemment une relation entre x et y plus faible que celle contenue dans l'équation $y = 2x$. Kuipers, montre plus généralement que les équations différentielles qualitatives déduites d'un système d'équations différentielles ordinaires sont plus "faibles" que celles-ci. Il précise le sens de plus faible: plus faible signifie que toute solution du système différentiel satisfait nécessairement l'ensemble des contraintes qualitatives.

De la même manière par exemple dire qu'une courbe est croissante avec une asymptote est une information plus faible que celle fournie par la donnée complète d'un graphe.

On peut visualiser cela sur un dessin. Considérons la courbe associée à l'équation précédente.

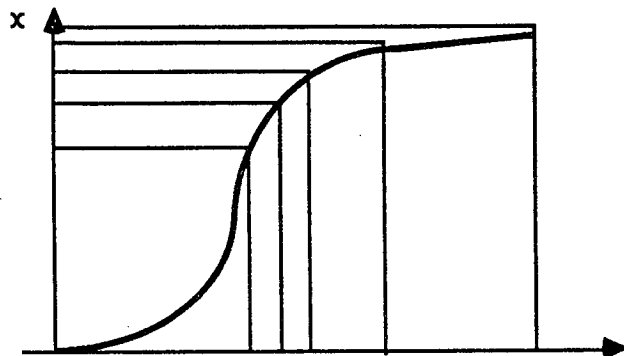


Fig7: graphe du modèle logistique

Ce graphe véhicule une information qualitative : courbe croissante avec inflexion et asymptote. L'information complète (quantitative) est la donnée de tous les points du graphe.

Remarque sur la définition d'un modèle

Il est à noter qu'un système d'équations différentielles est toujours défini à une reparamétrisation ou à un changement de variable près. Ainsi soient par exemple:

$$M1: \begin{cases} \frac{dx}{dt} = ax - bx^2 \end{cases}$$

$$M2: \begin{cases} \frac{dx}{dt} = ax(1 - \frac{x}{k}) \end{cases}$$

$$M3: \begin{cases} \frac{dx}{dt} = bRxS \\ \frac{ds}{dt} = -bxS \end{cases}$$

Les systèmes M1, M2, M3 sont équivalents.

On pourrait définir formellement cette définition d'équivalence entre systèmes par

E, U (resp. E', U') étant l'ensemble des variables d'état et des paramètres du système M (resp de M')
 $M \equiv M'$ ssi il existe une application linéaire bijective de $E * U$ dans $E' * U'$ telle que $F(M) = M'$

On devrait donc appeler en fait modèle la classe d'équivalence obtenue par cette relation d'équivalence. Par abus de langage, on confond un modèle et sa classe d'équivalence, étant entendu alors qu'un modèle est défini à une reparamétrisation ou un changement de variables près.

Ainsi les modèles M1, M2, M3 sont tous des représentants de la même classe d'équivalence: le modèle logistique. Une sous-classe de logistique est l'ensemble des représentants du modèle logistique équivalents à une reparamétrisation près. M1 et M2 appartiennent à la même sous-classe: logistic-forme1. Cette remarque triviale a toutefois son intérêt car elle est à l'origine de la multiplicité des interprétations biologiques qui peuvent être données pour un même modèle.

1.1.2. Connaissances superficielles

Il s'agit également ici de de deux types de connaissances, biologique ou mathématique. Ces connaissances concernent la courbe de réponse d'une part, la situation biologique d'autre part. Pour chacune d'elle, le biologiste dispose de deux sortes de connaissances: des connaissances générales sur des "pattern" de courbes et de situation, des connaissances spécifiques sur les caractéristiques de ses observations. Celles-ci lui permettent d'identifier respectivement un type de courbe ou un type de situation dans le catalogue adéquat de sa connaissance.

1.1.2.1 Courbe de réponse

Le biologiste a accès à des données expérimentales (chronologie concernant l'évolution d'une variable etc....). Il en extrait

- des informations qualitatives:

| | |
|---------------|---------------------|
| | <u>observations</u> |
| la courbe est | monotone |
| | croissante |
| | asymptote |
| | point d'inflexion |

A partir de ces observations et de ses connaissances sur les courbes de modèle, il identifie un "schéma" de courbe, ou "pattern" dans une hiérarchie de courbes. Le plus souvent, ce pattern est lui-même associé à un modèle ou un processus. (car les connaissances générales sur les modèles et les processus, incluent des associations de type modèle-courbe, processus-situation, qui permettent de proposer une classe de modèle ou de processus). La tâche principale impliquée dans ce processus d'identification est une tâche de classification (similaire aux processus de diagnostic, en général). La stratégie de résolution utilisée pour résoudre ce type de problème, consiste souvent à définir une hiérarchie dans laquelle chaque noeud contient la connaissance sous forme de règles de diagnostic du type: < symptômes > = > < évidence > d'un < concept de la hiérarchie >, et à procéder par pattern-matching.

Par exemple, si la courbe de réponse est monotone, croissante, avec une asymptote et un point d'inflexion, il en déduira qu'il peut s'agir de la "courbe logistique" (courbe correspondant au modèle logistique) ou de la courbe "gompertz". Par contre l'hypothèse, courbe "monomoléculaire" ou "monod" est rejetée. Ce type de connaissances superficielles est souvent représentée à l'aide d'un ensemble de règles du type:

SI croissante \wedge asymptote \wedge inflexion ALORS courbe-logistique
 SI croissante \wedge asymptote \wedge inflexion ALORS courbe-gompertz
 SI croissante \wedge asymptote \wedge inflexion ALORS \neg courbe-monod
 SI croissante \wedge asymptote \wedge inflexion ALORS \neg courbe-monomoléculaire

SI courbe-logistique alors modèle-logistique.
 SI courbe-gompertz alors modèle-gompertz

Le § 2.1.1 mettra en évidence les avantages du formalisme objet pour la représentation de ce type de connaissances: économie, clarté dans la représentation des concepts et des connaissances, et similitude avec le raisonnement humain.

- Il dispose d'informations quantitatives: la chronologie elle-même. Une fois choisie une classe de modèle, il procède à l'identification des paramètres de ce modèle à l'aide de ces données quantitatives.

1.1.2.1 Situation

Quelques notions biologiques usuelles sont précisées en ce qui concerne l'évolution d'une espèce:

- la notion de *croissance* ou de *décroissance* fait référence à la croissance ou à la décroissance ("dégradation", "mortalité") de l'espèce étudiée et non à l'évolution des autres espèces impliquées dans le phénomène.

- la notion de *régulation* fait référence à la présence de mécanismes ayant pour effet de limiter la croissance. Par exemple, il peut s'agir d'une situation où la croissance est auto-régulée. Il peut également s'agir de situations où la croissance est régulée par l'intervention de facteurs extérieurs. Certains facteurs extérieurs peuvent être des facteurs nécessaires à la croissance; on en distingue alors de deux sortes: les substrats et les catalyseurs.

- la notion de *substrat* correspond à un facteur extérieur nécessaire à la croissance, consommé ou assimilé par la biomasse. Il peut être indépendamment également produit par la biomasse (s- logistique4). On parle alors de *renouvellement* du substrat. Il s'agit de "ressources" si le substrat intervient dans une situation de croissance. Il s'agit de substance "toxique" si il intervient dans une situation de décroissance. Le système étant fermé (hypothèses de populations dans un milieu isolé) ces substrats sont forcément limitants.

- elle doit être opposée à la notion de *catalyseur* (ou enzyme) qui est également un facteur nécessaire à la croissance mais qui n'est pas consommé par la biomasse. Ils peuvent par contre être produits par la biomasse (s- logistique3.2). On dit également alors que le catalyseur est renouvelé.

- la notion de *saturation* correspond à un "ralentissement de dynamique" au cours du temps. Au bout d'un certain temps les changements Δx observés en fonction de Δt diminuent en valeur absolue au fur à mesure que t augmente.(cf [LEBRETON-MILLIER p.154]. Les causes de saturation sont diverses. En ce qui concerne la saturation par excès de substrat, il s'agit de mécanismes limitant les possibilités de consommation du substrat par la biomasse. Cette limitation peut être due à un mécanisme propre, c'est à dire au type de consommation elle-même (s- Monod). La saturation peut être également liée à la nature des autres mécanismes intervenant dans la croissance. Par exemple la saturation peut être observée lors de la consommation d'un substrat qui est renouvelé mais qui se dégrade simultanément, (s-monomoléculaire). Elle peut également être observée lorsque le substrat est renouvelé mais ne se dégrade pas. Elle est alors plus tardive (s- logistique4). Lorsqu'il y a consommation de substrat mais sans mécanisme indépendant d'évolution du substrat (ni renouvellement, ni dégradation), il peut également y avoir apparition d'une saturation plus tardive (s- logistique2).

Le biologiste a accès à des données biologiques sur la situation étudiée, par exemple, il sait que la

observations

la situation présente les caractéristiques: croissante
régulée
saturation-secondaire

A partir de ces informations et de ses connaissances sur les situations, il identifie un "schéma" de situation et en déduit une classe de modèle possible. Par exemple, le biologiste sait que la croissance en taille des vertébrés est liée à la présence d'un facteur hormonal de croissance. C'est pourquoi il trouvera justifié de la représenter par le modèle de gompertz.

1.2 Relations entre les différentes connaissances

On a montré que l'on pouvait distinguer selon le point de vue adopté, mathématique ou biologique, deux types de connaissances profondes (modèle, processus), et deux types de connaissances

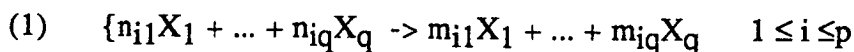
superficielles (courbe, situation). Les relations entretenues par ces diverses connaissances sont développées ci-dessous.

L'étude des interdépendances entre connaissances est une phase indispensable préalable à la représentation de ces connaissances dans un formalisme quelqu'il soit. Il est peut-être encore plus indispensable de faire ce travail d'analyse pour une représentation dans un formalisme objet à cause de la nécessité impérative de structurer les connaissances en "réseaux d'objets interconnectés".

1.2.1 Relation entre les connaissances profondes: liens entre processus et modèle

1.2.1.1 Relation processus-modèle

Ces deux représentations sont bien sûr étroitement liées puisque l'une et l'autre exprime la nature profonde des lois qui régissent le système. Les relations entre les systèmes d'équations différentielles multilinéaires de la cinétique chimique et le formalisme utilisé pour représenter les réactions est d'ailleurs connu depuis longtemps [Emanuel, Knorre]. Une procédure de traduction formalisée dès 1961 par Garfinkel permet d'associer au système de réactions:



le système d'équations différentielles :

$$(2) \quad \frac{dX_j}{dt} = \sum_{i=1}^p a_i(m_{ij} - n_{ij}) \prod_{k=1}^q (X_k)^{n_{ik}} \quad (1 \leq j \leq q)$$

On peut donc associer par cette procédure, à un processus P_i , un unique système d'équations différentielles M_i . Par extension, on peut définir une autre application qui lui associe toute une famille de systèmes M_j équivalents à M_i (remarque 1.1.1.2 page 13)

Exemple de relations entre le modèle logistique et différents processus:

Processus

Modèles

LOGISTIQUES

1 - logistique 1

$$M.L.1.1. : \left\{ \frac{dx}{dt} = rx \left(1 - \frac{x}{k}\right) \right.$$

$$P.L1 \left\{ X \frac{a}{b} \right.$$

----->

↑

$$M.L.1.2. : \left\{ \frac{dx}{dt} = ax - bx^2 \right.$$

2 - logistique 2

$$P.L.2 \left\{ \begin{array}{l} X + S \xrightarrow{b} (1 + R)X \\ \phantom{\xrightarrow{b}} \end{array} \right. \text{-----} > \quad M.L.2.: \quad \begin{cases} \frac{dx}{dt} = bRxS \\ \frac{ds}{dt} = - bxS \end{cases}$$

3 - logistique 3

$$P.L.3.2 \left\{ \begin{array}{l} nX + S \rightarrow (n+R) X+S \\ X+S \rightarrow 2S + X \\ S \rightarrow 0 \end{array} \right. \text{-----} > \quad M.L.3.2 \quad \begin{cases} \frac{dx}{dt} = rx^n \left(\frac{1}{x^{n-1}} - \frac{1}{Kx^{n-2}} \right) = r_1 x^n s \\ \frac{ds}{dt} = r_2 \left(\frac{1-n}{x^n} + \frac{n-2}{Kx^{n-2}} \right) \frac{dx}{dt} = - aS + bxS \end{cases}$$

(avec $n > 2$)

$n = 2$ (alors $b=0$)

$$P.L.3.2 \rightarrow P.L.3.1 \left\{ \begin{array}{l} 2X + S \xrightarrow{c} (2+R)X + S \\ S \xrightarrow{a} 0 \end{array} \right. \text{----} > \quad M.L.3.1. \quad \begin{cases} \frac{dx}{dt} = r_1 x^2 s = CRx^2 S \\ \frac{ds}{dt} = - aS \end{cases}$$

4 - logistique 4

$$P.L.4 \left\{ \begin{array}{l} X + S \xrightarrow{b} (1+R)X \\ X \xrightarrow{a} \frac{1}{R}S \end{array} \right. \text{-----} > \quad M4 \quad \begin{cases} \frac{dx}{dt} = bRxs - ax \\ \frac{ds}{dt} = - bxs + \frac{a}{R}x \end{cases}$$

Autres modèles:

GOMPERTZ

$$M.G.1. \quad \left\{ \frac{dx}{dt} = rx \left(1 - \log \frac{x}{K} \right) \right.$$

$$P.G. \left\{ \begin{array}{l} X + F \xrightarrow{a} (1+R)X+F \\ F \xrightarrow{b} 0 \end{array} \right. \text{-----} >$$



$$M.G.2. : \quad \begin{cases} \frac{dx}{dt} = xf \\ \frac{df}{dt} = - af \end{cases}$$

JOHNSON

$$\text{P.J. } \begin{cases} X + 2F \xrightarrow{a} (1+R)X + 2F \\ 2F \rightarrow 0 \end{cases} \text{ ----- } > \quad \text{M.J. } \begin{cases} \frac{dx}{dt} = axf^2 \\ \frac{df}{dt} = -af^2 \end{cases}$$

LUNDQVIST

$$\text{P.L. } \begin{cases} X + cF \rightarrow (1+R)X + cF \\ cF \rightarrow 0 \end{cases} \quad \text{M.LU. } \begin{cases} \frac{dx}{dt} = axf^c \\ \frac{df}{dt} = -af^c \end{cases}$$

On visualisera les correspondances sur la figure ci-dessous en ce qui concerne le modèle logistique:

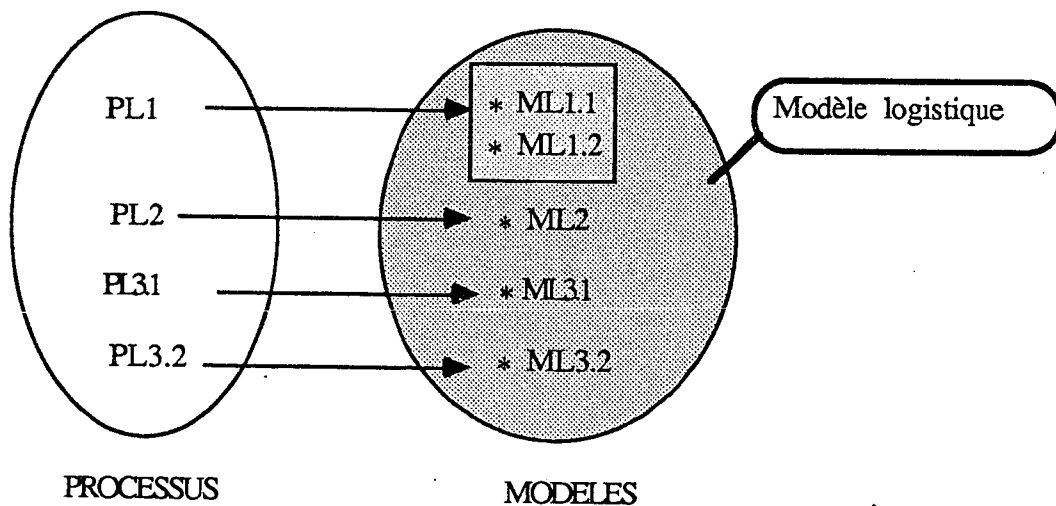


Fig8: Correspondance processus, modèle

A un processus P_i est associé un représentant M_i unique d'une classe de modèle.

$$P_i \rightarrow M_i$$

On peut donc lui associer une classe unique \mathcal{M} de modèles (et donc toute une famille de représentants M_j dans cette classe)

A un processus biologique P correspond un seul modèle mathématique \mathcal{M} (au sens classe d'équivalence) mais une infinité de représentants équivalents M_j .

$$1 \text{ processus } P_i \rightarrow 1 \text{ modèle } \mathcal{M}.$$

Corollaire: plusieurs processus P_i peuvent avoir la même image \mathcal{M} .

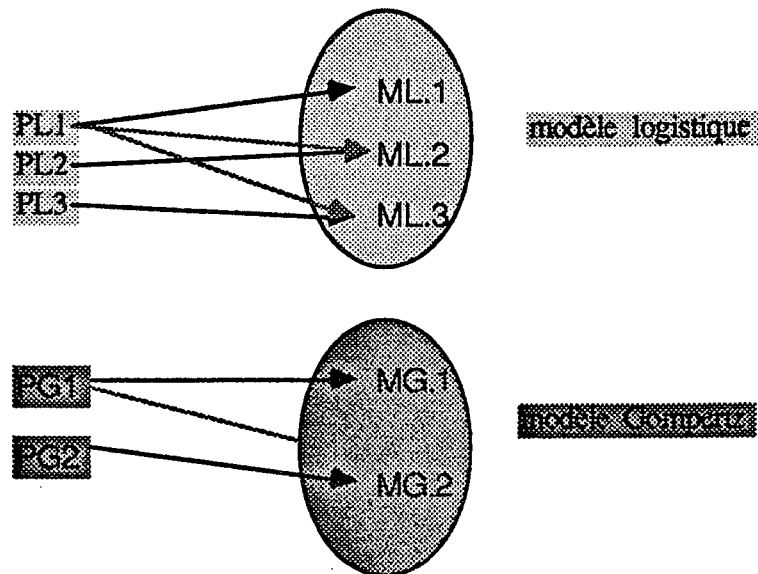


Fig9: Plusieurs processus P_i peuvent avoir la même image \mathcal{M} .

1.2.1.2 Etude de la relation inverse: modèle - > processus

La correspondance modèle, processus n'est pas unique. Plus précisément, à un représentant d'une classe de modèle, on ne peut associer un processus de manière unique.

Les exemples donnés sur ce problème sont souvent ambigus. Il y a parfois confusion entre un représentant d'une classe \mathcal{M} et cette classe.

Il est en effet évident qu'étant donné l'infinité de changement de variables possibles, une infinité de systèmes différentiels sont équivalents à un système différentiel donné, et donc une infinité de processus peuvent être associés à une classe de modèles. Ceci vient d'être noté dans le corollaire ci-dessus (plusieurs processus P_i ont la même image \mathcal{M}). Sur la figure ci-dessus à la classe \mathcal{M} du modèle logistique, peut être associée une infinité de processus différents : P1, P2, P3, etc.... Il s'agit là de l'étude de la relation inverse de l'application $P_i \rightarrow \mathcal{M}$.

Il est par contre plus intéressant d'étudier la relation inverse de $P_i \rightarrow M_j$ et de noter qu'à un système différentiel donné, peut être également l'image d'une infinité de processus possibles.

Exemple :

Soient les processus P1 et P2 :

$$P1 : X \xrightarrow{b} p_1 S$$

$$P2 : \begin{cases} X \xrightarrow{b} p_1 S \\ X \xrightarrow{c} p_2 S + X \end{cases}$$

Les systèmes correspondants sont respectivement :

$$M1 : \left\{ \frac{dx}{dt} = -bx \right.$$

$$M2 : \begin{cases} \frac{dx}{dt} = -bx \\ \frac{ds}{dt} = bp_1 + cp_2 \end{cases}$$

Les modèles associés à P1 et P2 sont donc identiques si $bp_1 + cp_2 = 0$ soit

$$p_2 = -\frac{b}{c} bp_1$$

Cet exemple montre que deux processus distincts P1 et P2 peuvent avoir pour image le même système différentiel M:

$$\begin{array}{l} P1 \setminus \\ M1 = M2 \\ P2' \end{array}$$

On pourrait exhiber de manière plus générale une CNS pour que deux processus correspondent au même modèle, sous forme d'un système algébrique (reliant les constantes de réaction et les coefficients stoechiométriques; ce point a été abordé dans la thèse [PAVE 80]).

Définition d'une relation d'équivalence entre processus :

Deux démarches sont possibles:

On peut définir une relation d'équivalence \mathcal{R} sur l'ensemble des processus par

$$P \mathcal{R} P' \text{ SSI } M \sim M'$$

Dans ce cas par, en notant \mathcal{P} , l'ensemble quotient obtenu, la correspondance $\mathcal{P} \rightarrow \mathcal{M}$ est par définition bijective. Il serait intéressant de trouver une signification biologique à cette relation \mathcal{R} .

Ce qui semble plus intéressant en fait serait de définir la relation \mathcal{R} de manière intrinsèque (c'est à dire en lui donnant une définition "biologique") et d'étudier la bijectivité de la correspondance entre processus et modèle sur les ensembles quotients obtenus \mathcal{M} et \mathcal{P} . C'est-à-dire de démontrer que la correspondance entre \mathcal{M} et \mathcal{P} est bijective. (On pourrait par exemple expliciter la signification biologique de la relation mentionnée plus haut sur les constantes de réaction et les coefficients stoechiométriques).

On a donc vu qu'un modèle peut être l'image de plusieurs processus distincts :

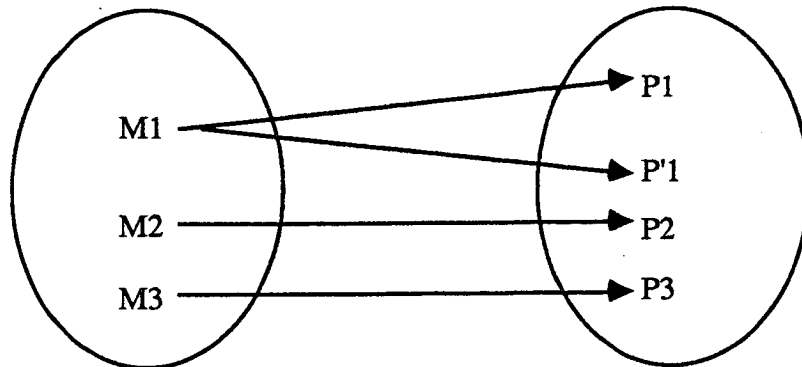


Fig11: relation inverse, modèle-processus

En conclusion :

la correspondance de $P \rightarrow \mathcal{M}$ est une application (à un processus P correspond une classe \mathcal{M} unique) non bijective (une classe \mathcal{M} peut être l'image de plusieurs processus).

1.2.2. Relations entre les connaissances profondes et les connaissances superficielles.

A une classe de modèle \mathcal{M} correspond une courbe \mathcal{C} par exemple, au modèle logistique correspond la courbe logistique (on peut voir cette question plus finement cf. Houllier).

A un processus P correspond une situation S par exemple, au processus P.L.1. : $x \frac{a}{b} 2x$ correspond une compétition intra-spécifique.

En tenant compte de l'étude du paragraphe précédent :

- à un processus P correspond une classe de modèle \mathcal{M}
- à un modèle \mathcal{M} correspond une courbe \mathcal{C}
- à un processus P correspond une situation S

soit :

| Point de vue / Connaissances | BIOLOGIQUE | MATHEMATIQUE |
|------------------------------|---------------|----------------------------------|
| PROFONDES | 1 processus P | 1 classe \mathcal{M} de modèle |
| SUPERFICIELLES | 1 situation S | 1 courbe \mathcal{C} |

\downarrow (from P to S) \rightarrow (from P to \mathcal{M}) \downarrow (from \mathcal{M} to \mathcal{C})

Fig12: Associations causales entre les différentes connaissances

1.2.3 Relations entre les connaissances superficielles: courbe et situation

Un type de courbe n'est compatible qu'avec certains types de situation et réciproquement. Par exemple une courbe avec asymptote n'est compatible qu'avec une situation régulée. Plus la description d'une courbe est raffinée, plus la description de la situation est précise. Par exemple la courbe logistique qui est une courbe monotone croissante avec asymptote et inflexion au milieu de la courbe, n'est cohérente qu'avec des situations de type croissance régulée avec saturation moyenne. Courbe et situation sont des connaissances étroitement corrélées par le biais du réseau d'associations causales ci-dessus. Elles doivent donc être cohérentes.

1.2.4 Exemple d'associations causales pour le modèle logistique

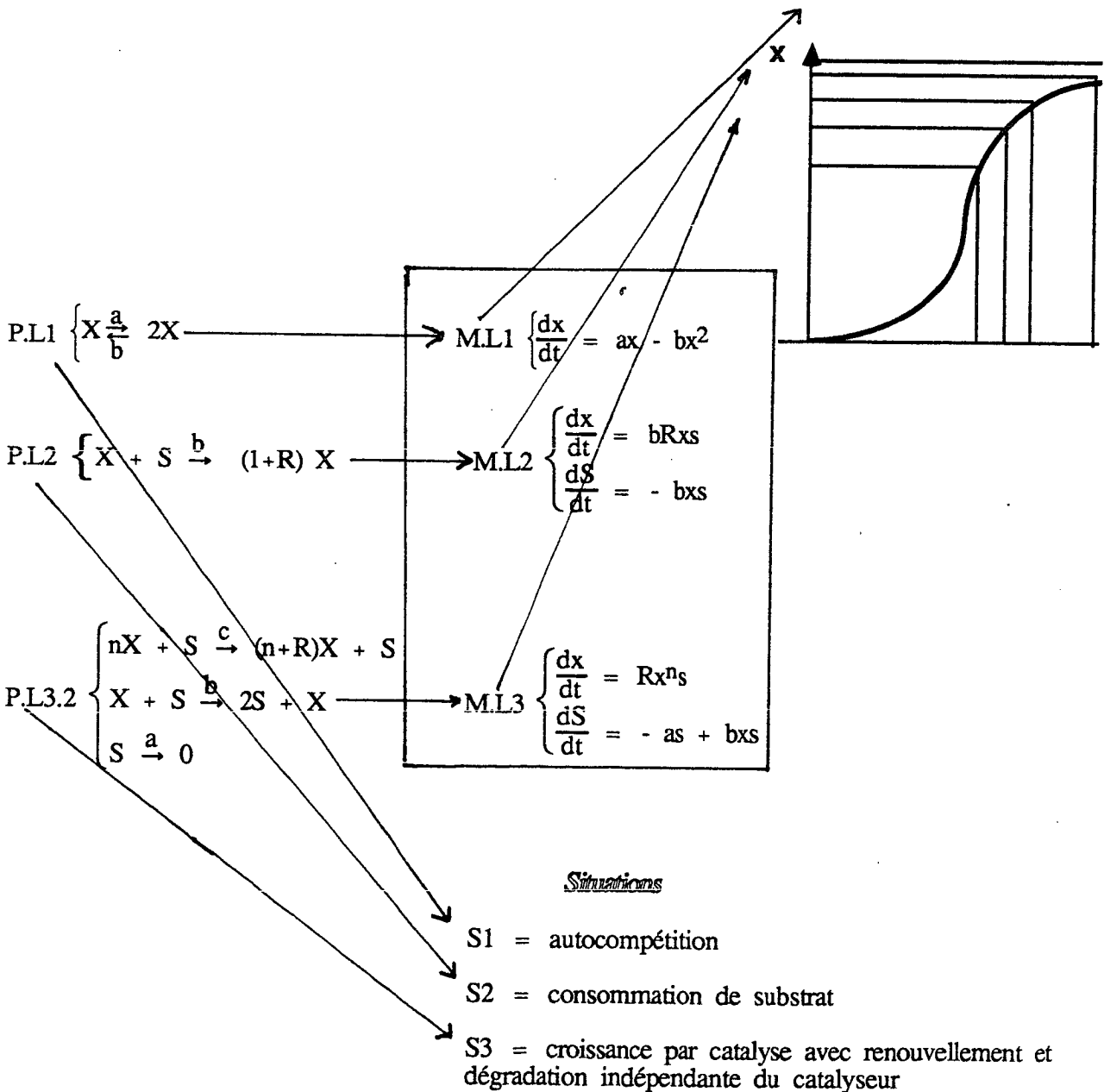
Processus

Modèle

Courbe

M: modèle logistique

C: courbe logistique



Le réseau ci-dessus peut être lu de différentes manières en formulant soit les relations entre un processus et les autres connaissances :

On peut associer au processus P.L1

- la situation S1
- le modèle ML1
- la courbe C

de même on peut associer au processus P.L2

- la situation S2
- le modèle ML2
- la courbe C

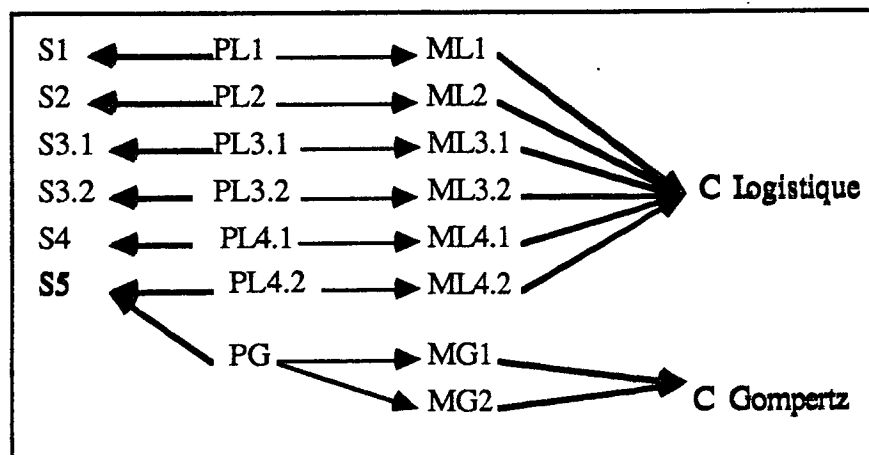
etc...

soit les relations entre un modèle et les autres connaissances :

On peut associer au modèle \mathcal{M}

- les processus PL1, PL2, PL3
- les situations S1, S2, S3
- la courbe C

On notera que le réseau général d'associations est complexe puisqu'une courbe peut-être associée à plusieurs situations et inversement, à une situation donnée peut correspondre différentes courbes. Par exemple, la figure 13 montre que les courbes logistique et gompertz peuvent toutes deux être associées à la même situation (S5) de croissance limitée par un catalyseur se dégradant indépendamment.



S: situation
 PL: processus logistique
 PG: processus gompertz
 C: courbe

Fig13: Réseau d'associations entre les différentes connaissances

La raisonnement utilisé en modélisation est basé sur ce réseau de relations puisqu'il s'agit pour le biologiste d'inférer les modèles \mathcal{M} ou les processus P susceptibles d'engendrer une courbe C et une situation S données. Comment ce réseau d'associations causales est-il exploité dans la pratique courante de modélisation? Ce point est abordé en 1.3

1.3 Analyse du raisonnement de formulation d'un modèle

La formulation d'un modèle consiste à trouver l'ensemble des modèles possibles pouvant engendrer des manifestations données.

Pour résoudre ce problème, on conduit un raisonnement en deux étapes: hypothèses et test.

La démarche habituelle pour la formulation du modèle repose sur un raisonnement "d'inférence abductive". On rappelle rapidement ci-dessous la distinction entre syllogismes déductif et abductif, par une présentation de ces syllogismes à l'aide de règles:

raisonnement déductif: étant donné le fait "A" et la règle "A -> B", inférer "B".

raisonnement abductif: étant donné le fait "B" et la règle "A -> B", inférer "A plausible".

On notera que dans le syllogisme déductif la flèche "->" fait référence à l'implication logique, alors que dans le syllogisme abductif elle fait plutôt référence à une notion d'association causale entre A et B. D'autre part, comme le souligne Reggia J. [REGGIA, 86]:

"The term 'abduction' refers to any reasoning in which the goal is to derive the best explanations for a given set of observed facts. Tasks that are abductive in nature typically involve probabilistic, context sensitive disambiguation of problem features using associative knowledge. They involve the construction of a solution, as well as selection from alternative solutions".

Le problème de la modélisation correspond bien à un tel cadre.

On peut distinguer deux types de problèmes, celui du choix et celui de la construction d'un modèle:

Le choix d'un modèle au sein d'une base prédéfinie de modèles, peut être classé dans la catégorie des problèmes de diagnostic (selection d'une solution).

La construction d'un modèle peut être placée dans le cadre des problèmes de conception (construction d'une solution).

On abordera d'abord la question du choix et de la validation de ce choix.

1.3.1 Choix du modèle et validation

1.3.1.1 Choix

Le raisonnement de choix d'un modèle peut être rapproché du raisonnement de diagnostic sur deux points:

Un processus de diagnostic est, en général, complexe mais au coeur de ce processus se trouve une tâche centrale qui consiste à identifier la position d'un objet (i.e. **classifier**) au sein d'un ensemble de classes (classification). Ce problème est largement abordé dans le domaine médical. De nombreuses publications ont été faites sur le sujet. Citons pour mémoire, "Representation and strategy in diagnostic problems solving" [WORTMAN, 66], ou la revue récente "Review of Causal Reasoning In Medical Diagnosis de [PATIL, 86].

Le raisonnement de diagnostic est généralement considéré comme un **processus séquentiel d'hypothèses et test**. Comme le souligne [REGGIA, 86] "... diagnostic reasoning is generally accepted to be a sequential hypothesis-and-test (hypothetico-deductive) process during which the diagnostician conceptually constructs a "model" of the underlying causative disorders..." Il précise également que le cycle d'inférences peut être décomposé en trois phases essentielles qui se recourent :

- évocation d'hypothèses
- évaluation ou construction de ces hypothèses
- génération de questions.

L'approche en modélisation est très similaire (cf §3.1: different steps in the modeling process [PIERRET, 88]):

- choix du modèle,
- validation du modèle (simulation, identification, étude qualitative)
- suggestion de nouveaux protocoles expérimentaux.

Le paragraphe précédent, a précisément montré l'importance des associations causales entre modèle, processus, situation, courbe. Dans la pratique courante, le choix d'un modèle repose sur l'inférence du modèle d'après l'identification (qualitative ou quantitative) d'une courbe. Cette démarche repose sur un raisonnement abductif puisqu'il est de la forme :

"le modèle \mathcal{M}_i est capable de produire la courbe C_i , et on sait que la courbe C observée est du type C_i , donc peut-être que le modèle \mathcal{M}_i est la cause de C ".

Dans un système à base de règles, on pourrait dire qu'elle correspond à l'activation par chaînage arrière d'un paquet de règles du type :

SI modèle \mathcal{M}_i ALORS courbe C_i .

1.3.1.2 Validation du modèle

L'étape de validation correspond à la prise en considération des connaissances biologiques sur le système. Cette démarche repose sur un raisonnement abductif-déductif puisqu'il est de la forme :

"le processus \mathcal{P}_i est capable de produire le modèle \mathcal{M}_i , et on sait que le modèle \mathcal{M}_i est plausible, donc peut-être que le processus \mathcal{P}_i est la cause de \mathcal{M}_i ". On infère donc \mathcal{P}_i . Sachant que le processus \mathcal{P}_i produit la situation S_i , et que \mathcal{P}_i est plausible, on en déduit S_i . Il y a alors

confrontation de la situation observée et de S_i .

Le raisonnement correspondrait donc cette fois à une activation toutes les règles ci-dessous, par chaînage arrière puis avant :

| | |
|---|-----------|
| SI modèle \mathcal{M}_i ALORS courbe C_i | (arrière) |
| SI processus \mathcal{P}_i ALORS modèle \mathcal{M}_i | (arrière) |
| SI processus \mathcal{P}_i ALORS situation S_i | (avant) |

1.3.1.3 Exemple

Soit ci-dessous la liste des règles qui seraient nécessaires pour représenter les connaissances précédentes (cf page 24), en ce qui concerne le modèle logistique (ML) et le modèle monomoléculaire (MM):

| | | |
|------------------------|--|---------------------------------|
| règles logistiques | $\left\{ \begin{array}{l} P1 \rightarrow ML1 \text{ (règle 1)} \\ P1 \rightarrow S1 \text{ (règle 2)} \\ P2 \rightarrow ML2 \text{ (règle 3)} \\ P2 \rightarrow S2 \text{ (règle 4)} \\ P3 \rightarrow ML3 \text{ (règle 5)} \\ P3 \rightarrow S3 \text{ (règle 6)} \end{array} \right.$ | ML1 \rightarrow C1 (règle 7) |
| | | ML2 \rightarrow C1 (règle 8) |
| | | ML3 \rightarrow C1 (règle 9) |
| | | |
| | | |
| règles monomoléculaire | $\left\{ \begin{array}{l} P4 \rightarrow MM1 \text{ (règle 10)} \\ P4 \rightarrow S4 \text{ (règle 11)} \\ P5 \rightarrow MM2 \text{ (règle 12)} \\ P5 \rightarrow S5 \text{ (règle 13)} \end{array} \right.$ | MM1 \rightarrow C2 (règle 14) |
| | | MM2 \rightarrow C2 (règle 15) |
| | | |
| | | |

et soient les faits:

cas 1: C_1, S_1 (ou S_2 ou S_3)

cas 2: C_1, S_4

Dans la pratique usuelle le modélisateur procède en deux temps:

- le choix du modèle repose d'abord uniquement sur l'ensemble des règles 7, 8, 9, 14, 15. Comme la courbe observée est C_1 , les modèles ML1, ML2, ML3 sont sélectionnés comme "candidats". (En pratique, l'un des modèles ML1, ML2 ou ML3 est alors arbitrairement retenu.)

- une fois le modèle choisi, pour le valider, on procède à l'identification des paramètres, puis on essaye de voir si l'interprétation biologique du processus sous-jacent au modèle est conforme à la situation observée (cf exemples dans [PAVE 88]). La validation correspond donc à prendre en compte les règles 1, 2, 3, 4, 5, 6.

Dans le cas 1 par exemple, le modèle logistique est retenu dans l'étape de choix du modèle, il est puis validé. Dans le cas 2, le modèle logistique est rejeté.

Généralement, on procède à l'identification de la courbe de réponse, sans même avoir procédé au préalable à sa validation biologique. Un bon ajustement des données est même souvent le premier critère de choix, on se pose alors seulement la question de la valeur "biologique" du modèle.

La séparation des points de vue mathématique et biologique et l'aspect séquentiel de la démarche a plusieurs conséquences:

On observe un certain nombre de déconvenues, notamment quand la modélisation est confiée à un spécialiste "mathématicien". Le modèle qu'il propose ne peut pas toujours être validé par le biologiste. C'est le problème qui a été rencontré par exemple, lors de l'utilisation de Cosmos [Hamrouni, 79]. Cet outil d'aide à la formulation de modèles est uniquement basé sur le point de vue mathématique (ajustement de courbes). Les responsabilités concernant le choix des compartiments, autrement dit le choix des processus, est entièrement laissé à la seule responsabilité de l'utilisateur. On trouvera ci-dessous un exemple précis concernant l'utilisation de COSMOS pour la modélisation du stockage de la sérotonine dans les plaquettes humaines [LAUNAY, 88]. Différents modèles ont été testés à l'aide de COSMOS. Le choix de la structure du modèle a été complètement arbitraire, tant pour le choix du nombre de compartiments (deux, trois ou quatre compartiments) que pour celui des échanges (choix et sens des flèches). Au vu de l'optimisation,

le modèle à quatre compartiments a été retenu (Fig 11). Les biologistes n'ont pu accorder une interprétation biologique qu'au compartiment 3 correspondant aux observations recueillies dans le granule dense. Aucune interprétation cohérente avec leurs hypothèses sur la structure du système biologique (Fig12) n'a pue être trouvée pour les autres compartiments. En effet, il était impossible d'identifier le compartiment extérieur dans le modèle proposé: la dynamique connue pour ce compartiment n'était compatible avec aucune des dynamiques calculées pour les compartiments 1, 2, ou 4.

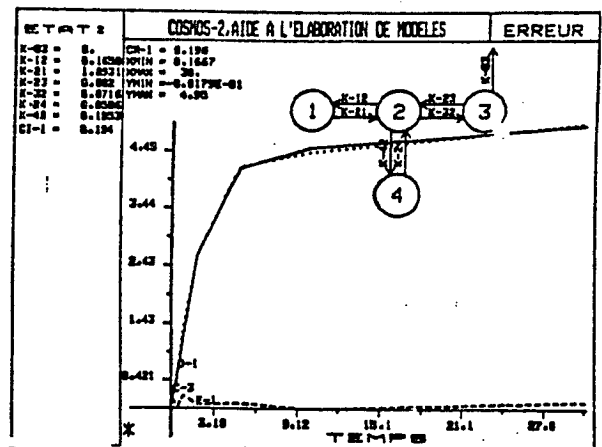
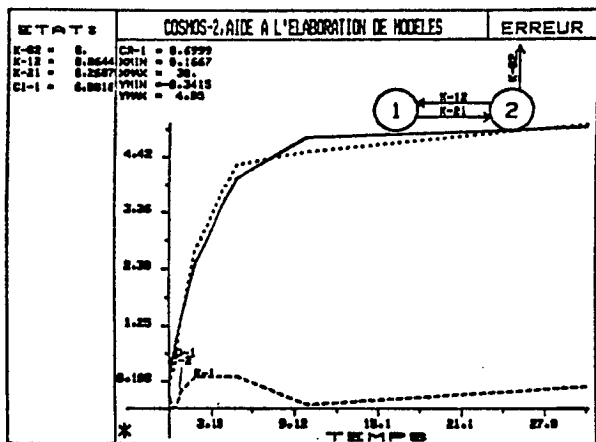
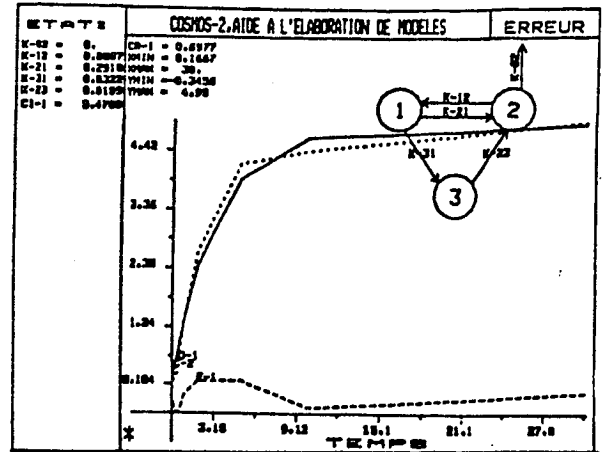
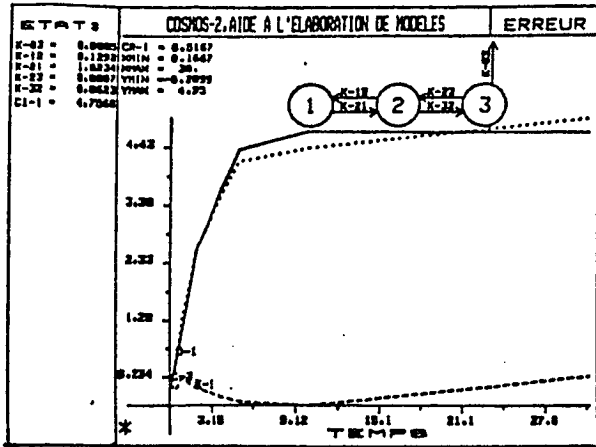


Fig14 Différents modèles proposés avec COSMOS

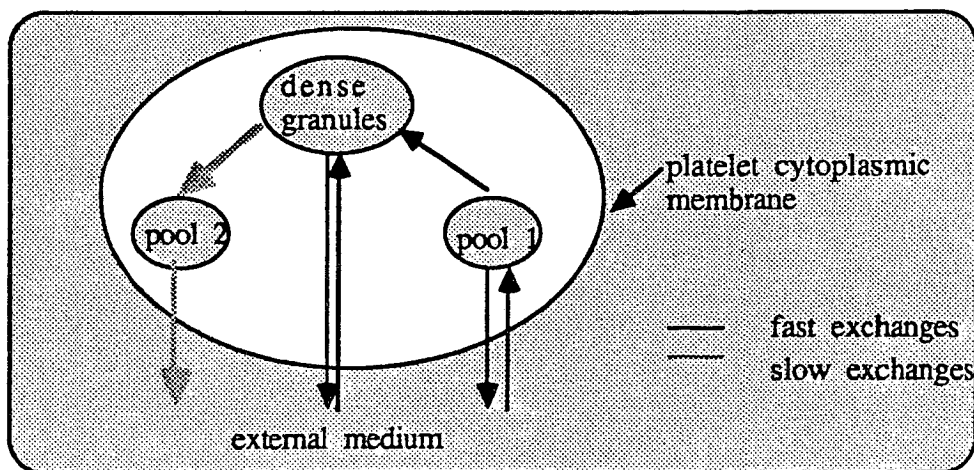


Fig. 15 Modèle de Costa du stockage de la sérotonine dans les plaquettes humaines

On procède ainsi à l'identification de différents modèles candidats, même si l'un d'eux n'est pas valide sur le plan de son interprétation biologique, d'où une perte de temps. Bien souvent d'ailleurs l'identification des paramètres d'un tel modèle conduit à des valeurs de paramètres inacceptables du point de vue de leur signification biologique.

Les différents points de vue de connaissances doivent être utilisés conjointement pour un meilleur choix d'un modèle,

1.3.2 Construction d'un modèle

La construction d'un modèle est une activité complexe, puisqu'un modèle est l'expression d'un bilan correspondant à la combinaison de plusieurs processus biologiques. Ce problème est délicat puisque chaque processus altère les connaissances superficielles observées: la situation et la courbe de réponse sont des observations globalisant les effets individuels de chacun des processus impliqués. La démarche de construction d'un modèle est en général complexe mais au cœur de ce processus se trouvent trois tâches centrales:

- analyse décrementale,
- inférence abductive (classification)
- suivie d'une phase d'incrémentation.

On prendra pour exemple, celui de la construction d'un modèle pour le système prédateur-proie. Il s'agit naturellement d'un exemple d'école, servant uniquement à reconstituer les différentes étapes de la démarche.

1. ANALYSE DECREMENTALE

Dans un premier temps, on procède à l'analyse du système. Cette analyse conduit à la **décomposition** du système en sous-systèmes élémentaires. On raisonne alors en considérant chacun des sous-systèmes comme un système isolé. Par exemple, on considère le système isolé: prédateur (S3), le système isolé proie (S1), le système isolé prédateur-proie (S2). L'heuristique de décomposition d'un système n'est pas approfondie ici. On notera seulement que le choix des sous-structures est un choix fondamental, correspondant à un raisonnement complexe, puisqu'il conduit à des hypothèses fortes sur la structure du système. Par exemple, certaines variables peuvent être considérées comme "sûres", prédateur, proie, (cf par exemple [LAVERGNAT, 79], d'autres variables, variables intermédiaires introduites pour rendre compte de tous les phénomènes, sont "moins sûres" (par exemple l'introduction d'un substrat limitant dans le système prédateur-proie permet de générer des solutions oscillantes amorties). Il est essentiellement guidé par l'identification de situations biologiques élémentaires prototypiques contribuant à la situation globale, et/ou l'identification de courbes élémentaires prototypiques contribuant à la courbe de réponse observée. Ces situations élémentaires étant elles-mêmes associées à des processus élémentaires, et les courbes élémentaires associées à des modèles élémentaires.

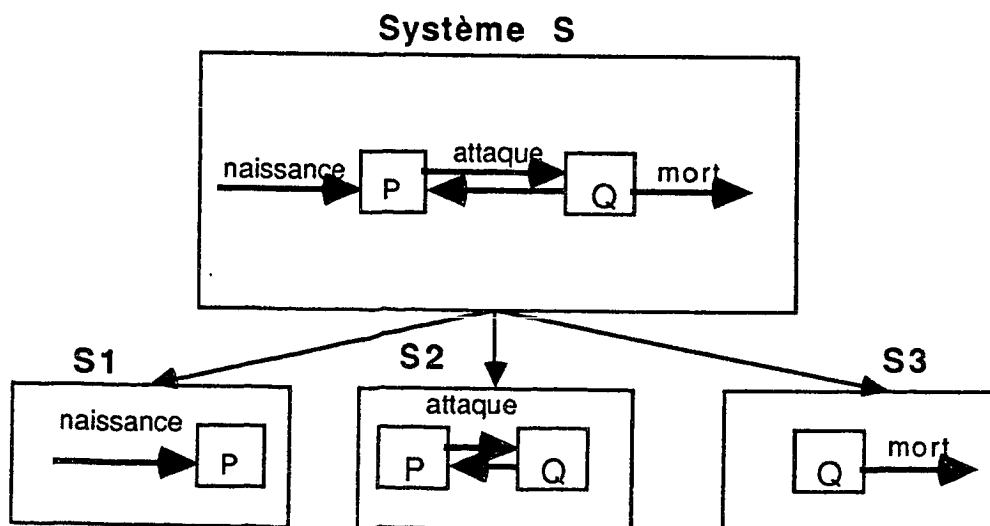


Fig16: décomposition du système S prédateur-proie en trois sous systèmes isolés S1, S2, S3

2. HYPOTHESES ELEMENTAIRES et TEST:

Un raisonnement similaire à celui décrit au paragraphe 1.3.1, basé sur l'ensemble des associations causales décrites précédemment, conduit au choix d'un processus élémentaire (cf base de processus élémentaires §2.1.2.5) ou d'un modèle pour chaque sous-système. Le résultat est donc un ensemble de processus biologiques ou de modèles élémentaires pouvant engendrer l'ensemble des manifestations élémentaires observées sur le sous-système isolé: phénomène élémentaire, courbe élémentaire. Différentes hypothèses peuvent être retenues pour chacun des sous-systèmes. Par exemple, différents modèles (ou processus) peuvent être plausibles pour (S1): modèle exponentiel, modèle logistique etc.... On conduit alors un raisonnement conduisant au choix et à la validation d'un modèle élémentaire telle qu'il a été exposé ci-dessus.

Par exemple,

le modèle exponentiel M_1 : $\frac{dP}{dt} = aP$ (correspondant à un processus de naissance: $P \xrightarrow{a} 2P$ cf. processus élémentaire (p7) page 40) est choisi pour modéliser le système proie (S1).

le modèle exponentiel M_2 : $\frac{dQ}{dt} = -bQ$ (correspondant à un processus de mort: $Q \xrightarrow{b} \emptyset$, cf. processus élémentaire (d7) page 42) est choisi pour modéliser le système prédateur (S2).

le modèle M_3 : $\begin{cases} \frac{dP}{dt} = -cR_2 PQ \\ \frac{dQ}{dt} = dR_2 PQ \end{cases}$ (correspondant à un processus de prédation:

$P + Q \xrightarrow{c} (1+R_1)Q + (1-R_2)P$, processus élémentaire du type (l1) page 49) est choisi pour modéliser le système (S3).

Le choix de ces processus repose sur les hypothèses suivantes:

en l'absence de prédateur, la croissance de la proie peut être assimilée à une croissance naturelle d'une population

tandis qu'en l'absence de proies, la croissance du prédateur peut être assimilée à un processus de mortalité.

Enfin, la situation de prédation est assimilée à la consommation d'une espèce par une autre espèce ce qui conduit au choix d'un processus de consommation.

3. AGREGATION

Après vérification de la cohérence entre les différentes hypothèses (processus élémentaires), l'ensemble de ces hypothèses (processus élémentaires) est agrégé. Un processus est ainsi défini par **combinaison** des processus élémentaires retenus. Il lui est associé, un modèle exprimant le bilan de ces différents processus.

Exemple:

Un nouveau processus \mathcal{P} est défini par **agrégation** des processus élémentaires (p7), (d7), (l1) soit dans le formalisme pseudo-chimique, le processus:

$$\mathcal{P} \begin{cases} P \frac{a}{b} > 2P \\ P+Q \frac{c}{b} > (1+R) Q \\ Q \frac{b}{b} > 0 \end{cases}$$

Il lui correspond le modèle:

$$\begin{cases} \frac{dP}{dt} = aP - cPQ \\ \frac{dQ}{dt} = -bQ + dPQ \end{cases}$$

A ce modèle est associé un ensemble de courbes C d'évolution des variables en fonction du temps, ou un plan de phase Φ .

A ce processus est également associé une situation globale:

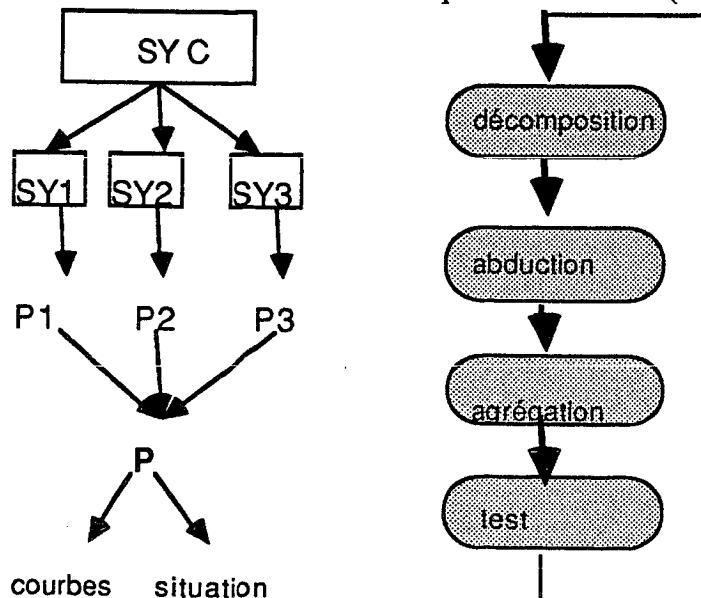
$$\mathcal{S} \begin{cases} \text{s-exponentielle+} \\ \text{s-logistique2} \\ \text{s-exponentielle-} \end{cases}$$

4. TEST

Le test du modèle correspond à la confrontation des propriétés du modèle et des propriétés attendues. On analyse ainsi:
 ce qui fait défaut,
 ce qu'il y a en trop,
 ce qui ne coïncide pas exactement avec la réalité.

Elle consiste essentiellement en une comparaison du comportement qualitatif du modèle et du comportement qualitatif attendu (comparaison des courbes observées aux courbes C ou Φ). On tente alors de modifier soit la structure globale du système, soit les hypothèses faites sur les sous-processus.

Comme on le voit, la construction d'un modèle est un processus itératif (essais-erreurs).



1.4 Conclusion

Le schéma suivant illustre les différentes connaissances impliquées et leurs étroites relations d'interdépendance:

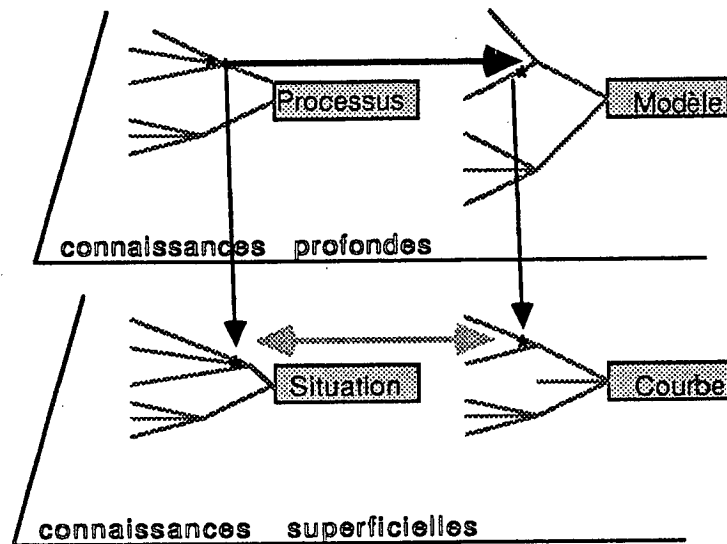


Fig17: Différents niveaux de connaissance.

L'analyse des connaissances (structure et raisonnement) a mis en évidence :

- l'importance de l'aspect prototypique et hiérarchique des connaissances (ce point est explicite au §2).
- l'importance de la notion de point de vue
- l'importance des interrelations entretenues par les différentes connaissances.
- l'importance de la prise en compte dans le raisonnement, de ces différents points de vue et de leurs interrelations.
- la nature abductive du raisonnement de modélisation.

En résumé, il s'agit de connaissances structurées en de multiples hiérarchies qui peuvent être définies de manière indépendantes. Un réseau d'associations causales réunit les différents objets de ces hiérarchies. Cette structuration des connaissances en multiples hiérarchies est le support d'un mécanisme de raisonnement abductif basé sur les associations.

L'un des problèmes que soulève la représentation de ce type de connaissances est l'absence de support théorique. Comme le souligne Reggia J. [REGGIA, 86]:

"association-based abductive models of diagnostic inference, in marked contrast to models using statistical patterns classification and rule-based deduction, do not have a readily identifiable, well-developed theoretical foundation."

Il apparaît cependant clair qu'une représentation à l'aide de règles serait lourde, coûteuse (cf. 1.1.3.2.-2: 15 règles au minimum pour 2 modèles!), peu intelligible et fondamentalement mal adaptée à ce type de connaissances. La connaissance concernant un prototype serait dissiminée à travers toute la base au lieu d'être regroupée sur une seule entité. De plus, des informations de cette

17.05.88

nature, concernant typiquement des associations causales, sont plus volontiers représentées sous forme descriptive comme le permettent par exemple les formalismes centrés objet. Enfin un raisonnement de "choix" basé sur de telles relations causales apparaît bien correspondre à un raisonnement de classification (par filtrage) sur plusieurs hiérarchies d'objets interconnectées.

La partie suivante développe donc l'organisation de la base de connaissances dans le formalisme de représentation centrée-objet choisi: les schémas (SHIRKA). Elle montre l'adéquation de ce modèle 'objet' au type de problème posé:

- représentation de connaissances de nature prototypique et hiérarchique.
- représentation des relations causales entre les différents objets par la définition d'objets de plus haut niveau.
- raisonnement de multi-classification sur cette hiérarchie de plus haut niveau.

2. REPRESENTATION DES CONNAISSANCES

On présente successivement la manière dont on a structuré les connaissances profondes, superficielles et la manière de conduire le raisonnement avec cette organisation des connaissances.

2.1 Représentation des connaissances profondes: modèle, processus

2.1.1 Le formalisme mathématique: la famille *modèle*

La définition de l' objet générique *modèle* est approximativement la suivante:

```
{modèle
  sorte-de      =  objet
  var-indep     $un      var-indep
  état          $liste-de état
  paramètre     $liste-de paramètre
  entrée        $liste-de entrée
  sortie        $liste-de sortie
  équation      $liste-de equa-dif
  cond-ini      $un      atome
  prototype     $un      modèle
  mesures       $un      {chronique
                          intervalle $var-nom inter
                          pas        $var-nom dt}

  simulation    $un      chronique
                $sib-exec
                {GEAR
                  équations $var <- équation
                  intervalle $var <- inter
                  pas        $var <- dt}
  identification $liste-de nombre
                $sib-exec
                {ICSE
                  ... ..
                }
  points d'équilibre $liste-de ... .. point-équilibre}}
```

Cet objet est la racine d'une hiérarchie. Cette classe est partitionnée d'une part, en sous-classes de modèles prédéfinis dans la base, d'autre part une sous-classe particulière permet de construire de nouveaux modèle par composition de sous-modèles:

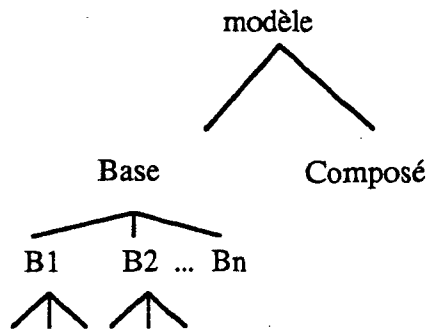


Fig18: Structure de la base de modèles

2.1.1.1 Prototypes de modèle

Une classe de modèles prédéfinie de la base est un prototype de modèle. On définit ainsi un certain nombre de prototypes correspondant aux modèles usuels utilisés en dynamique des populations: les classes modèle logistique, modèle gompertz, modèle monod etc....

Pour caractériser chacune des classes on définit un représentant privilégié dit **instance prototypique**. par exemple, la sous-classe `logistic-forme1` est définie à l'aide de l'instance d'équation différentielle associée à l'équation $\frac{dx}{dt} = ax(1 - \frac{x}{k})$; $x(0) = x_0$

Ces instances prototypiques permettent :

- . de savoir si un modèle introduit par l'utilisateur correspond à un modèle de la base,
- . de créer une instance d'une classe de modèle par "duplication" du prototype,
- . de pouvoir donner un exemple de modèle pour chaque classe.

Ce mode de représentation des classes par un représentant canonique n'a de sens que si on définit une relation d'équivalence entre les modèles, c'est-à-dire si on possède un mécanisme de mise en correspondance (ce qui implique une étape de simplification formelle). Un algorithme de reconnaissance a donc été implémenté. Il utilise la forme pseudo-canonique d'une expression obtenue à la sortie du simplificateur. Les spécifications du simplificateur ont été fixées en liaison étroite avec ce type de fonctionnalités. [c.f. PIERRET 88].

2.1.1.2 Classes et instances

Exemple : schéma de classe du modèle logistique défini à l'aide de son instance prototypique

```
{modèle-logistique
  sorte-de      =      modèle-différentiel}

{logistique-forme1
  sorte-de      =      modèle-logistique
  prototype     $défaut prototype-logistic-1 }
```

Prototype-logistic-1 est une instance prototypique définissant la classe modèle-logistique-forme1 (par définition, elle est instance de la classe logistique-forme1:

```
{prototype-logistic-1
  est-un        =      logistique-forme1
  état          =      $x
  paramètres    =      $a $k
  équations     =      equ-logistic-1 }
```

(equ-logistic-1 est une instance prédéfinie dans la base de connaissance des équations différentielle)

Exemple d'instance de logistique

Soit le modèle logistique défini par

$$\frac{dx}{dt} = 1.1 * x * (1 - \frac{x}{100})$$

$$x(0) = 50$$

où x représente la taille d'un rat musqué mesurée en cm étudiée pendant 36 mois
 t représente le temps, le domaine d'étude est $[0+\infty]$

Ce modèle est une instance particulière de la classe `logistic-forme1`. Il sera défini par l'instance:

```
{exemple-logistique-forme1
  est-un      =      logistique-forme1
  état       =      $taille-rat-musqué
  paramètres =      1. 1 100.
  équations  =      equ-dif1
  cond-ini   =      50}
```

```
{equ-dif1
  est-un      =      equation-dif
  membre-gauch =      diff1
  membre-droit =      expr1}
```

membre de gauche de l'équation différentielle *equ-dif1*:

```
{diff1
  est-un      =      diff
  arguments   =      $taille-rat-musqué $t}
```

membre de droite de *equ-dif1*

```
{expr1
  est-un      =      mult
  arguments   =      1.1 $taille-rat-musqué}
```

```
{expr2
  est-un      =      moins
  arguments   =      1 $expr3}
```

```
{expr3
  est-un      =      div
  arguments   =      $rats-musqués 100. }
```

variables:

```
{$taille-rat-musqué
  est-un      =      taille
  domaine de définition =      0 60
  unité       =      cm
  dimension   =      $L
  mesures     =      longitudinales }
```

```
{$t
  est-un      =      temps
  signifiant  =      temps
  domaine de définition =      R+
  dimension   =      $T
  unité       =      mois }
```

2.1.1.3 Combinaison déclaratif et procédural

Une méthode d'intégration (GEAR par exemple) ou d'identification est définie par défaut pour un modèle en général (cf la définition de modèle). Une méthode particulière peut être redéfinie pour chaque classe de modèle particulière. La méthode RK4 ou EULER peut suffire pour une classe de modèles donnés. Il est même possible d'attacher à un modèle le code explicite correspondant.

Les exemples suivants illustrent comment peut être attachée à chaque modèle, sa méthode spécifique de calcul pour l'identification ou la simulation. Par exemple la méthode d'identification "programme-logistique" sera directement codée sur le modèle logistique, la méthode "programme-gompertz" pour le modèle de Gompertz. Ces méthodes sont activées à la demande du système, uniquement au moment où la valeur des paramètres du modèle est demandée.

```
{modele
  sorte-de      =      objet
  equations     $liste-de equation
  integration   $default GEAR}
```

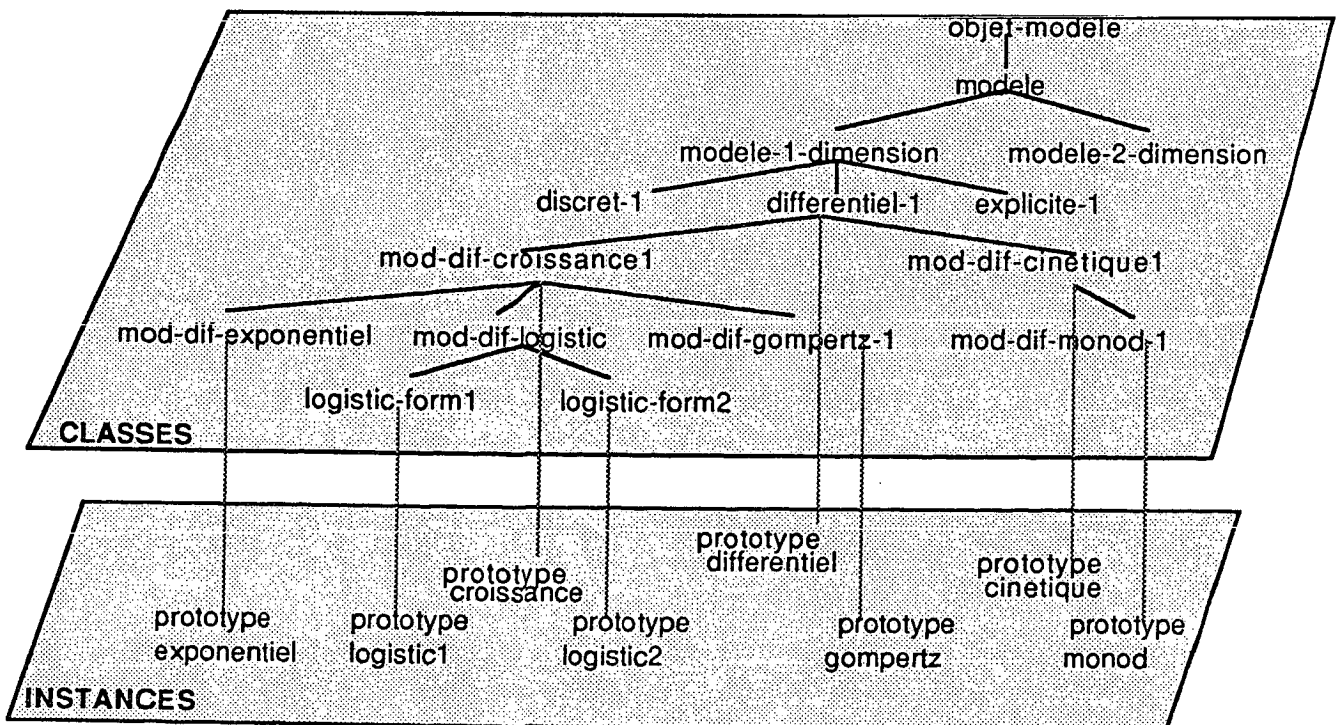
La méthode appropriée à un objet, est ainsi directement codée sur l'objet.

```
{model-logistique
  sorte-de      =      model
  indent       $sib-exec
                {programme-logistique
                  equ $var-list < - equations
                  res $var- >      identification}

{model-gompertz
  sorte-de      =      model
  indent       $sib-exec
                {programme-gompertz
                  equ $var-list < - equations
                  res $var- >      identification}
```

2.1.1.4 Organisation hiérarchique des modèles

Plusieurs points de vue peuvent être choisis pour organiser la base de modèles. La hiérarchie peut être organisée selon le comportement qualitatif des modèles, la méthode d'identification etc... La présentation ci-dessous de la branche objet-model n'est qu'une ébauche d'une telle base. La figure 13 ne montre donc que la méthode de représentation d'une telle hiérarchie de prototypes, par l'utilisation d'instances prototypiques.



2.1.2 Le formalisme biologique: la famille *processus*

2.1.2.1 Définition de l'objet *variable-biologique*

Un objet biologique peut être vu selon plusieurs points de vue:

- du point de vue fonctionnel. Ce peut être un catalyseur, un substrat, un prédateur, une proie etc... On définit donc différentes classes de fonctions, *catalyseur, substrat, prédateur, proie* etc...

On peut noter que dans un même système un objet peut avoir simultanément plusieurs fonctions. Par exemple, soit le système

mulot -----lièvre -----lynx

Le processus général est composé de deux sous-processus de type prédateur-proie. Dans le premier le lièvre a pour fonction prédateur, dans le second il a pour fonction proie.

- du point de vue phylogénétique: il s'agit d'un vertébré, d'une plante etc... Conformément à la sémantique biologique, différentes spécialisations sont proposées: les espèces étudiées peuvent être des instances des classes : *organe, organisme, population* .

On définit donc plusieurs objets génériques: un objet *variable-biologique*, un objet *fonction*.. Chacun de ces objets est la racine d'une hiérarchie spécifique (Fig 14). Un objet de "plus haut niveau" permet de réunir ces différents points de vue sur le même objet: l'objet *entité*. Il a pour attribut une *vraiabie-biologique* et une *fonction biologique*. Les espèces en présence sont des instances de cette classe:

```
{entite
  sorte-de      =      objet;
  variable-biologique  $un  variable-biologique
  fonction-bio   $un  fonction-bio }
```

Par exemple, le lièvre sera représenté pour le premier processus par:

```
{entite#1
  est-un        =      entite;
  variable-biologique  =      lievre#1
  fonction-bio   =      proie }
```

tandis que dans le second processus, il sera représenté par l'instance:

```
{entite#2
  est-un        =      entite;
  variable-biologique  =      lievre#1;
  fonction-bio   =      predateur }
```

où lievre#1 représente une instance de mammifère:

```
{lievret#1
  est-un        =      mammifere;
  taille        =      $taille }
```

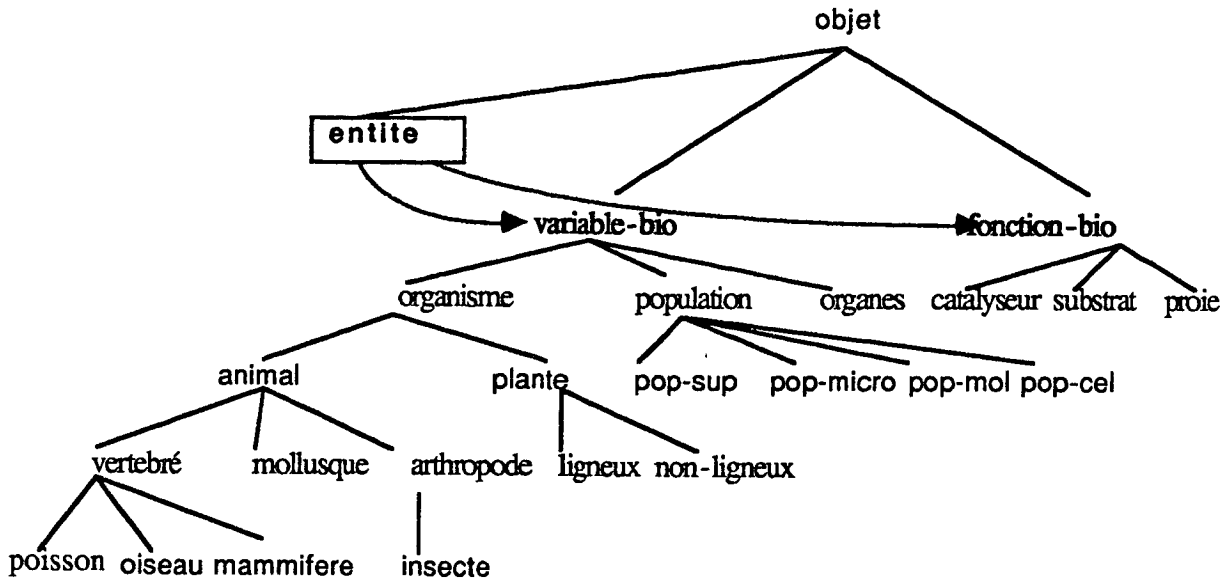


Fig 20: Organisation hiérarchique des variables biologiques et de leurs fonctions

Les variables mathématiques et les variables biologiques sont reliées entre elles par des attributs. La classe *variable-biologique* possède un attribut spécifique: "variables" permettant de préciser le ou les grandeurs quantifiables de la classe faisant l'objet de l'étude. Il s'agit généralement de variables morphométriques décrivant un individu, quantité de produit (pour un substrat), effectif pour une population etc...Par exemple, l'attribut "variables" permet d'associer à un organisme, sa biomasse

Exemples

Variables biologiques

{X
 est-un = **variable-bio**
 signification = cellule
 variables = \$x}

{rat#1
 est-un = **mammifère**
 signification = mulot
 variables = \$p \$t
 habitat = "étang"
 etc..... }

Variables mathématiques

{\$x
 est-un = **longueur**
 signification = diamètre-cell
 unité = μ}

{\$p
 est-un = **poids}**
 {\$t
 est-un = **taille**
 signification = long-patte}

2.1.2.2 Définition de l'objet *réaction*

La classe réaction est caractérisée par trois attributs:

- membre de gauche,
- membre de droite,
- constante de réaction.

Chaque membre a pour valeur une liste de termes chimiques élémentaires.

Exemple

La réaction $X + S \xrightarrow{a} (1 + R)X$ a pour membre de gauche $X+S$, pour membre de droite $(1+R)X$, pour vitesse de réaction a . Le membre de gauche est composé de deux termes: X et S , le membre de droite d'un seul terme $(1+R)X$. Cet exemple correspond à l'instance de réaction :

```
(1) {reaction1
      est-un          = réaction
      mb-gche         = chim1 chim2
      mb-dte          = chim3
      constante       = a}

{chim1
  est-un          = réaction
  espece          = X
  coefficient      = 1}

{chim2
  est-un          = terme-chimique
  espece          = S
  coefficient      = 1}

{chim3
  est-un          = terme-chimique
  espece          = X
  coefficient      = plus 1}
```

où X et S sont des instances d'entité
 a et R sont des instances de paramètres
 plus 1 est une instance d'expression *plus* représentant l'expression $(1+R)$.

2.1.2.3. L' objet *processus*

La classe processus est un objet générique similaire à l'objet modèle.

Un processus est caractérisé par ses entités (variables biologiques) et par des réactions entre ses variables.

Comme on avait distingué *entrée*, *état* et *sortie* pour les variables un modèle, on distingue *variable-étudiée* et *influence* pour un processus. Les premières sont les espèces dont on étudie l'évolution, les seconds sont les facteurs à prendre en considération dans le processus. L'objet processus est ainsi défini par le schéma :

```
{processus
  sorte-de          = objet
  variables-étudiées $liste-de entite
  influences         $liste-de entite
  réaction          $liste-de réaction
  prototype         $un processus}
```

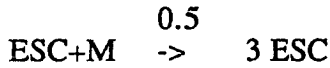
Exemple d'instance de processus:

```
{processus
  sorte-de          = objet
  variables-étudiées = ESC
  influences         = milieu-culture-1
  réaction          = réaction#1}
```

2.1.2.4 Prototypes de processus

Une classe de processus prédéfinie de la base est un prototype de processus.

Par exemple, le processus



est une instance particulière du processus général de croissance par consommation de substrat dont la structure est définie par:



Pour caractériser cette classe on définit un représentant privilégié dit **instance prototypique**.

2.1.2.5 Organisation hiérarchique de la base des processus

A) Présentation:

La classe *processus* est d'une part la racine d'une hiérarchie de prototypes. D'autre part, il est possible de construire de nouveaux processus à l'aide de manipulations symboliques prédéfinies sur l'ensemble des processus, par exemple, *juxtaposition de processus*, ou *ajout de termes*. Un processus ainsi construit par combinaison de sous-processus, est instance d'une sous-classe particulière, *processus-composé*.

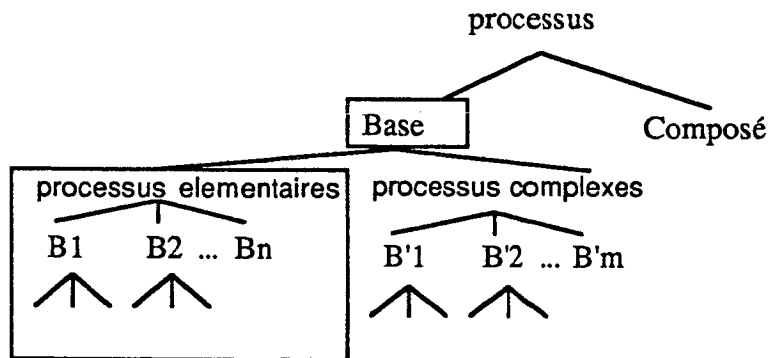


Fig20: Structure de la base de processus

B) La base de prototypes:

Les processus de base B_1, B_2 etc... sont des processus prédéfinis. Il en existe de deux sortes, les

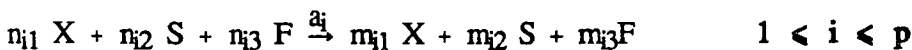
processus simples ou complexes. Les premiers seront appelés processus "élémentaires" (de la même manière que dans l'arbre du calcul formel il y a des opérateurs prédéfinis élémentaires: +, -, *, diff). Les seconds, processus "composites" (complexes) de la base, sont des combinaisons de processus élémentaires. (de la même manière, l'expression $a \cdot x + b$ est une expression prédéfinie dans la base des expressions, obtenue par combinaison des opérateurs élémentaires + et *).

La représentation des différentes classes de processus a nécessité de formaliser précisément les connaissances et ainsi de faire une typologie des processus prédéfinis. Celle-ci montre la nature fondamentalement hiérarchisée des processus. On trouvera ci-dessous une **typologie des processus**. L'organisation des processus élémentaires est présentée en premier, puis celles des processus complexes de la base.

I. PROCESSUS A UNE POPULATION

On considère les processus biologiques mettant en cause au plus trois espèces biologiques X, S, F. Une seule variable est a priori l'objet de l'étude, les deux autres variables S et F sont des variables influençant l'évolution de X. On appellera de tels processus, des processus à une population. Cette restriction à deux facteurs influents est motivée par la clarté de l'exposé mais la typologie proposée peut être aisément généralisée à plusieurs facteurs. On pourra en suivant la même méthodologie organiser la base de processus à deux populations.

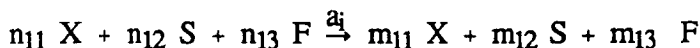
La forme générale d'un tel processus est donc:



On commence par définir les processus élémentaires de la base.

I.1. Définition des processus élémentaires à une population:

Un processus élémentaire est un processus impliquant au plus une réaction:



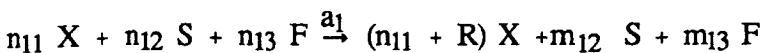
On notera qu'il est clair d'après ces définitions, que tout processus à une population est soit un processus élémentaire, soit un processus obtenu par combinaison de sous-processus élémentaires. Parmi les processus élémentaires, on distingue deux sous classes en fonction du signe de m_{11} : les processus de production ($m_{11} = n_{11} + R$, soit $m_{11} \geq 0$) et les processus de dégradation (m_{11}

$$= n_{11} - R, \text{ soit } m_{11} \leq 0).$$

Processus de production

1. Production de biomasse ou processus de croissance (p1)

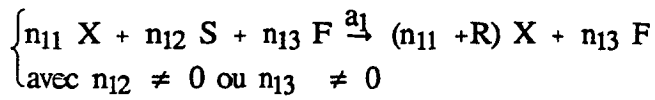
Un tel processus est défini par une réaction de la forme



Les n_{ij} sont des coefficients positifs ou nuls. On distingue donc différents cas d'après la nullité ou non des n_{ij} .

1.1. Production régulée (p2)

Un tel processus correspond au cas $n_{12} \neq 0$ ou (non exclusif) $n_{13} m_{13} \neq 0$. Il est donc défini par une réaction de la forme



S s'appelle un substrat et F un facteur de croissance ou catalyseur

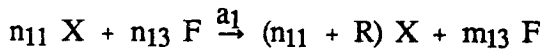
On distingue deux cas de croissance régulée :

$$\boxed{1} \quad n_{12} = 0 \quad n_{13} \neq 0$$

$$\boxed{2} \quad n_{12} \neq 0 \quad n_{13} = 0$$

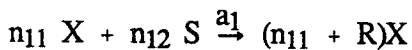
1.1.1. Production régulée par catalyseur (p3)

Il correspond au cas $\boxed{1}$ soit :



1.1.2. Production régulée par consommation de substrat (p4)

Il correspond au cas $\boxed{2}$ soit :

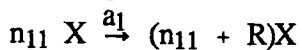


1.1.2.1 Production régulée par saturation (liée à un excès de substrat)



1.2. Production non régulée (p5)

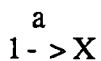
Un tel processus correspond au cas $n_{12} = n_{13} = 0$ soit



On distingue deux cas: $n_{11} = 0$ ou $n_{11} \neq 0$

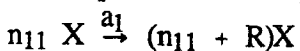
1.2.1 Production non régulée indépendante de la biomasse (p6)

pour $n_{11} = 0$, le processus correspond à une croissance par apport extérieur; on adopte la notation:



1.2.2 production non régulée dépendant de la biomasse (p7)

un tel processus correspond à une réaction de la forme:

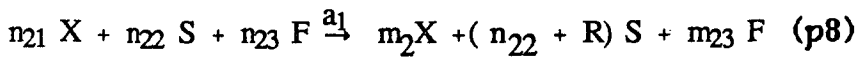


avec $n_{11} \neq 0$

2. Production de substrat ou renouvellement

Remarque: la notion de substrat ou de catalyseur n'a de sens que par rapport à un processus donné, en particulier il faut qu'il existe un processus de croissance ou de décroissance régulée pour que ces notions aient un sens.

Le substrat peut être impliqué dans un autre processus indépendant du premier. On dira de ce processus qu'il est un processus de production ou de renouvellement de substrat s'il est de la forme:

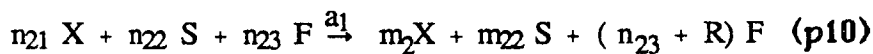


ou apport extérieur:



3. Production de facteur:

La définition est similaire:



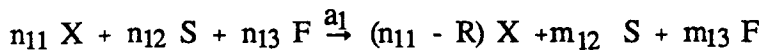
ou apport extérieur:



Processus de dégradation ou mortalité

1. Dégradation de biomasse ou mortalité (d1)

Un tel processus est défini par une relation de la forme



où les n_{ij} sont des coefficients positifs ou nuls. On distingue donc comme précédemment: différents cas. Les définitions sont similaires.

1.1 décroissance régulée (d2)

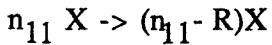
1.1.1 décroissance par consommation de substrat (substance toxique) (d3)

1.1.2 décroissance par catalyse (d4)

1.2 décroissance non régulée (d5) ou mortalité spontanée de la biomasse

1.2.1 décroissance non régulée dépendant de la biomasse (d6)

C'est un processus défini par une réaction de la forme:

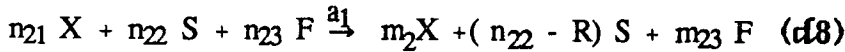


1.2.1 décroissance non régulée indépendante de la biomasse (d7)

C'est un processus défini par une réaction de la forme:



2. Dégradation de substrat

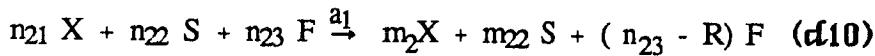


ou mortalité spontanée



3. Dégradation du facteur de croissance

La définition est similaire:



ou

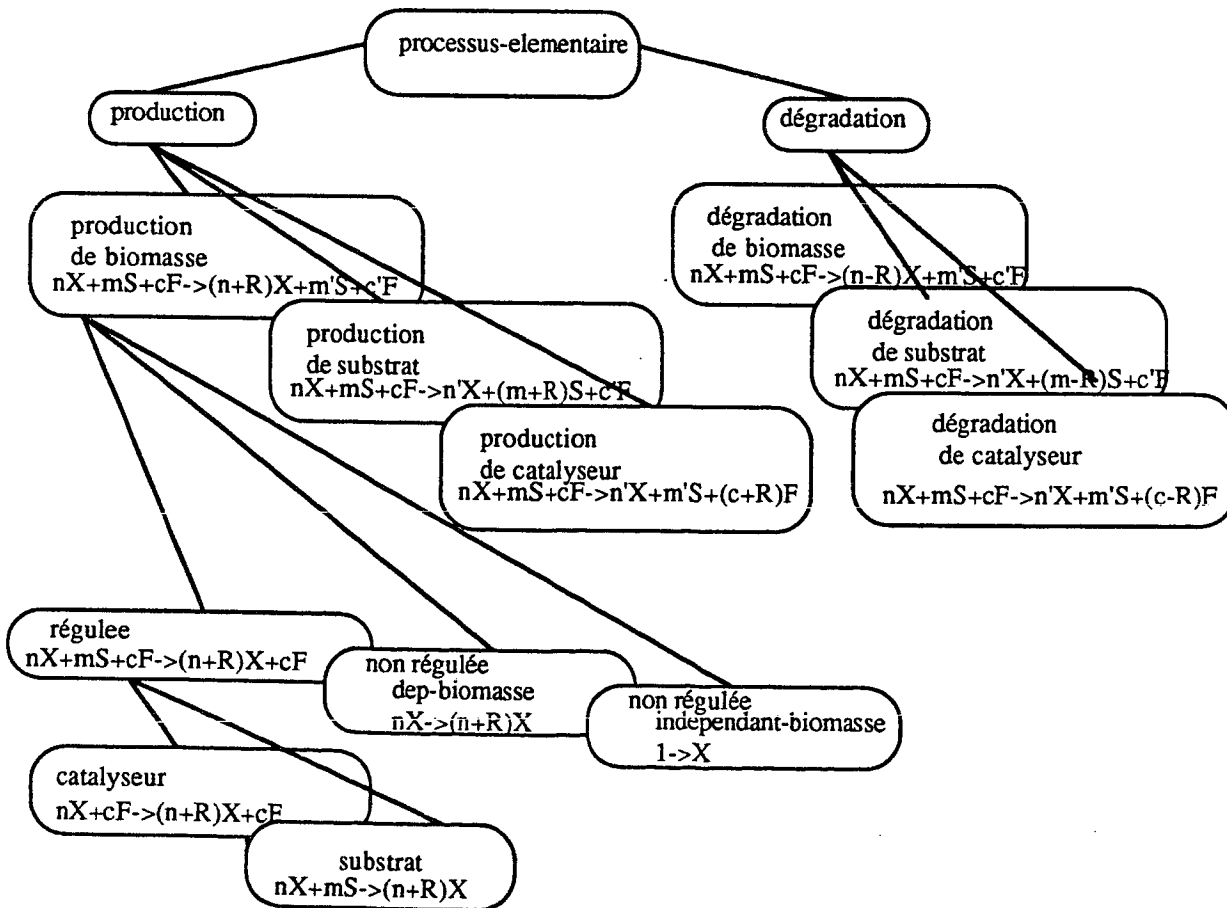


Fig 21 Organisation hiérarchique des processus élémentaires

La combinaison de ces différents processus élémentaires permet de définir des processus "complexes". Parmi ces processus complexes, un certain nombre sont des processus classiques en dynamique des populations. Une typologie de ces processus est proposée en se basant sur la nature des processus élémentaires qui les composent.

I.2. Définition de processus complexes prototypiques à une population:

On appelle processus complexe, un processus prototypique de la base, composé d'un ou plusieurs processus élémentaires précédemment définis:

C1. Processus de croissance :

c'est un processus composé d'au moins un processus de production de biomasse c'est à dire d'un processus du type (p1). Il peut être composé d'autres processus de type quelconque excepté (p1).

On fait en effet l'hypothèse que (p1) "globalise" la production de biomasse. Il est donc défini par l'ensemble des processus:

$$\begin{cases} n_{11}X + n_{12} S + n_{13} F \xrightarrow{a_1} (n_{11} + R) X + m_{12} S + m_{13} F \\ n_{i1} X + n_{i2} S + n_{i3} F \xrightarrow{a_i} n_{i1}X + m_{i2} S + m_{i3}F \quad 2 \leq i \leq p \end{cases}$$

C1.1 Processus de croissance régulée par au moins un facteur

c'est un processus composé d'au moins un processus de production régulée de biomasse c'est à dire d'un processus du type (p2). Elle est donc composée de processus du type:

$$\begin{cases} n_{11} X + n_{12} S + n_{13} F \xrightarrow{a_1} (n_{11} + R) X + n_{13} F & (p2) \\ \text{avec } n_{12} \neq 0 \text{ ou } n_{13} \neq 0 \\ \text{et} \\ n_{i1} X + n_{i2} S + n_{i3} F \xrightarrow{a_i} n_{i1}X + m_{i2} S + m_{i3}F & 2 \leq i \leq p \end{cases}$$

On devrait distinguer à ce niveau les cas 1 ou 2 facteurs. (régulation unique, le cas régulation multiple est volontairement omis pour la clarté de l'exposé).

C1.1.1 Croissance par consommation de substrat

c'est un processus composé d'au moins: un processus (p4)

C1.1.1.1 Croissance par consommation d'un substrat renouvelé

c'est un processus composé d'au moins: un processus (p4)
et d'un processus de type (p8) ou (p9)

C1.1.1.1.1 Croissance par consommation de substrat renouvelé non accumulable:

c'est un processus composé d'au moins: un processus (p4)
et d'un processus de type (p8) ou (p9)
et d'un processus de type (d8) ou (d9)

C1.1.1.1.2 Croissance par consommation de substrat renouvelé non accumulé:

c'est un processus composé d'au moins: un processus (p4)
et d'un processus de type (p8) ou (p9)
mais pas de processus de type (d8) ou
(d9)

C1.1.1.2 Croissance par consommation de substrat non renouvelé

c'est un processus composé d'au moins: un processus (p4)
et ne comportant aucun processus de type
(p8) ou (p9)

C1.1.1.3 Croissance par consommation de substrat dégradé

c'est un processus composé d'au moins un processus (p4) et d'un processus de type (d8) ou
(d9)

C1.1.1.4 Croissance par consommation de substrat non dégradé

c'est un processus composé d'au moins un processus (p4) et d'aucun processus de type (d8) ou
(d9)

C1.1.2 Croissance régulée par catalyseur

c'est un processus composé d'au moins un processus (p3).
On distingue de la même manière selon le type des autres relations impliquées dans le processus:

C1.1.2.1 Croissance par catalyseur renouvelé (p3) et (p10 ou p11)

C1.1.2.2 Croissance par catalyseur non renouvelé: p3 et pas (p10 ou p11)

C1.1.2.3 Croissance par catalyseur dégradé (p3) et (d10 ou d11)

C1.1.2.4 Croissance par catalyseur non dégradé (p3) et pas (d10 ou d11)

C1.2 Croissance autorégulée

processus composé au moins des processus: (p1)
et (d1)

C1.3 Croissance non régulée

C1.3.1 Croissance non régulée indépendante de la biomasse

processus composé d'au moins un processus (p6)

C1.3.2 Croissance non régulée dépendant de la biomasse

processus composé d'au moins un processus (p7)

Cette liste n'est pas exhaustive, on peut définir d'autres classes par combinaison des processus élémentaires, par exemple:

Croissance avec mortalité spontanée de la biomasse: (p1) et (d6) ou (d7).

Etant donné le nombre important de combinaisons possibles (1336), on se limite aux processus usuels de dynamique des populations. Il est entendu que tout processus ne figurant pas dans la base, peut être obtenu par combinaison de ces processus élémentaires. Il peut donc être représenté comme sous-classe de la classe *processus-composé*.

On donnera ci-dessous quelques exemples parmi les processus les plus fréquemment rencontrés en dynamique des populations et donc définis comme prototypes de la base:

P. monomoléculaire1:

| | |
|--------|--------------|
| 1- > S | de type (p9) |
| S- > X | de type (p4) |
| X- > 0 | de type (d1) |

Il s'agit donc d'un processus de croissance régulée par consommation de substrat renouvelé, accumulable avec dégradation spontanée de la biomasse. C'est donc une sous-classe de

P. monomoléculaire2:

| | |
|--------|--------------|
| 1- > X | de type (p6) |
| X- > 0 | de type (d1) |

Il s'agit donc d'une sous-classe de croissance indépendant de la biomasse, non régulée (C1.1.1.1.2), avec mortalité.

P. Monod

$$X + \frac{S}{K} \rightarrow (1+R) X \quad \text{de type (p4')}$$

c'est une sous-classe de la classe des processus de croissance avec saturation

P. logistique1

$$X \rightarrow 2X$$

c'est une sous-classe de la classe ? des processus de croissance autorégulée

17.05.88

P. logistique2

$$X+S \rightarrow (1+R) X \quad \text{de type (p4')}$$

c'est une sous-classe de la classe (C.1.1.1) des processus de croissance par consommation de substrat

P. logistique3.1

$$2X+ F \rightarrow (2+R)X + F \quad \text{de type (p3)}$$

$$F \rightarrow 0 \quad \text{de type (d11)}$$

C'est une sous-classe de la classe (C.1.1.2.4) des processus de croissance régulée par catalyseur dégradé.

P. logistique3.2

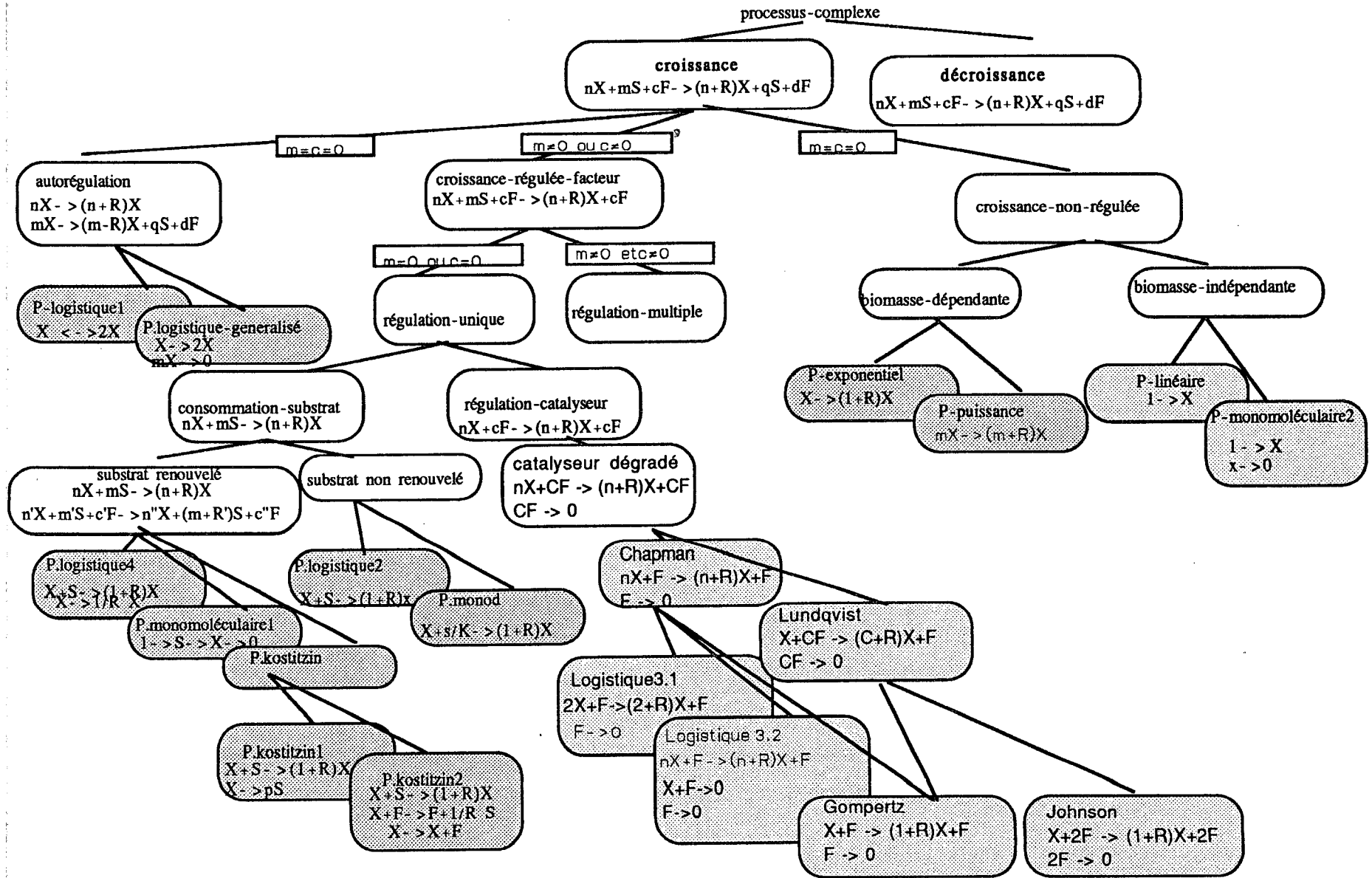
$$nX+ F \rightarrow (n+R)X + F \quad \text{de type (p3)}$$

$$X+F \rightarrow 2F + X \quad \text{de type (p10)}$$

$$F \rightarrow 0 \quad \text{de type (d11)}$$

C'est une sous-classe de la classe (C.1.1.2.4) des processus de croissance régulée par catalyse avec dégradation du facteur (il y a en plus production du facteur par la biomasse).

2.1.2.5 Approche d'organisation hiérarchique d'une base de processus à une population

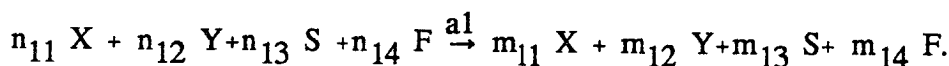


II. PROCESSUS A DEUX POPULATIONS.

Il s'agit de processus biologiques concernant deux espèces biologiques X et Y en interaction: (relations interspécifiques entre deux populations). Ces deux variables font l'objet de l'étude, d'autres espèces S et F peuvent être impliquées dans le processus mais il ne s'agit que de variables influençant l'évolution de X (elles jouent un rôle similaire à celui des paramètres ou du contrôle dans un modèle).

II.1 Processus élémentaires à deux populations.

Un processus élémentaire à deux populations est un processus impliquant au plus une réaction du type:

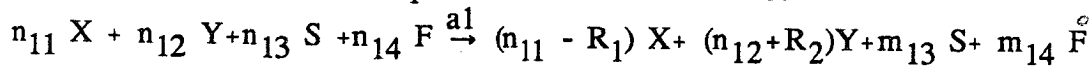


On distingue selon les signes de m_{11} et m_{12} , trois types de processus élémentaires:

| | |
|------------------------------------|--------------------------------|
| $m_{11} \geq 0$ et $m_{12} \leq 0$ | interaction de type prédatif |
| $m_{11} \leq 0$ et $m_{12} \leq 0$ | interaction de type compétitif |
| $m_{11} \geq 0$ et $m_{12} \geq 0$ | interaction de type mutualiste |

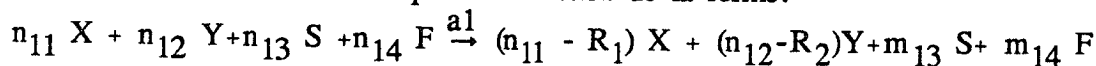
Interaction prédatrice (i1)

Un tel processus est donc défini par une réaction de la forme:



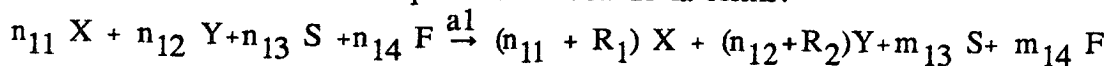
Interaction compétitive (i2)

Un tel processus est donc défini par une réaction de la forme:



Interaction mutualiste (i3)

Un tel processus est donc défini par une réaction de la forme:



Une hiérarchie de ces processus peut être construite en suivant la même démarche que celle suivie en I.1 pour les processus élémentaires à une population.

II.1 Processus composites prototypiques à deux populations.

Un processus composite à deux populations est un processus prototypique (prédéfini) obtenu par combinaison d' au moins:

- un processus élémentaire prototypique à deux populations

ou

-de deux processus élémentaires prototypiques à une population concernant respectivement des populations distinctes.

On remarquera dans cette organisation, le rôle essentiel des processus élémentaires. On aurait pu se limiter d'ailleurs à ne représenter que ces processus primitifs puisque tout processus composite peut être obtenu à partir de ces processus. Il a paru toutefois important de représenter dans la base de prototypes un certain nombre de processus "complexes" classiques en biologie tant dans le cadre des processus à une population: processus de consommation, de catalyse etc....que dans le cadre des processus à plusieurs populations: prédation avec auto-limitation, sans auto-limitation, avec immigration, avec seuil pour les proies, compétition directe ou indirecte etc... Ce point ne sera pas exposé plus avant ici.

C) Les opérateurs prédéfinis:

Certains opérateurs ont été définis de façon à pouvoir procéder à des opérations élémentaires sur les processus:

- L'opérateur juxtaposition (+)

Soient les processus P_i définis par:

| | | | |
|---------|--------------------|---|-----------|
| $\{P_i$ | est-un | = | processus |
| | variables-étudiées | = | X_{ij} |
| | influences | = | G_j |
| | réaction | = | R_{ij} |

alors $P = \sum_j P_i$ est défini par

| | | | |
|-------|--------------------|---|---------------|
| $\{P$ | est-un | = | processus |
| | variables-étudiées | = | $\cup X_j$ |
| | influences | = | $\cup G_j$ |
| | réaction | = | $\cup R_{ij}$ |

Exemple:

{ P₁

| | | |
|-------------------|---|-------------------|
| est-un | = | processus |
| variable-étudiées | = | X |
| influences | = | S |
| reactions | = | X+S -a-> (1+R) X} |

{ P₂

| | | |
|-------------------|---|-------------------|
| est-un | = | processus |
| variable-étudiées | = | Y |
| influences | = | S |
| reactions | = | Y+S -a-> (1+R) Y} |

Si $P = P_1 + P_2$ alors P est défini par

{ P

| | | |
|-------------------|---|---------------------------------------|
| est-un | = | processus |
| variable-étudiées | = | X Y |
| influences | = | S |
| reactions | = | X+S -a-> (1+R) Y+S -a-> (1+R) Y} |

- L'opérateur ajout-terme(@)

Exemple

{ P₁

| | | |
|-------------------|---|--------------------------------|
| est-un | = | processus |
| variable-étudiées | = | X |
| influences | = | S ₁ |
| reactions | = | X+S ₁ -a-> (1+R) X} |

{ S₂

| | | |
|----------|---|------------|
| est-un | = | entité |
| fonction | = | influence} |

S₂@P₁ = P est défini par

{ P₁

| | | |
|-------------------|---|--|
| est-un | = | processus |
| variable-étudiées | = | X |
| influences | = | S ₁ S ₂ |
| reactions | = | X+S ₁ +S ₂ -a-> (1+R) X} |

$P_1 @ S_2 = P$ est défini par

{ P_1

| | | |
|-------------------|---|--------------------------|
| est-un | = | processus |
| variable-étudiées | = | X |
| influences | = | $S_1 S_2$ |
| reactions | = | $X+S_1 -a-> (1+R) X+S_2$ |

Les processus et les opérateurs primitifs qui ont été définis, jouent un rôle essentiel pour la construction de nouveaux modèles.

2.2 Représentation des connaissances superficielles: courbe, situation

Il existe deux grandes familles d'informations qualitatives, les informations concernant la courbe de réponse et celles concernant la situation biologique.

2.2.1 Courbes de réponse

Les classes ont été définies en prenant en compte les caractéristiques habituellement utilisées par le biologiste pour identifier une courbe par exemple:

- croissance ou décroissance
- existence ou non d'une asymptote
- nature de la concavité de la courbe: tournée vers les $y>0$, $y<0$, nulle ou alternée ce qui permet l'inférence de l'existence ou non d'un point d'inflexion
- position du point d'inflexion: partie inférieure, milieu, ou partie supérieure de la croissance etc... (pour plus de détails se référer à la définition des schémas de ces classes)

Exemple de classe:

```
{sigmoïde-avec-asymptote
  sorte-de           = asymptote-avec-inflexion;
  nb-pts-inflexion  $domaine 1;
  position-pt-inflexion $un symbole
                    $domaine milieu inf sup}
```

Exemple d'instance:

```
{ courbe1
  sorte-de           = courbe;
  optimum           = non ;
  croissante        = vrai;
  asymptote         = vrai;
  inflexion         = vrai;
  nb-inflexion      = 1;
  position-inflexion = milieu}
```

L'organisation hiérarchique proposée pour les courbes de réponse est la suivante

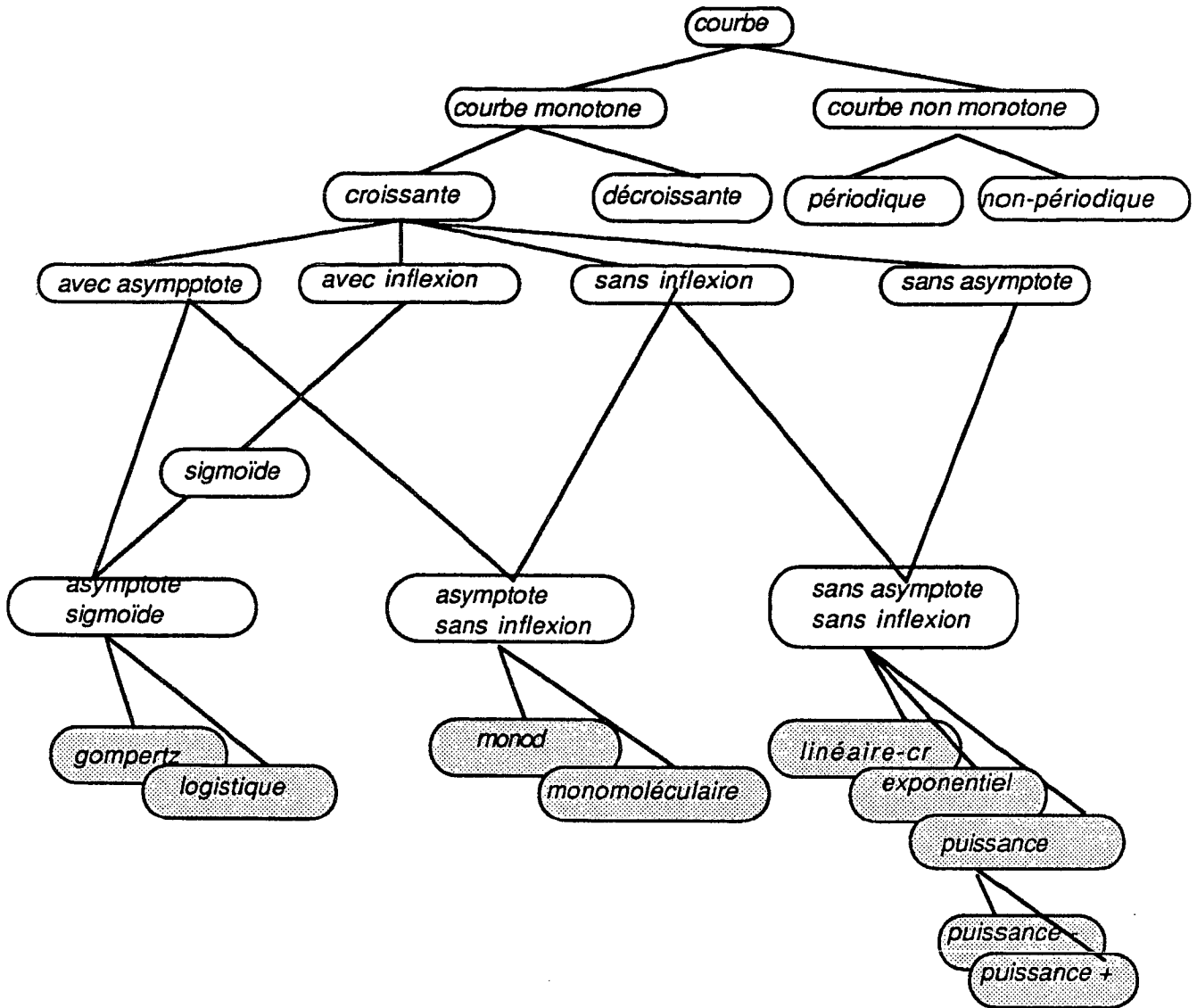


Fig 22: Hiérarchie des courbes de réponse

Le choix par le biologiste d'une courbe dans un catalogue (§ 1.1.2.) correspond à un raisonnement de classification d'une instance dans cette hiérarchie.

Une instance de courbe est donc d'abord créée:

```

Classif:
cr-inst courbe-de-reponse - cr2
  Classe: courbe-de-reponse
  optimum? -
  -> cr2
  
```

Elle est ensuite classée dans la hiérarchie à l'aide d'un algorithme de classification. L'algorithme de classification demande les valeurs d'attributs de l'instance à classer et parcourt la hiérarchie pour identifier les classes sûres, possibles, ou impossibles. Les classes "sûres" sont les classes avec lesquelles l'instance est totalement compatible. Les classes "possibles" sont les classes avec lesquelles l'instance n'est pas incompatible, mais pour lesquelles l'information n'est pas suffisante pour déterminer la totale compatibilité ou l'incompatibilité. Les classes "impossibles" sont les classes incompatibles avec les valeurs des attributs de l'instance. On notera que les classes dites "sûres" ne

sont en fait que des classes **plausibles** (en raison du raisonnement fondamentalement abductif de l'algorithme de classification).

Les principes de cet algorithme sont développés dans [PIVOT, 87]. On en rapportera très rapidement certaines caractéristiques ici. Il est pour cela d'abord nécessaire de préciser quelques termes:

père-de:

C est père-de $C_k \iff C_k$ est sorte-de C

spécialiser:

spécialiser une instance I rattachée à une classe C , dans une classe C -cible revient à se prononcer sur le caractère "sûr", "possible", "impossible" de I pour C -cible.

Une condition nécessaire pour que I puisse être instance de C -cible est qu'elle puisse être instance de ses pères. Cette propriété est donc vérifiée sur les pères par un appel récursif de la spécialisation vers tous es C_k , pères de C -cible (remontée récursive dans le treillis, en profondeur d'abord).

On vérifie que I satisfait également les restrictions propres à C -cible, soit:

$$I \text{ instance de } C\text{-cible} \iff \begin{cases} I \text{ instance des } C_k, \text{ pour tout père } C_k \text{ de } C\text{-cible} \\ \text{et} \\ I \text{ vérifie les restrictions propres de } C\text{-cible} \end{cases}$$

La classification d'une instance I dans un groupe C consiste à un parcours par niveaux descendants de toutes les classes C_i , du groupe ou de la famille privé des groupes C_k déterminés "impossibles", avec appel de la spécialisation de I dans C_i .

On notera que chaque fois qu'une spécialisation de I vers C_k est finie une explication est associée à C_k pour I .

Le dialogue est donc du type suivant:

Classif:

class-inst cr2

classe??

nom de la classe origine a partir de laquelle on veut classer l'instance: un symbole

classe? *courbe*

guide??

mode guide si vrai (par défaut): un boolean

guide? -

avec-inference? *faux*

optimum? *neant*

croissante? *vrai*

asymptote-sup? *vrai*

concavite? *alternee*

nb-pts-inflexion? *1*

symetrique/pt-inflexion? *faux*

Les classes sûres:

-> courbe courbe-monotone croissante
cr-avec-pt-inflexion asymptote-sigmoïde gompertz

Les classes impossibles:

-> courbe-non-monotone décroissante
asymptote-sans-inflexion sans-asymptote-sans-inflexion logistique

Ici par exemple, la courbe "gompertz" a été retenue tandis que la courbe "logistique" a été exclue. Toutes les deux satisfont en effet les caractéristiques d'une sigmoïde avec asymptote mais la valeur *faux* de l'attribut symétrique/pt-inflexion a permis d'exclure la logistique.

On trouvera en annexe des exemples de sessions concernant la classification et d'explication.

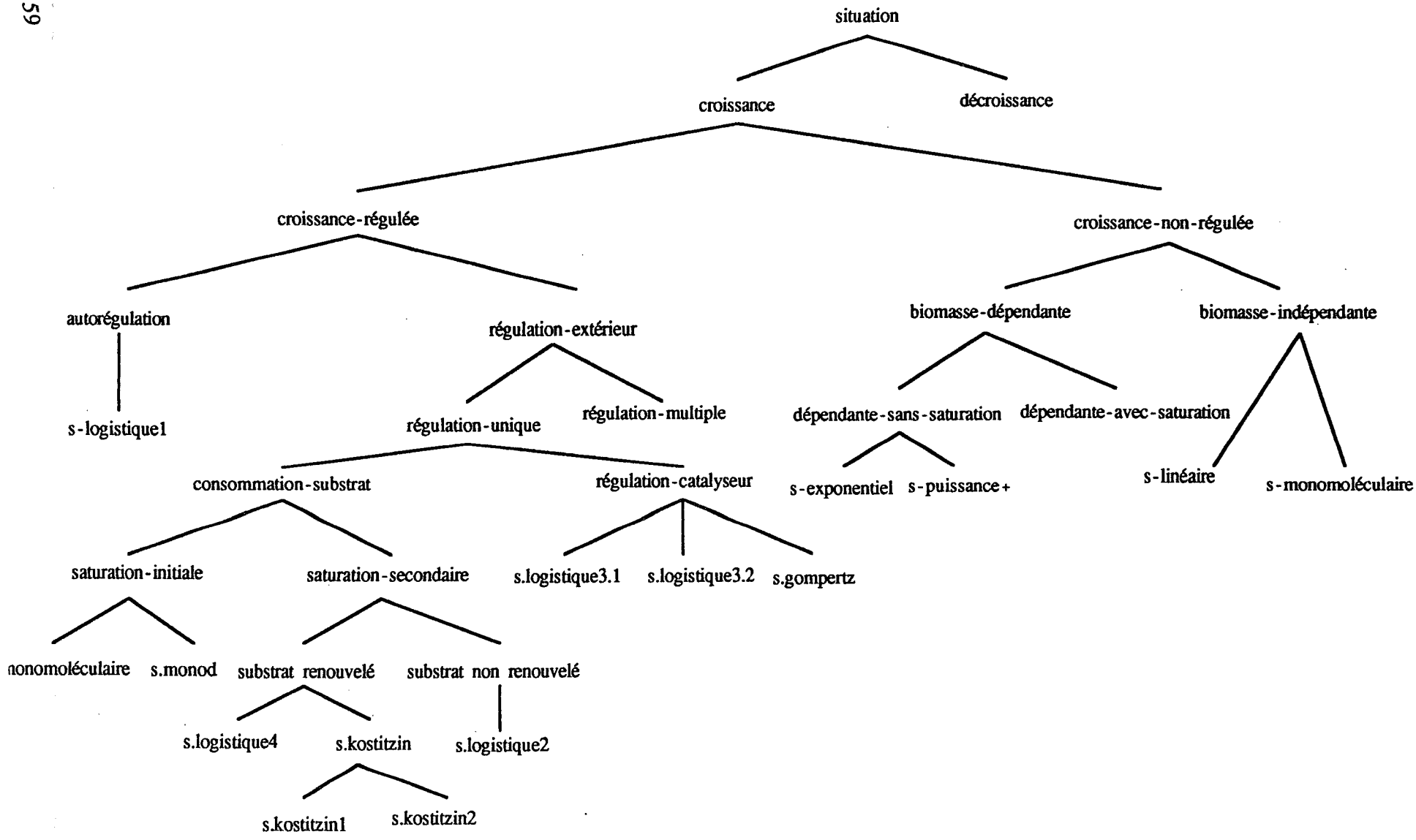
2.1.2 Situation

Les classes ont été définies en prenant en compte les caractéristiques habituellement utilisées par les biologistes pour identifier une situation.

Exemple:

```
{croissance-regulee
  sorte-de           = croissance;
  regulee            $domaine vrai;
  nb-facteur-limitant $un entier
                    $com "nombre de facteurs limitant la croissance";
  saturation         $domaine vrai;
  nature-saturation $un symbole
                    $domaine initiale lente moyenne brutale
  $com " initiale: saturation initiale i.e. d'emblée dès le début de la croissance"
      " lente: saturation n'apparaît qu'après un certain temps mais en début de
        croissance"
      " moyenne pareil, mais en milieu de croissance"
      " brutale saturation observée brutalement en fin de croissance "}
```

L'organisation hiérarchique proposée pour les situations est la suivante (Fig 17: Hiérarchie des situations concernant l'étude d'une population)



17.05.88

Le raisonnement du biologiste correspond au choix d'une situation dans un catalogue de situations prototypiques qu'il a mémorisées (§1.1.3.2). Il correspond donc bien à un raisonnement de classification d'une instance dans cette hiérarchie.

Le mode d'utilisation du système est le même que celui décrit plus haut:

Une instance de situation est d'abord créée:

Classif:

cr-inst situation - sit2

Classe: situation

croissance? -

-> sit2

L'algorithme de classification demande à nouveau ici les valeurs d'attributs et parcourt la hiérarchie pour identifier les classe sûres, possibles, ou impossibles. Le dialogue est du type suivant:

Classif:

class-inst cr2

classe? ?

nom de la classe origine a partir de laquelle on veut classer l'instance: un
symbole

classe? situation

guide? ?

mode guide si vrai (par default): un boolean

guide? -

avec-inference? faux

croissante? vrai

régulée? vrai

nb-facteur-limitant? 1

nature-facteur? substrat

renouvelé? vrai

dégradé? faux

Les classes sûres:

-> situation croissance croissance-régulée regulation-unique

consommation-substrat substrat-renouvelé s-logistique4

Les classes impossibles:

-> décroissance

croissance-non-régulée regulation-multiple

regulation-catalyseur s-gompertz substrat-non-renouvelé s-logistique2

s-monod

substrat-renouvelé-dégradé s-monomoléculaire

Ici par exemple, la situation "gompertz" a été exclue tandis que la situation "logistique" a été retenue. Toutes les deux satisfont en effet les caractéristiques d'une croissance régulée mais la nature catalytique du facteur a permis d'exclure la logistique. La distinction entre situation monomoléculaire et logistique n'est pas évidente: il s'agit toutes deux de croissance par consommation d'un substrat renouvelé. La valeur de l'attribut dégradé permet toutefois de ne retenir que la situation logistique

2.3 Coexistence des différents points de vue et des différents niveaux de connaissance

L'analyse des connaissances (§1.2) a permis de montrer que les connaissances étaient structurées en de multiples hiérarchies dont les noeuds étaient reliés par un réseau d'associations causales. On a mis en évidence qu'un modèle prototypique était ainsi associé d'une part à une courbe prototypique, d'autre part à toute une famille de processus et de situations prototypiques (exemple du modèle logistique §1.2.4). On a également montré au §1.2.3 que les connaissances superficielles étaient fortement interdépendantes (par l'intermédiaire des connaissances profondes: structure et mécanismes qui régissent un système). Il s'agit donc de représenter et d'assurer la cohérence de ces différentes de connaissances.

L'analyse du raisonnement (§1.3) concernant la formulation d'un modèle a mis en lumière qu'il était déterminant de prendre en compte les différentes connaissances car négliger un des points de vue pouvait conduire à un modèle inapproprié.

On notera qu'on retrouve une problématique analogue dans les systèmes de diagnostic médical développés ces dix dernières années. La prise en compte de connaissances causales dans le raisonnement de diagnostic est apparue comme une nécessité [PATIL, 87], [PATIL, 86]. Les maladies peuvent être organisées selon plusieurs dimensions (*étiologique, anatomique, pathophysiologique*, qui ont pour conséquences différents degrés (*severité*) de dysfonctionnements, perçus à travers divers *signes* ou *symptômes*) et comme le souligne Patil: "... It is often possible, even desirable in the early phase of diagnosis, to characterize a clinical problem along these dimensions..... Les connaissances causales sont représentées de différentes manières selon les systèmes (réseau d'états pathophysiologiques dans CASNET, attributs *causé-par, cause-de, complication-de* de frames dans PIP etc... La base de connaissances de CADUCEUS [POPLE, 82] est très proche de l'esprit dans lequel la base a été structurée ici: organisation en multiple hiérarchies, les maladies de ces hiérarchies sont reliées par des relations causales elles-mêmes organisées hiérarchiquement. Connaissances profondes et connaissances superficielles sont utilisées conjointement dans le but d'obtenir un meilleur diagnostic.

Bien que l'objectif soit différent ici, puisque les connaissances profondes sont recherchées, on est confronté à un problème similaire, celui de la représentation des relations de causalité entre ces différentes connaissances.

L'idée ici est d'assurer la consistance des différentes connaissances prototypiques par la définition d'un nouvel objet, l'objet "système" et d'une hiérarchie d'objets construite à partir de cette racine. L'objet "système", peut être considéré comme un objet de "plus haut niveau" permettant d'agrèger les connaissances cohérentes. Chaque sous-classe est en quelque sorte un "monde de connaissances cohérentes".

Le raisonnement est réalisé par un mécanisme de classification sur cette hiérarchie des systèmes.

2.3.1 L'objet système

L'objet *système* synthétise le réseau d'associations causales présenté au §1. L'objet *système* a quatre attributs: courbe, situation, processus, modèle:

```
{système
  sorte-de      =   objet
  courbe        $un courbe
  situation     $un situation
  modèle        $un modèle
  processus     $un processus}
```

2.3.2 Hiérarchie des systèmes

Cet objet est la racine d'une famille de sous-classes. Deux branches sont issues de cette racine.

2.3.2.1 Systèmes prédéfinis

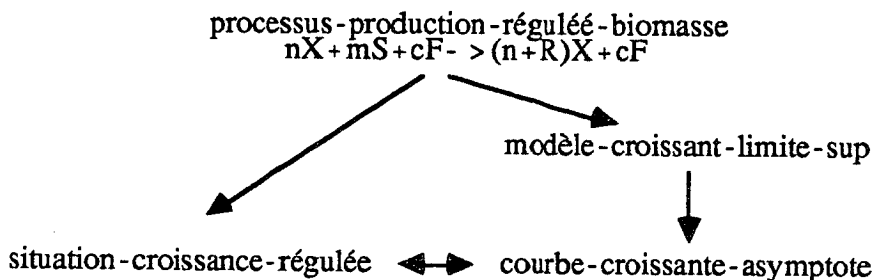
La première branche correspond à des systèmes prototypiques prédéfinis. Les connaissances sont ainsi organisées en "paquets" de connaissances. A partir des quatre hiérarchies précédentes une nouvelle hiérarchie est construite. L'arbre obtenu pour les systèmes est moins hiérarchisé puisqu'il organise les connaissances selon différents aspects simultanément.

Exemple de sous-système:

-Un exemple de système "élémentaire" est présenté, celui du *système-croissance-réglée*:

```
{système-croissance-réglée
  sorte-de = système;
  courbe   $domaine c-croissante-avec-asymptote;
  situation $domaine s-croissance-réglée;
  modèle   $domaine m-croissance-limite-sup;
  processus $domaine p-production-biomasse-regulee}
```

Cette classe *système-croissance-réglée* est une représentation du réseau d'associations



- système-logistique:

On peut se rapporter à l'exemple du §1.2.3 où le réseau d'associations est explicité. Le sous-système logistique est défini par

```
{système-logistique
  sorte-de = système;
  courbe   $domaine c-logistique;
  situation $domaine s-logistique1 s-logistique2 s-logistique3.1 s-logistique3.2 s-
  s-logistique4;
  modèle   $domaine modèle-logistique;
  processus $domaine p-logistique1 p-logistique2 p-logistique3.1 p-logistique3.2}
  p-logistique4}
```

Les différentes classes c-logistique, s-logistique1 s-logistique2 s-logistique3.1 s-logistique3.2, s-logistique4, modèle-logistique, p-logistique1 p-logistique2 p-logistique3.1 p-logistique3.2, p-logistique4 sont définies dans leur hiérarchie respective. (§2.1 et §2.2).

- système-linéaire:

```
{système-linéaire-croissant
  sorte-de = système-croissance;
  courbe   $domaine c-linéaire-cr;
  situation $domaine s-linéaire-cr;
  modèle   $domaine m-linéaire-cr;
  processus $domaine p-linéaire-cr}
```

2.3.2.2 Systèmes construits

La seconde branche correspond aux systèmes construits par combinaison de sous-systèmes prédéfinis. Elle permet de représenter des systèmes complexes n'existant pas dans la base de connaissance.

On définit une opération dite de "juxtaposition" de systèmes (notée +)

$$\text{Soient les systèmes } S_{y_i} \left\{ \begin{array}{l} C_i \\ S_i \\ P_i \\ M_i \end{array} \right.$$

La juxtaposition de plusieurs sous-systèmes est l'opération qui aux sous systèmes S_i associe le

$$\text{système } S_y \left\{ \begin{array}{l} C \\ S \\ P \\ M \end{array} \right. \text{ défini par : } P = \Sigma P_i$$

M est obtenu par traduction de P, C est la courbe associée à M, S est la situation associée à P.

2.3.3 Raisonnement

La représentation et la structuration des connaissances proposée permet de répondre à des questions qui seraient plus difficiles à résoudre dans d'autres formalismes de représentation:

- inférence d'un objet après multi-classification
- validation des faits
- informations incomplètes
- cohérence des connaissances

L'objet système, en réunissant les différentes connaissances sur un même objet permet en effet:

- l'inférence d'un objet par multi-classification:

La structuration des connaissances en treillis distincts, contrôlés par un treillis de "plus haut niveau" est le support d'un mécanisme de raisonnement original. Un algorithme de "multi-classification" permet en effet d'identifier la position d'une instance de la classe système à partir de la position de chacun des objets qui la caractérisent dans sa propre hiérarchie.

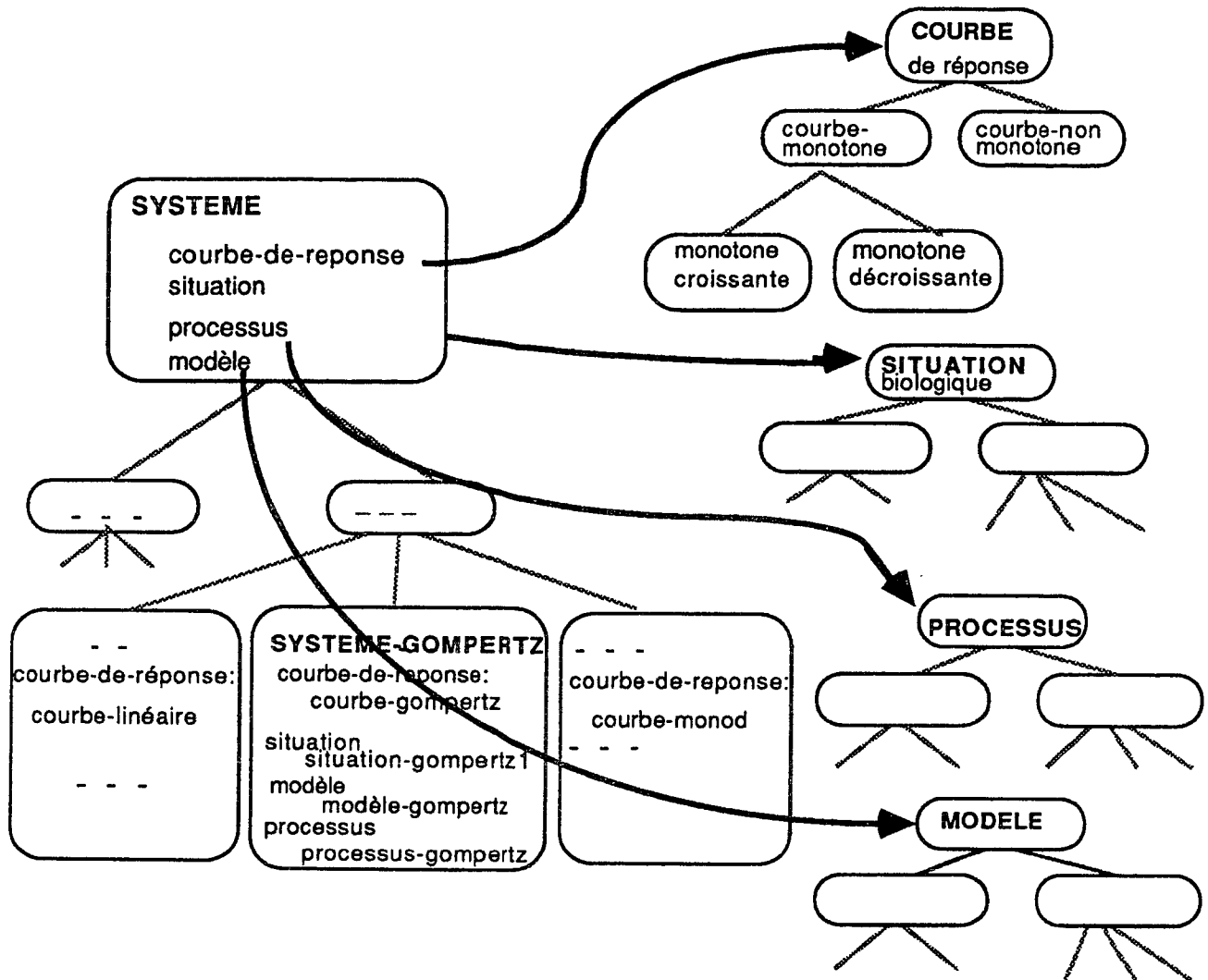


Fig 18: Raisonnement de classification sur de multiples hiérarchies.

Ceci permet de mener le raisonnement conduisant au **choix du modèle** sur la base de différentes connaissances simultanément. Le choix du modèle est assuré par l'algorithme de classification décrit plus haut mais travaillant cette fois sur la hiérarchie de systèmes ce qui implique de mener une classification sur les différentes hiérarchies définies précédemment (Fig 18).

L'utilisateur définit une instance de système correspondant à son problème. Il y a quatre entrées possibles des données correspondant respectivement à la courbe, à la situation, au modèle, au processus.

1. Si par exemple, les données concernent la courbe et la situation, une classification de l'instance de système est réalisée à partir de ces données. L'algorithme de classification parcourt d'abord la hiérarchie des courbes pour classifier le type de courbe observé (cf \$), puis celle des situations (\$). Une ou plusieurs classes de système sont ainsi proposées ou rejetées. Le ou les modèles correspondants sont alors déduits par héritage.

2. Si un modèle est entré, un algorithme de reconnaissance permet de voir s'il existe déjà dans la base.

3. Si un processus est entré, le modèle correspondant est généré par un algorithme de traduction.

- la vérification des informations:

les informations concernant un point de vue sont confrontées à celles fournies par l'autre aspect. Par exemple, une croissance par consommation de substrat avec saturation par excès de substrat est incompatible avec une courbe de réponse présentant une inflexion. Si un utilisateur donne simultanément ces deux informations, il n'obtient que des classes de système "impossibles". Il est donc mis en garde sur la valeur de ses observations.

- de compléter un type d'information incomplète:

si les informations fournies sont incomplètes, elles peuvent être complétées par les informations concernant l'autre aspect. Ceci permet de classer une instance par raffinement de la classe retenue.

Exemples:

1. Soit le dialogue précédent sur la situation. Il se peut que le biologiste ne sache pas répondre à la question "régulée?". L'existence d'une asymptote, elle, permet de répondre à cette question et donc de positionner le système dans la classe des systèmes régulés.

2. Soit le dialogue précédent concernant la courbe de réponse. Il se peut que l'utilisateur ne sache pas répondre à la question "concavité?". C'est le cas par exemple où la courbe observée présente une concavité négative sur son horizon d'étude mais le biologiste estime ses données insuffisantes pour conclure (il se pourrait qu'il y ait un point d'inflexion avant le début du recueil des données). Le type de courbe choisi à ce stade est donc simplement croissante avec asymptote. Sans autre assistance, il faudrait alors tester quantitativement les modèles correspondant aux deux hypothèses: gompertz, logistique, monod, monomoléculaire. Si par contre, le biologiste sait qu'il s'agit d'une croissance par consommation d'un substrat qui est renouvelé et qui se dégrade simultanément, l'hypothèse du point d'inflexion peut être exclue et la classe courbe avec asymptote sans inflexion aussi. Ceci permet donc de classer en raffinant.

- d'apporter une aide à la construction d'un nouveau modèle;

Exemple d'école:

On prend comme exemple celui de la construction du modèle de Lotka-Volterra à partir des systèmes élémentaires de la base (on suppose donc dans le cadre de cet exemple que le modèle de Lotka-Volterra n'existe pas dans la base)

Les informations de surface concernent la courbe de réponse observée et la description de la situation.

Une instance de système est donc créée, elle réunit ces deux types d'information:

```
{système#1
  est-un      =  systeme
  courbe      =  courbe#1
  situation   =  situation#1 }
```

Si la courbe observée est périodique, courbe#1 est une instance de courbe du type:

```
{courbe#1
  est-un      =  courbe-non-monotone;
  périodique =  vrai}
```

La situation en ce qui concerne la proie est une situation de croissance régulée par une interaction de type prédation entre deux populations (prédateur et proies), de même pour la proie.

```
{situation#1
  est-un =  situation-prédation}
```

Différentes juxtapositions de systèmes élémentaires sont essayées, par exemple soient les systèmes élémentaires:

{sy#1
 est-un = système-croissance-exponentiel
 courbe = C_1 : c-croissante-exponentiel;
 situation = s-croissance-non-regulée-biomasse-dep
 processus = $X - a - > 2X$;
 modele = $\frac{dx}{dt} = ax$ }

{sy#2
 est-un = système-decroissance-exponentiel
 courbe = C_2 (c-decroissante-exponentiel);
 situation = s-decroissance-non-regulée-biomasse-dep
 processus = $Y - b - > 0$;
 modele = $\frac{dy}{dt} = -by$ }

{sy#3
 est-un = système-interaction-prédation;
 courbe = C_3 ;
 situation = s-interaction-prédation;
 processus = $X+Y - c > (1+R) Y$
 modele = $\left\{ \begin{array}{l} \frac{dx}{dt} = -cxy \\ \frac{dy}{dt} = dxy \end{array} \right\}$ }

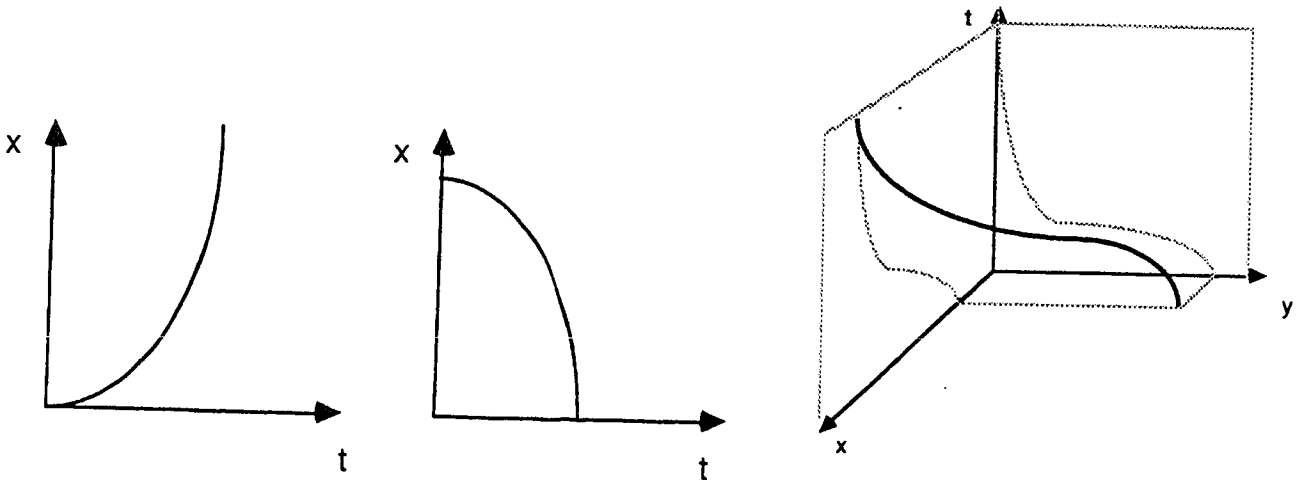


Fig20: Courbes C_1, C_2, C_3

Soit $S = sy_1 + sy_2 + sy_3$

Par définition de l'opération de composition des systèmes, le système $S \begin{cases} C \\ S \\ P \\ M \end{cases}$ est obtenu par

juxtaposition des trois processus précédents, M est le modèle obtenu par traduction, S et C sont respectivement la courbe et la situation associée à ce processus composé:

{S

| | | | |
|-----------|---|---|---|
| est-un | = | système | |
| courbe | = | C; | |
| situation | = | s-prédation | |
| processus | = | $\begin{cases} X - a - > 2X \\ X+Y - c > (1+R) Y \\ Y - b - > 0 \end{cases}$ | |
| modele | = | $\begin{cases} \frac{dx}{dt} = ax - mxy \\ \frac{dy}{dt} = -by + rxy \end{cases}$ | } |

Le système ainsi construit est cohérent avec les inforlations fournies: courbe périodique, situation de prédation.

Supposons par contre que les données aient été différentes. Par exemple, supposons que la courbe de observée pour X ait été non périodique et ait présenté des oscillations amorties, le système ainsi construit n'aurait pas été satisfaisant puisque la courbe correspondante, elle, serait périodique en raison de la nature mathématique du modèle M . D'autres combinaisons peuvent être alors essayées. Il suffit pour cela de choisir d'autres systèmes élémentaires, c'est à dire d'autres processus élémentaires dans la base de processus primitifs. Supposons par exemple que le système $sy\#1$ soit remplacé par

{sy' #1

| | | |
|-----------|---|---|
| est-un | = | système-croissance-logistique |
| courbe | = | C_1 (courbe-logistique); |
| situation | = | s-croissance-autolimitaion |
| processus | = | $X < -\frac{a}{b} > X$ |
| modele | = | $\frac{dx}{dt} = r x (1 - \frac{x}{k})$ |

Cette fois, le système construit est satisfaisant, puisqu'il présente des oscillations amorties;

Cet exemple illustre bien comment une légère modification structurelle (modification ici d'un seul système constitutif, le système $sy\#1$) peut avoir d'importantes conséquences sur le comportement du système global: l'introduction d'une autolimitation de la croissance des proies suffit à empêcher l'existence de solutions périodiques. On obtient alors au contraire un système asymptotiquement stable vers lequel les solutions convergent de manière oscillante amortie ou même exponentiellement. **Le choix des systèmes élémentaires est donc une phase primordiale dans le procédé de construction de nouveaux modèles.** Ce choix est conduit par le raisonnement précédemment décrit.

Conclusion

On a vu comment la représentation centrée-objet est le support de mécanismes de raisonnement originaux, en particulier en ce qui concerne le choix "multi-critères" d'un objet dans une base. La structuration des connaissances proposée a permis de répondre à différentes questions: validation et cohérence des informations, inférence de connaissances profondes à partir de connaissances

superficielles, coexistence de connaissances qualitatives et quantitatives. Des solutions intéressantes ont pu être apportées tant pour le choix d'un modèle dans une base de prototypes que pour la construction d'un nouveau modèle en dynamique des populations.

Ce type d'approche pourra être étendue plus généralement à la formulation de modèles de croissance dans d'autres domaines. Une analyse similaire peut probablement être menée en économétrie par exemple.

Les perspectives de ce travail concernent maintenant d'une part la construction de nouveaux modèles. Le problème de l'analyse décrementale d'un système doit être approfondi. Le travail de transfert et de la modélisation de l'expertise sur ce point doit être poursuivi. Il doit être également poursuivi en ce qui concerne les autres étapes du processus de modélisation, notamment pour l'étude qualitative d'un modèle. Celles-ci nécessitent la structuration d'une base de méthodes dans le formalisme objet.

Une question reste ouverte, celle de la structuration d'une base de connaissances pour l'aide à la conception de modèles en général. Des solutions peuvent être envisagées par l'introduction, par exemple, de différents formalismes de représentation intermédiaires à caractère relativement général (formalisme compartimental, contraintes qualitatives, systèmes dynamiques etc...). Une assistance au choix du formalisme adéquat devra être proposée. Des opérations élémentaires comme celles qui ont été présentées ici pour la composition de systèmes devront être définies pour chacun de ces formalismes.

La conceptualisation d'une base de connaissances avec un modèle purement objet est un art difficile. Ce travail montre toutefois les apports d'une telle approche tant pour l'expert que l'informaticien.

BIBLIOGRAPHIE

- [AIKINS, 83]
Aikins J.
Prototypical knowledge for expert systems. Artificial Intelligence, 20, 1983, pp 163-210
- [CHANDRESEKERAN, 83]
Deep versus compiled knowledge approaches to diagnostic problem solving.
Int. J. Man-Machine Studies (1983) 19,425-436
- [CLANCEY, 81]
Clancey W.J., Letsinger R.
NEOMYCIN: Reconfiguring a rule-based expert system for application for teaching
7th IJCAI, Vancouver B.C., Canada, 829-836, 1981
- [CLANCEY, 85]
Clancey W.J., Letsinger R.
Heuristic Classification
A.I. Journal 27(4), 289-350, 1985
- [DAVIS, 83]
Davis R.
Reasoning from first principles in hardware trouble shooting.
Int. Journal of Man-machines Studies, 19, 403-423, 1983.
- [EMANUEL et KNORRE, 75]
Emanuel N., Knorre D.
Cinétique chimique
Ed MIR, Moscou, 1975
- [FIKES & KEHLER, 85]
Fikes R., Kehler T. *The role of frame-bases representation in reasoning*. Communications of the ACM, Vol 28, sept 1985, pp 904-920

18.05.88

[FORRESTER, 61]

Forrester J.W. : *Industrial Dynamics*. MIT Press, 1961

[GARFINKEL, 62]

Garfinkel D.

Digital Computer of Simulation of an Ecological System based on a modified Mass Action Law. *Ecology*, 45, 502, 507, 1962

[HAMROUNI, 79]

Hamrouni M.K.2 : *Etude et développement d'un système informatique d'aide à l'élaboration de modèles en biologie*.

[HART, 81]

Hart Peter E.

Directions for A.I. in the Eighties.
SIGART Newsletter, 79, january.

[KLEER et BROWN, 81]

DeKleer, J. & Brown J.S.

Mental Models of physical mechanisms and their acquisition. In ANDERSON J. R, Ed., *Cognitive Skills and Their Acquisition*, pp285-309.
Hillsdale, New Jersey: Erlbaum Associates.

[KUIPERS, 84]

Kuipers B.

Commonsense Reasoning about Causality: Deriving Behavior from Structure
Artificial Intelligence, 24, 1984

[KUIPERS, 87]

Kuipers B.

New Reasoning methods for Artificial Intelligence in Medecine.
Int. J. Man-Machine Studies, 27, 708-718, 1987

[LAUNAY, 88]

Launay G.

Identifiability of a model of compartmentation of serotonin in human platelets.
in Proc. of The Eighth Conference on Analysis and Optimization of Systems, Antibes, June 8-10, 1988.

[LAVERGNAT, 79]

Lavergnat G.

Elaboration d'un modèle chimiluminescence.

Actes du colloque Elaboration et Justification des modèles. Applications à la biologie.
Delattre P. et Thellier M., Maloine Ed, Paris, 1979.

[MINSKY, 75]

Minsky M. : *A framework for representing Knowledge* . The Psychology of Comp Vision
P.H. Winston ED., McGrawHill 1975.

[ODUM, 72]

Odum H.T. : *An energy language for ecological and Social systems: its physical basis*.

Systems Analysis and simulation in Ecology B.C.Patten Academic Press 1972, pp.139-211

[PATIL, 86]

Patil S.

Review of Causal Reasoning in Medical Diagnosis

18.05.88

in proc. of The tenth annual symposium of IEEE Computer Applications in Medical Care, Washington, 1986.

[PATIL, 87]

Patil S.

A Case Study on Evolution of System Building Expertise: Medical Diagnosis.

AI in the 1980s and Beyond- An MIT Survey

ed. Grimson W. Eric L. and Patil Ramesh S., The MIT Press, Massachussets, 1987.

[PAVE, 88]

Interprétation et construction de modèles de la dynamique des populations à l'aide de schémas fonctionnels.

Cahiers Edora N°1, INRIA, Mars 1988

[PAVE & RECHENMANN, 86]

Pavé A., Rechenmann F.]

Computer aided modelling in biology an artificial intelligence approach.

Artificial Intelligence and Simulation, SCS Rev , 1986

[PIERRET, 87]

Pierret C.

Vers l'intégration des objets symboliques et biologiques dans EDORA : expressions, modèles, processus, systèmes.

Cahiers d' EDORA N°1, INRIA, mars 1988

[PIERRET, 87]

Pierret C.

Object-centered Knowledge Representation for biological Modeling.

International Symposium on A.I., Expert Systems and Languages in Modelling and Simulation, Elsevier Sc. Publ. Barcelona, 1987

[PIERRET, 88]:

Pierret C

Knowledge Representation for computer aided modelling in Biology.

IMACS World Congress'88, Paris, July 18-22,1988

[ALENGRY & PIERRET, 88]

Alengry P. - Pierret C.

Object-centered representation as a tool for supporting the diagnosis activity.

ERGO/IA'88, Biarritz, France, October 4-6,1988

[PIERRET, 88]

Pierret C.

Contribution of an object-centered representation for computer aided modeling in biology..
(soumis)

[ALENGRY & PIERRET]

Alengry P. - Pierret C.

Analyse de procédures de résolution d'accidents. (rapport de recherche, à paraître)

[ALENGRY & PIERRET]

Alengry P. - Pierret C.

Recueil et formalisation d'expertise à l'aide du concept de classification. (rapport de recherche, à paraître)

[PIVOT, 87]

Pivot B. -Prokop M.

Définition et réalisation d'une fonctionnalité de classification dans Shirka.

18.05.88

Institut National Polytechnique de Grenoble, juin 1987.

[POPLE, 82]

Pople H.

Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics.

In Szolovits P., ed., *Artificial Intelligence in medicine*. Westview Press, Boulder, Col, 1982.

[PUCCIA & LEVINS, 85]

Puccia Charles J., Levins R.

Qualitative Modeling of Complex Systems

Harvard University Press, Cambridge, Massachusetts, and London, England, 1985

[RECHENMANN, 85]

Rechenmann F.

SHIRKA: Mécanismes d'inférence sur une base de connaissances centrée-objet. Cinquième congrès AFCET-ADI-INRIA. Reconnaissance des formes et Intelligence Artificielle, Nov 1985

[RICHMOND, 85]

Richmond B.

STELLA. User's Guide. High-Performance Systems,
Dartmouth College, New Hampshire, USA, 1985.

[STEELS, 87]

Steels L.

The Deepening of Expert Systems.

AICOM Vol 0 N°1 Aug 1987

[STEFIK, 83]:

Stefik M., Bobrow D.G., Mittal S., Conway L.

Knowledge programming in LOOPS: report on an experimental course. The A.I. magazine, Vol.4, N°3, 1983, pp 3-13

[SWARTOUT, 81]

Swartout W.

Producing Explanations and Justifications of Expert Consulting Programs
Ph.D. diss, Massachusetts Institute of Technology, LCS.

[SZOLOVITS & SCHWARTZ, 81]

Patil R., Szolovits P., Schwartz W.,

Causal understanding of patient illness in medical diagnosis.

Proc IJCAI-81, 893-899, August 1981

[WIELINGA & BREUKER, 84]

Wielinga B.J., Breuker J.A.

Interpretation of Verbal Data For Knowledge Acquisition.

ECAI 84, 145-172, 1984

[WORTMAN, 66]

Wortman P.

Representation and Strategy in diagnostic problem solving.

Human Factors 8:48, 1966.

[ZEIGLER, 86]

Zeigler B. P.

Towards a knowledge based implementation in multifaceted modelling methodology.

In *A.I. Applied to Simulation*, Kerckhoffs E.J.H., Vansteenkiste G.C., Zeigler B., Ed

18.05.88

SCS, Simulation Series Vol. 18, N°1, 1986.

ANNEXE

Exemples de sessions :

Ecran 1, Ecran 2: classification d'une instance de courbe dans la hiérarchie des courbes

Ecran 3: classification d'une instance de situation dans la hiérarchie des situations

Ecran 4: explication de la classification de la situation ci-dessus

Ecran 5: classification d'une instance de système dans la hiérarchie des systèmes

```

Shirka:cr-inst
nom-classe? courbe
Classe: courbe
Sous-classe? -
Nom de l'instance? c1
optimum? -
-> c1
Shirka:classif
  Classif:class
  instance? c1
  classe? <
  optimum? non
  croissante? vrai
  asymptote? vrai
  concavite? y+
  nb-pts-inflexion? 1
  position-pt-inflexion? -
Les classes sures:
-> courbe courbe-monotone monotone-croissante
croissante-avec-asymptote asymptote-avec-inflexion
  Les classes possibles:
-> sigmoide-avec-asymptote c-logistique c-gompertz
  Les classes impossibles:
-> courbe-non-monotone monotone-decroissante
croissante-sans-asymptote asymptote-sans-inflexion
Classif:|

```

```

  Classif:-
Shirka:cr-inst
nom-classe? courbe
  Classe: courbe
Sous-classe? -
Nom de l'instance? c2
optimum? -
-> c2
Shirka:classif
  Classif:class
  instance? c2
  classe? <
  optimum? non
  croissante? vrai
  asymptote? vrai
  concavite? y+
  nb-pts-inflexion? 1
  position-pt-inflexion? milieu
Les classes sures:
-> courbe courbe-monotone monotone-croissante
croissante-avec-asymptote asymptote-avec-inflexion
sigmoide-avec-asymptote c-logistique
  Les classes impossibles:
-> courbe-non-monotone monotone-decroissante
croissante-sans-asymptote asymptote-sans-inflexion c-gompertz
Classif:|

```

```

-> sit2
Shirka:class
instance? sit2
classe? situation
avec-inference? <
monotone? vrai
croissance? vrai
regulee? vrai
saturation? vrai
nb-facteur-limitant? 1
nature-saturation? ??
domaine: initiale lente moyenne brutale
nature-saturation? brutale
nature-facteur? substrat
renouvele? -
  Les classes sures:
-> situation situation-monotone croissance croissance-regulee
regulation-exterieur
  Les classes possibles:
-> regulation-unique consommation-substrat saturation-secondaire
s-kostitzin
  Les classes impossibles:
-> situation-non-monotone decroissance croissance-non-regulee
autoregulation regulation-multiple regulation-catalyseur
saturation-initiale s-logistique4 s-logistique2
Shirka:

```

```

Shirka:expl
instance? sit2
classe? s-gompertz
L'instance sit2 pourrait appartenir a la classe regulation-unique
La valeur de l'attribut renouvele est inverifiable
L'instance sit2 ne peut pas appartenir a la classe
regulation-catalyseur
La valeur de l'attribut nature-facteur est substrat
alors que l'admissibilite est pour: catalyseur
Shirka:expl
instance? sit2
classe? s-logistique4
L'instance sit2 pourrait appartenir a la classe regulation-unique
La valeur de l'attribut renouvele est inverifiable
L'instance sit2 ne peut pas appartenir a la classe s-logistique4
La valeur de l'attribut nature-saturation est brutale
alors que l'admissibilite est pour: moyenne
Shirka:expl
instance? sit2
classe? s-monod
L'instance sit2 pourrait appartenir a la classe regulation-unique
La valeur de l'attribut renouvele est inverifiable
L'instance sit2 ne peut pas appartenir a la classe saturation-initiale
La valeur de l'attribut nature-saturation est brutale
alors que l'admissibilite est pour: initiale
Shirka:

```

```

courbe? -
-> syst1
Shirka: classif
      Classif: class syst1 <
      situation? §
Ilot situation:
      monotone? vrai
      croissante? vrai
      nb-facteur-limitant? 1
      regulee? vrai
      saturation? vrai
      nature-saturation? initiale
      nature-facteur? substrat
      renouvele? faux
      Les classes sures:
      -> situation-non-monotone decroissance
      croissance-non-regulee autoregulation
      regulation-multiple regulation-catalyseur
      saturation-secondaire s-monomoleculaire1
      Quel est votre
      choix? s-monod
      courbe? §
Ilot courbe:
      optimum? vrai
      valeur non admissible
      optimum? non
      croissante? vrai
      asymptote? vrai
      concavite? y-
      nb-pts-inflexion? 0
      pente-mi-asymptote? -
      Les classes sures:
      -> courbe courbe-monotone monotone-croissante
      croissante-avec-asymptote
      Les classes possibles:
      -> asymptote-sans-inflexion c-monod c-mono-mol
      Les classes impossibles:
      -> courbe-non-monotone monotone-decroissante
      croissante-sans-asymptote asymptote-avec-inflexion
      Quel est votre
      choix? c-monod
      Les classes sures:
      -> systeme systeme-monod
      Les classes impossibles:
      -> systeme-gompertz systeme-logistique
Classif:

```

