BASE DE CONNAISSANCES SUR LES CORAUX DE LA REUNION ET DES MASCAREIGNES

G. FAURE : Université de Montpellier II- Institut des Sciences de l'Ingénieur, 34O95, Montpellier &Université de Provence(Marseille I) « Dynamique des Récifs »UPRESA 6019

N. CONRUYT : Université de La Réunion- IREMIA- B.P. 7151 - 97415- St-Denis - La Réunion.

M. PICHON : EPHE, Université de Perpignan, Avenue de Villeneuve, 66025- Perpignan.

M. GUILLAUME : BIM, MNHN, URA 699-CNRS, 55, rue de Buffon, 75OO5- Paris.

D. GROSSER : Université de La Réunion - IREMIA- B.P. 7151- 97415- St-Denis - La Réunion.

Résumé

Une base de connaissances sur les coraux de La Réunion et de l'ensemble des Mascareignes est en cours de création. Elle a pour support une collection d'environ 3000 spécimens appartenant à 58 genres et 185 espèces. Cette base de connaissances constitue un outil informatique d'aide à l'identification et à la classification des spécimens de Scléractiniaires de l'archipel (Réunion, Maurice, Rodrigues). Dans le but de mettre en oeuvre la méthode scientifique en biologie (conjecturer et tester), notre approche suit le processus d'acquisition des connaissances. Celle-ci est divisée en trois étapes, dont la première constitue la partie la plus importante de la méthode.

- Acquisition des connaissances (création d'un arbre de description, constitution d'une base de cas)
- Traitement des connaissances (classification, identification).
- Validation.

A ce jour, seule la famille des Pocilloporidae est enregistrée dans la base de connaissances. L'ensemble des Astrocoeniidae et des Thamnasteriidae (2 genres, 9 espèces), des Fungiidae (6 genres et sous genres, 11 espèces), est actuellement expertisé. L'expertise de la famille des Poritidae est en cours de réalisation, de même que celle des Agariciidae

Mots-clés : base de connaissances, Scléractiniaires, Mascareignes, modèle descriptif, traitement des connaissances, validation.

Abstract

A Scleractinian knowledge base is coming true from a Mascarene Archipelago (Reunion, Mauritius, Rodriguez islands) corals collection(about 3000 specimens belonging to 58 genera and 185 species). It consists of a computer aided systematics tools for describing, classifying and identifying coral specimens. In order to apply the scientific method in biology (conjecture and test), our approch follows a natural process of knowledge acquisition which is divided in three steps :

- Acquisition of descriptive model and examples (cases base).
- Processing of this knowledge (classification and identification purposes).
- Validation.

At the moment, the family Pocilloporidae is the only one to be recorded in the knowledge base, but for others, such as Astrocoeniidae and Thamnasteriidae (including 2 genera, 9

species) Fungiidae (6 genera and sub-genera, 11 species). families, the first step (acquisition of a descriptive model) is now effected. The valuation of the families Poritidae, Agariciidae is on the way to be achieved.

Keys-words : knowledge base, Scleractinian-corals, Mascarene Archipelago, descriptive model, processing of the knowledge, validation.

1.- INTRODUCTION

La mise en oeuvre d'outils informatiques permettant la gestion de collections, la description, la classification des espèces du monde vivant, apparait aujourd'hui comme une nécessité dans le domaine de la biodiversité. L'existence d'une importante et récente collection de Scléractiniaires (environ 3000 spécimens appartenant à 58 genres et 185 espèces, déposés à l'Université de La Réunion, au Muséum de St-Denis, au MNHN de Paris, et récoltés dans les trois iles de l'archipel : Réunion, Maurice, Rodrigues) (G.FAURE., 1982), constitue un patrimoine scientifique qu'il convenait de conserver et de valoriser, en utilisant les techniques informatiques appropriées que constituent les bases de données et de connaissances (CONRUYT N., 1994)

Les bases de données scientifiques sur la biodiversité, qui se développent actuellement en biologie, sont mises sur support CD-ROM (ex. : ETI en Hollande, ICLARM aux Philippines, etc...) ,ou sont accessibles par INTERNET (ex. : HBS à Hawaii, AIMS et CSIRO en Australie, etc...). Elles constituent des moyens de diffusion de l'information sur les espèces biologiques.

Contrairement à ces systèmes d'information bibliographique et/ou taxinomique qui nécessitent de connaître le nom (ou taxon) du spécimen, **les bases de connaissances** offrent une aide à l'identification pour ceux qui ne connaissent pas le nom d'un échantillon, et une aide à la classification pour les spécialistes,dont le domaine n'est que partiellement étudié (ce qui est vrai pour une grande partie de la faune marine, ainsi également , que pour la flore et la faune tropicales terrestres).

2.- BASE DE CONNAISSANCES SUR LES CORAUX DE LA REUNION ET DES MASCAREIGNES.

2.1.- Justificatifs

Une dizaine d'experts se partagent la connaissance des coraux dans le monde. Dans quelques années, du fait de leur départ, l'expertise accumulée pour les identifier ne sera plus disponible . Il ne restera que les collections et les monographies et la mémoire littéraire pour reconnaître les espèces.

Connaissant le rôle très important des écosystèmes récifaux pour l'équilibre de la planète, et les dégradations continuelles opérées par l'homme sur son milieu naturel, il est impératif de pouvoir transmettre ce savoir aux personnes chargées de la biodiversité des milieux récifaux ainsi qu'aux futurs systématiciens.

2.2.- Objectifs.

- Identification des espèces de coraux de La Réunion et de l'ensemble de l'archipel des Mascareignes.
- Création d'outils pédagogiques et multimédia de formation à l'expertise.
- Valorisation et gestion des collections de spécimens dans les museums.
- Nouvelle approche de la systématique des coraux assistée par ordinateur.
- Travail de recherche collaborative à distance.
- Couplage avec les systèmes d'informations existants sur les coraux.

3.- METHODOLOGIE.

La gestion des connaissances en sciences de la vie, nécessite de s'appuyer sur une démarche expérimentale. Les experts en sciences de la vie appliquent une méthode inductive fondée sur l'observation des faits, l'énoncé d'hypothèses, la pratique de tests expérimentaux pour mettre ces hypothèses à l'épreuve.

Nous avons construit un atelier de gestion de bases de connaissances itératif qui reproduit ces trois étapes sous la forme suivante (Fig. 1).

- Acquisition des connaissances.
- Traitement des connaissances
- Validation des connaissances et itération.

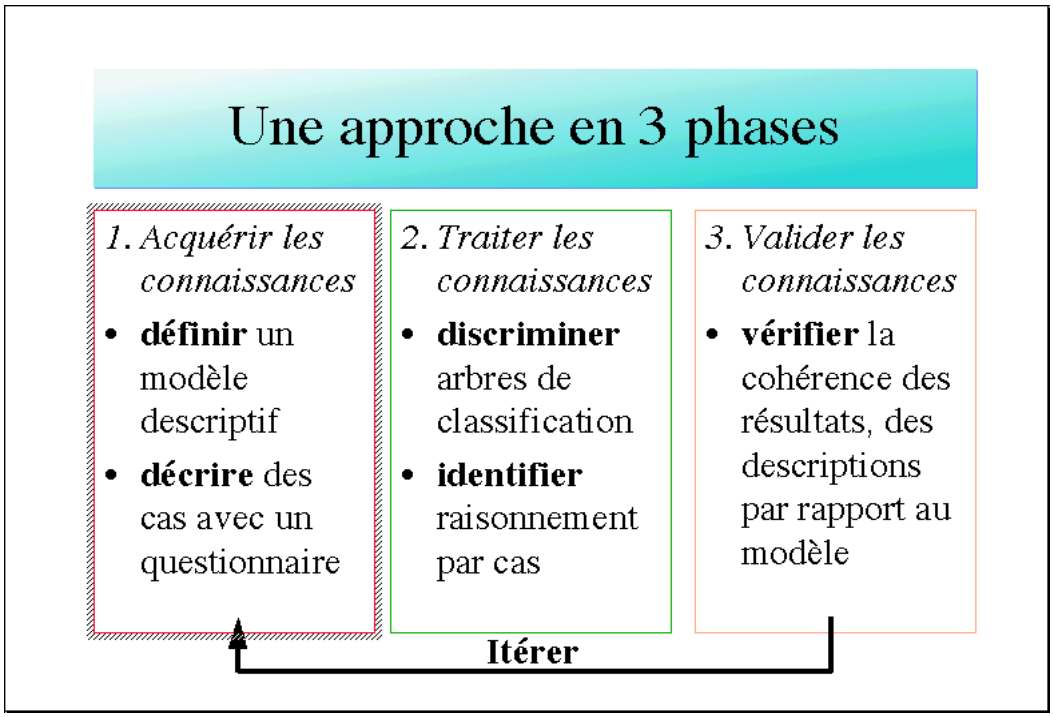

Fig.1.- Méthodologie de gestion des connaissances .

A l'aide de ces outils, l'expert peut appliquer la méthode scientifique en biologie : **expérimenter** (apprendre des règles de classification à partir de l'arbre inductif) et **tester** (mettre à l'épreuve les hypothèses inductives par de nouvelles observations). Le non-spécialiste va pouvoir consulter les **systèmes expert** résultant, avec une forte probabilité de succès (par détermintion déductive ou identification comparative).

3.1- Acquisition des connaissances.

Le but de cette phase est de construire une base de descriptions qui soient comparables entre elles. Ces descriptions structurées et pré-classées par l'expert ,permettent de constituer une base de cas (90 cas actuellement pour le seul genre ***Pocillopora***) . Lors de la phase d'acquisition des connaissances, nous distinguons l'étape d'acquisition du modèle descriptif (*l'observable*), de l'étape d'acquisition des descriptions (*l'observé*).

Le modèle descriptif représente tout ce qui est *observable* pour notre domaine d'étude. Sa réalisation constitue une étape fondamentale de la méthode (CONRUYT N. *et al.*, 1997 a,b). De sa qualité, dépend la validité de la base de connaissances. L'expert (ou mieux le groupe d'experts) doit (outre ses compétences) collationner un très grand nombre d'informations bibliographiques, de terrain (variabilité intra-spécifique ,notions d'écomorphes et de géomorphes), utilisation et redescription des types... Il doit, dans un deuxième temps, discerner et hiérarchiser les caractères spécifiques et génériques significatifs. La définition du

modèle est représentée sous forme d'un schéma structuré de tous les objets, attributs et valeurs possibles du domaine, ce dernier constituant la racine de ce que l'on nomme l'*arbre de description* . Pour construire cet arbre, nous suivons certaines *logiques descriptives* en sciences de la vie, avec des règles de bon sens pour décrire une espèce :(dé)composition, point de vue, spécialisation, multi-instanciation, conditions contextuelles (LE RENARD J. & CONRUYT N., 1994). Pour la famille des **Pocilloporidae**, par exemple, 45 objets et 97 attributs ont été recencés pour 4 genres et 14 espèces (Fig.2). Ensuite, un questionnaire est construit automatiquement à partir du modèle descriptif. Le biologiste utilise celui-ci comme un guide d'observation pour acquérir des descriptions observées et constituer une base de cas cohérente par rapport au modèle (Fig. 3 **a** et **b** ; Fig.4 **a** et **b**).

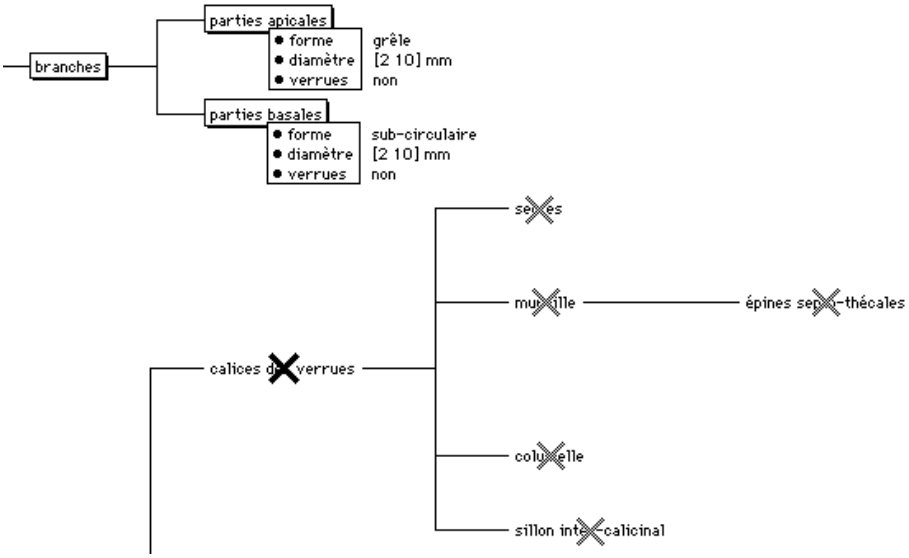

Fig.2.- Extraits de l'arbre de description du genre ***Pocillopora***

Fig. 3. **a** - Une vue locale de l'objet « calices de l'apex » chez *P. damicornis acuta* : **caractéristiques**

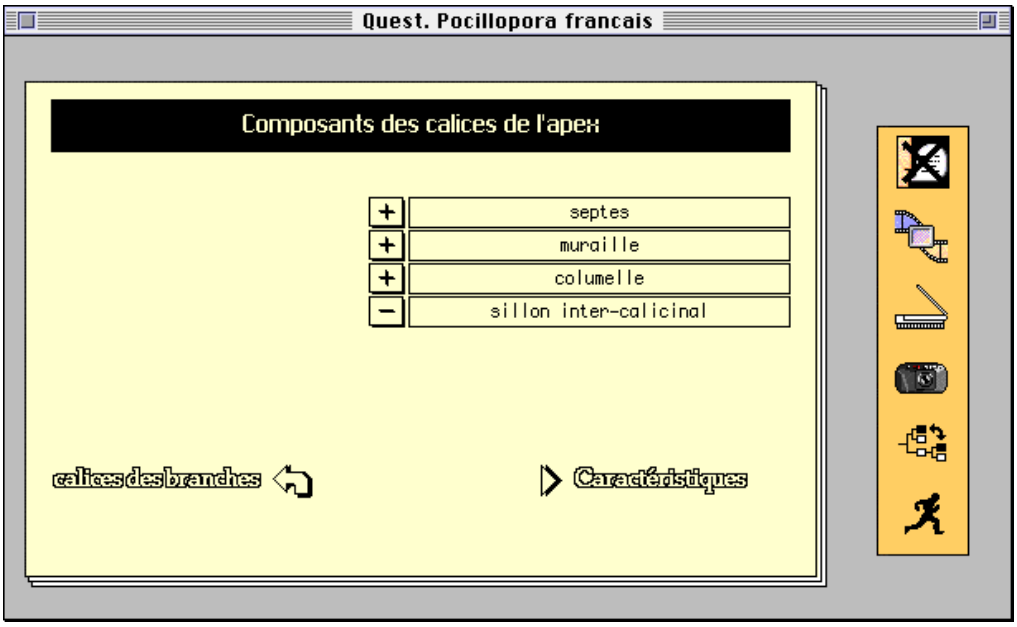

Fig. 3. **b** - Une vue locale de l'objet « calices de l'apex » chez P.*damicornis acuta* :**composants.**

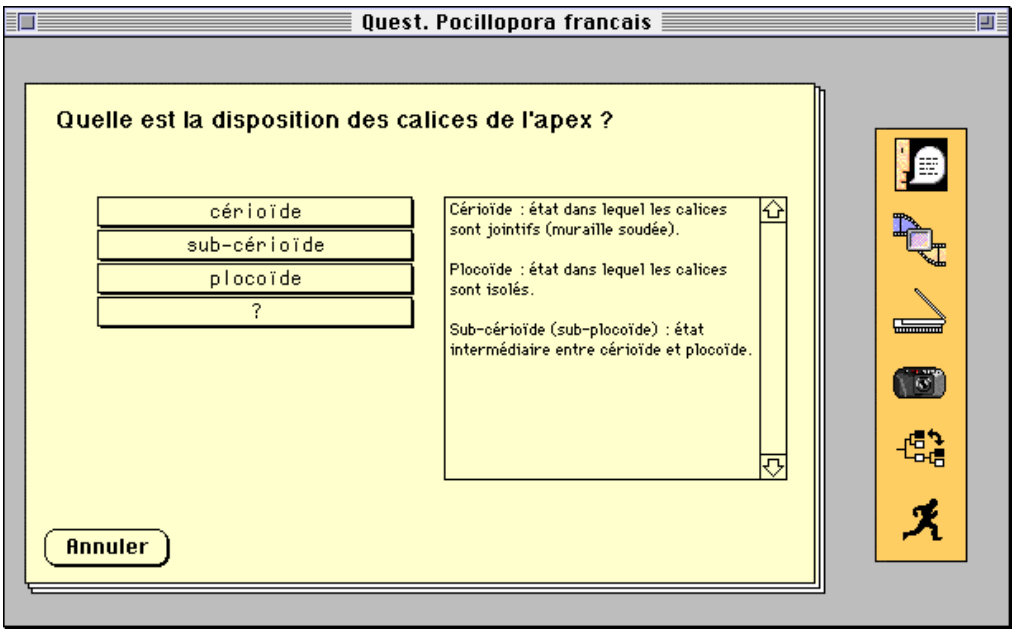

Fig. 4 **a-** Un exemple d'attribut **commenté**

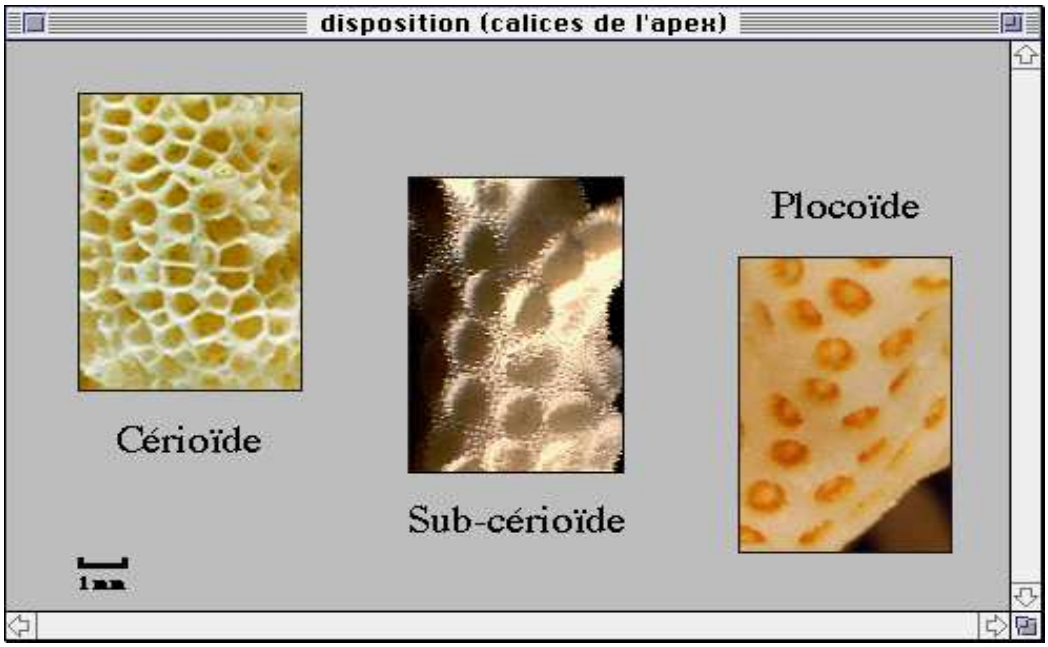

Fig.4. **b** - Un exemple d'attribut **illustré.**

3.2.- Traitement des connaissances.

Les cas sont alors traités selon deux technologies complémentaires en fonction de l'objectif poursuivi. Pour la **classification**, un arbre de décision est construit. A partir des descripions (représentation en extension), une méthode inductive fondée sur l mesure d'entropie et du gain d'information fabrique une caractérisation des classes par un ensemble de régles .Chaque chemin depuis l racine vers les feuilles de l'arbre de décision est une règle de classification (également appelée *diagnose*).

Pour les ***Pocillopora*** (genre de coraux très courant dans les Mascareignes), nous donnons à titre indicatif l'arbre de décision suivant, qui les classifie les 9 descriptions des types (espèces et écomorphes) (Fig.5).

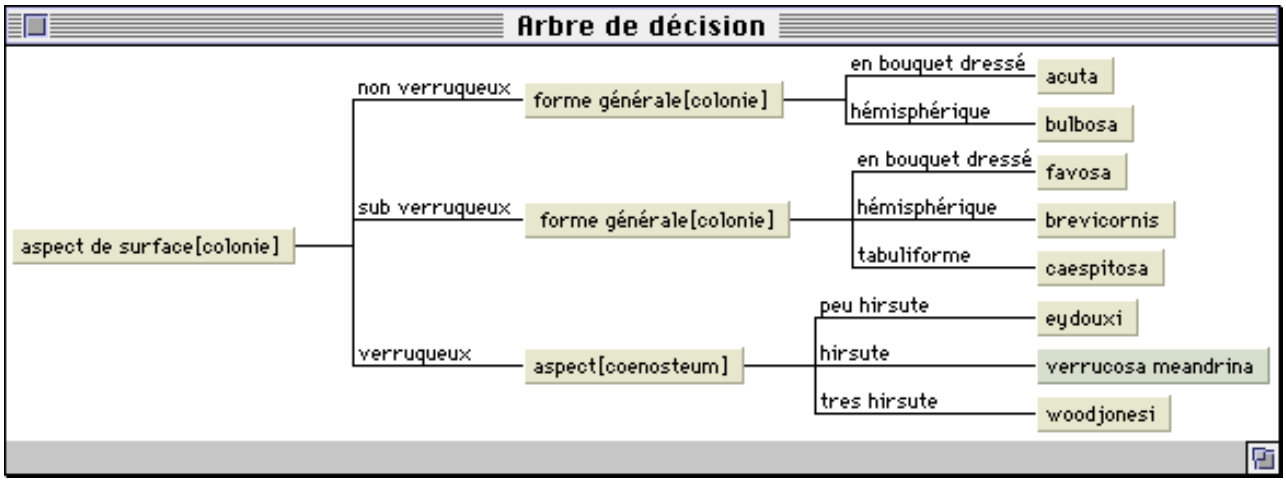

Fig.5.- Exemple d'arbre de décision pour classifier le genre ***Pocillopora.***

Cet arbre de classification peut être utilisé pour déterminer une nouvelle observation. Nénmoins, lorsque l'utilisateur répond « inconnu » à une quesion, la consultation de cet arbre est idadaptée.

Pour l' **identification**, le raisonnement à partir de cas est utilisé. Etant donné un ensemble d'exemples, notre méthode extrait dynamiquement le critère le plus efficace à partir d'une liste ordonnée de tests, après chaque réponse de l'utilisateur (Fig. 6). Les cas sont sélectionnés en fonction de cette réponse. Si la réponse est inconnue, le second test le plus discriminant est proposé à l'utilisateur, et ainsi de suite. D'autres méthodes faisant appel à des mesures de ressemblance peuvent être utilisées (identification par comparaison).

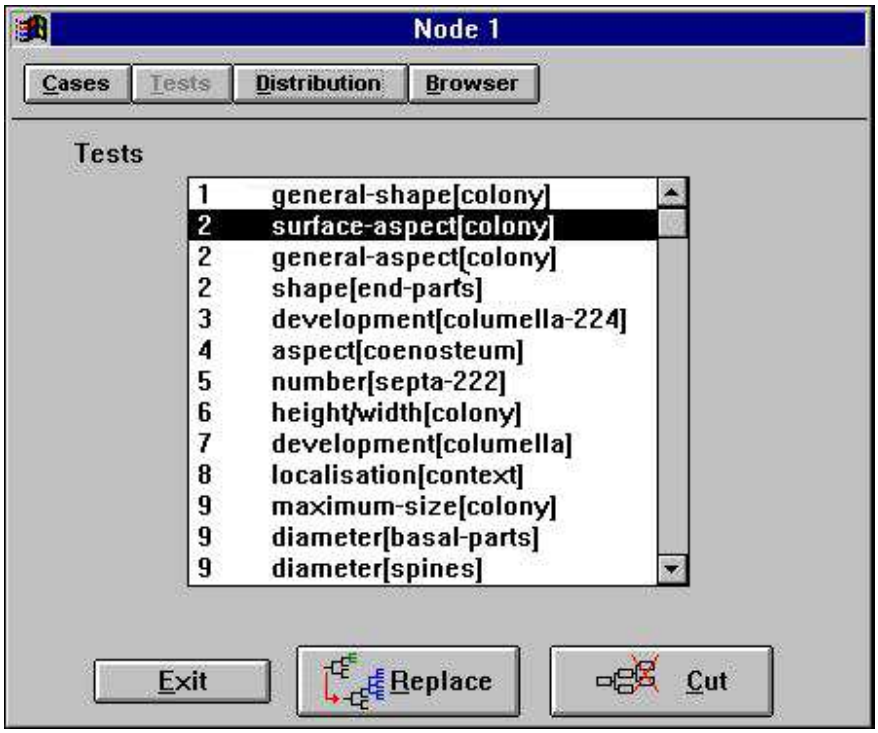

Fig. 6.- La liste des tests ordonnés à un noeud (ici la racine) est choisie par l'intermédiaire d'une mesure de discrimination inter-classes. A chaque noeud, l'attribut le plus discriminent est choisi pour séparer les cas.

3.3. Validation des connissances et itération.

Pour la classification, l'expert va pouvoir analyser les règles produites. Il contrôlera si les diagnoses son significatives pour séparer les espèces voisines. Il pourra aussi détecter certaines erreurs dans ses descriptions ayant servi à générer l'arbre de décision.

Pour l'identification, une phase de test est nécessaire pour contrôler la fiabilité des résultats à l'aide du raisonnement par cas. Cette épreuve de vérité permet de savoir si le système d'identification est robuste ou non (tolérance aux bruits). Dans une phase d'**expérimention,** différents utilisateurs (novices du domaine, autres experts) utilisent le questionnaire avec des spécimens,soit non décrits initialement dans la base de cas, soit déjà décrits par d'autres expérimentateurs. Les résultats du sytème expert sont alors comparés aux identifications de l'expert. L'idéal serait de pouvoir pour chaque description d'espèce, constituer un « pool » d'experts, afin de renforcer la « robustesse » de l'arbre de description.

Il s'en suit une phase d'**évaluation** et de **discussion**, où l'expert va pouvoir détecter certaines incohérences dans le modèle descriptif et/ou dans la base de cas (mauvais choix de caractères, illustrations non significatives, vocabulaire imprécis et/ou insuffisant, etc…).Une concertation entre les utilisateurs et l'expert permet d'identifier les différences d'interprétation du vocabulaire et des illustrations.

La méthode proposée donne donc la possiblité à l'expert de mettre à jour les connaissances observables et observées en fonction des résultats (classification et identification) et

d'améliorer son modèle descriptif de manière itérative. Par exemple, il va pouvoir redéfinir les attributs et les valeurs, restructurer les objets du modèle descriptif, faire de nouvelles illustrations plus explicites (dessins), de nouveaux commentaires, etc... Dans le cas des ***Pocillopora***, les interrogations les plus nombreuses ont porté : sur la féfinition de la partie apicale des branches, le caractère hirsute du coenosteum, la densité des épines septothécales, le développement de la columelle...

4.- DISCUSSION, CONCLUSION.

A ce jour, seule la famille des **Pocilloporidae** est enregistrée dans la base de connaissances. L'ensemble des **Astrocoeniidae** et des **Thamnasteriidae** (2 genres, 9 espèces), des **Fungiidae** (6 genres et sous genres, 11 epèces), est actuellement expertisé. L'expertise de la famille des **Poritidae** est en cours de réalisation de même que celle des **Agariciidae.**

L'une des difficultés majeure rencontrée lors de la phase d'expertise est due à l'extrême variabilité intra-spécifique, voire intra-coloniale que presentent les espèces à large répartition écologique et/ou géographique. Par ailleurs, et pour les mêmes raisons, il existe dans la littérature de nombreuses lacunes et contradictions. Par exemple, la fréquence et le développement de la columelle, qui aujourd'hui sont des caractères qui discriminent chez le genre ***Pocillopora*** le groupe *verrucosa-meandrina* (chez qui elle est absente ou abortive), du groupe *eydouxi-woodjonesi* (chez qui l'axe central est au contraire bien développé). Dans la description originale de P. *eydouxi* (MILNE EDWARDS H. & HAIME J., 1860), la columelle est absente, mais l'examen du type (MNHN, Paris, n° 434, Collection ME & H) montre à l'évidence une columelle toujours présente et bien développée... A l'inverse, ces mêmes auteurs mentionnent la présence chez P.*verrucosa* d'une « saillie columellaire bien développée ».

REFERENCES

CONRUYT N.,1994-Amélioration de la robustesse des systèmes d'aide à la description,à la classification et à la determination des objets biologiques.*Thèse Doctorat* , Univ.Paris IX Dauphine,1-281

CONRUYT N., FAURE G., ANCEL G., LE RENARD J., GUILLAUME M., NAIM O., GRAVIER-BONNET N., GROSSER D., 1997 a -A knowledge base for corals of the Mascarene archipelago : Genus *Pocillopora. Proc. 8th. Int. Coral Reef Symp.* Panama « sous presse ».

CONRUYT N., GROSSER D., FAURE G., 1997 b -Ingénierie des connaissances en sciences de l a vie. Application à la systématique des coraux des Mascareignes. *Actes des journées Ingénieries des connaissances, IC' 97,* Roscoff, 513-525.

FAURE G., 1982.- Recherches sur les peuplements de Scléractiniaires des récifs coralliens des Mascareignes. *Thèse Doct Etat Sci. nat.,* Uni. Aix-Marseille II, 2, 1-219.

LE RENARD J. & CONRUYT N., 1994- On the representation of observational data used for classification and identification of natural objects. *IFC93, Lectures Notes in Artificial Intelligence,* Springer Verlag, 308-315.

MILNE-EDWARDS H. & HAIME J., 1860 - Histoire naturelle des Coralliaires T.III, Paris, 219 p.