Ingénierie des connaissances en Sciences de la vie : application à la systématique des coraux des Mascareignes

Noël Conruyt* — David Grosser* — Gérard Faure**

** IREMIA, Université de La Réunion*
97715 St Denis Messag. Cedex 9
LA REUNION

*** ISIM, Université de Montpellier II*
Pl. E. Bataillon, 34095 Montpellier

RESUME. *Dans les Sciences de la vie, les connaissances que l'on veut représenter, gérer et transmettre sont complexes. Par exemple en systématique, discipline scientifique dont l'un des buts est de décrire et d'étudier la diversité des êtres vivants, les descriptions d'espèces et de spécimens sont le plus souvent structurées, variables, erronées, etc.. Pour faire face à la description d'objets complexes, des méthodes d'ingénierie des connaissances se sont développées autour de l'apprentissage automatique et de l'analyse des données symboliques. Notre problématique est de concevoir un système robuste d'aide à la description, à la classification, et à l'identification des objets biologiques. Pour y répondre, nous proposons une méthodologie de construction de bases de connaissances évolutives fondées sur un processus itératif d'acquisition, de gestion et de transmission des connaissances. Une application sur les coraux des Mascareignes (Maurice, Réunion, Rodrigues) nous sert d'illustration. Nous discutons ensuite du problème du stockage, de la gestion et de la transmission des connaissances au sein de cette application, et nous montrons en quoi notre système permet de renforcer sa robustesse.*

MOTS-CLES : *systématique, modèle descriptif, questionnaire, description, induction, raisonnement par cas, robustesse, coraux des Mascareignes.*

1. Introduction

Dans les domaines industriels, il se développe actuellement des bases de données (relationnelles ou à objets) associées à des outils de traitement pour l'aide à la décision (raisonnement par cas, data mining) [ALT 95]. Ces derniers mettent l'accent sur le traitement de grands tableaux de données, mais dont la complexité de représentation n'est pas aussi élevée que celle trouvée dans les domaines naturels. En outre, leurs connaissances sont avant tout des "constructions humaines" établies à partir d'un modèle conceptuel fondé sur la déduction. D'autres méthodes proposent une démarche de modélisation comportementale adaptée à la résolution de problèmes industriels. Par exemple, CommonKADS structure les connaissances à différents niveaux (stratégie, tâche, inférence, domaine) et utilise l'inférence déductive pour réaliser des objectifs [WIE 92].

Inversement dans les Sciences de la vie, les experts doivent appliquer une méthode expérimentale de nature inductive fondée sur *l'observation des faits, la constitution d'hypothèses, et des tests expérimentaux pour les mettre à l'épreuve*. De plus, les

connaissances naturelles sont riches, variées (multiples exceptions), évolutives, dépendantes des techniques d'observation et soumises à des observateurs hétérogènes. Notre objectif est de permettre l'expression de cette diversité, de valoriser le travail de l'expert en l'aidant à mieux maîtriser son domaine, de transmettre et utiliser ses connaissances, et de mieux comprendre les systèmes naturels.

Le cogniticien doit donc adapter les outils à la réalité du domaine (et non l'inverse !), c'est-à-dire comprendre la démarche expérimentale des naturalistes. Ensuite, il doit intégrer le fait que la robustesse des résultats de classification et de détermination découle de la qualité des descriptions soumises au traitement (robustesse des descriptions).

Dans [CON 94], nous donnons une définition de la robustesse que nous qualifions d'empirique car basée sur les pratiques des utilisateurs : c'est l'ensemble des facteurs qualitatifs (compréhension, précision, exhaustivité, cohérence, ergonomie, redondance, tolérance aux bruits, fiabilité, mise à jour, objectivité, etc.) qui améliore l'acquisition des connaissances sur le domaine ou encore permet d'éliminer certaines faiblesses liées à l'utilisation des outils. Au cours de l'article, nous illustrerons ces différents facteurs de la robustesse par des exemples.

2. La démarche expérimentale des naturalistes

Tout au long de sa recherche, le systématicien apprend à reconnaître les espèces à partir d'un travail de bibliographie, de missions sur le terrain, dans les muséums, d'observations effectuées en laboratoire et d'échanges d'informations avec d'autres chercheurs. Il construit progressivement un modèle de description de son domaine qu'il applique à de nouvelles observations. Les erreurs d'identification l'amènent à remettre en cause son propre modèle. Il acquiert ainsi petit à petit une intuition du domaine, qui l'élève progressivement au rang d'expert reconnu par la communauté scientifique.

Ce cheminement procède par un aller-retour entre les informations reçues et son savoir-faire : en effet, l'expérience se nourrit de l'apport des nouvelles informations qui évoluent au fur et à mesure de l'amélioration des techniques d'observation (observations morphologiques au microscope, caractères biochimiques, génome, etc.). Néanmoins, avant de pouvoir obtenir de l'information sur un individu, il faut d'abord l'identifier par un nom. Ceci justifie le travail de classification préalable des échantillons à partir de leurs descriptions.

Or, pour les coraux, les descriptions sont fondées principalement sur un nombre insuffisant de caractères. La taxonomie de ces animaux s'effectue à partir des traits morphologiques du squelette, celui-ci présentant par ailleurs une très grande plasticité induite par les caractères environnementaux (hydrodynamisme, éclairement, etc.) [VER 76]. Ceci explique les différences d'appréciation sur le nombre d'espèces reconnues au niveau mondial par les différents spécialistes (de 500 à 900), du fait même du concept incertain de l'Espèce pour certains groupes. Les classifications évoluent donc au cours du temps en fonction des informations accessibles au systématicien. Par ailleurs, dans le cas des coraux, la plupart des individus sont regroupés au sein d'une colonie, qui résulte elle-même du bourgeonnement d'un individu souche. Ceci pose des problèmes de représentation et de traitement des connaissances.

Une première étude a été menée sur le genre *Pocillopora* [CON 96]. Il possède 5 espèces dans les Mascareignes dont l'une d'elle (*P. damicornis*) est représentée par 5 écomorphes (variétés adaptées à un certain type de milieu) [FAU 82].

L'objectif du système de gestion de la base de connaissances est d'apporter une aide à la description, à la classification et à l'identification des coraux des Mascareignes. Nous souhaitons aussi plus généralement contribuer à une recherche méthodologique sur l'informatique appliquée à la systématique afin de développer le dialogue entre chercheurs de ces deux disciplines (facteur de **compréhension**) [LEB 96].

3. Méthodologie de transfert des connaissances

Nous avons distingué deux types de connaissances, les informations contextuelles et le savoir-faire de l'expert. La base de connaissances que nous construisons est constituée par un noyau d'expertise en relation avec un ensemble d'informations de différentes natures, relatives au domaine (Fig. 1).

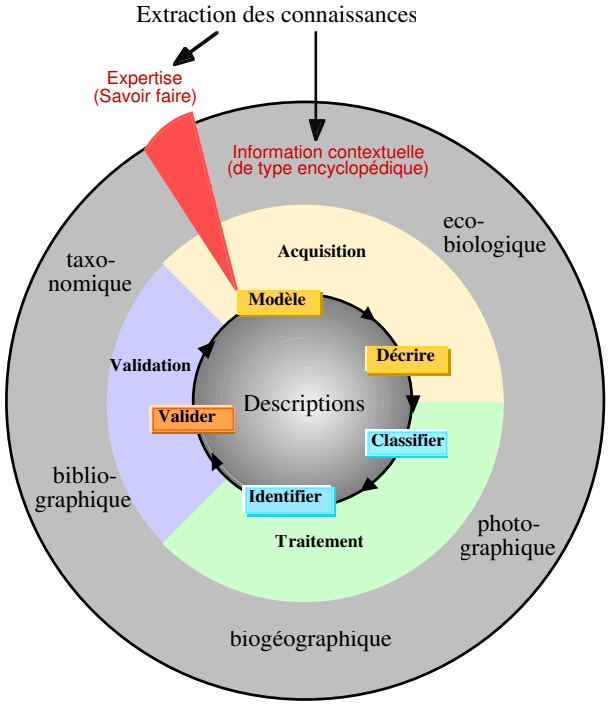

Figure 1. *Le processus itératif de transfert d'expertise*

3.1. Les informations contextuelles

L'expert dispose d'un ensemble d'éléments, qui peuvent être des références à des travaux et ouvrages existants, des données photographiques, géographiques, etc.. Même si ces connaissances n'interviennent pas directement dans le processus de description, de

classification et d'identification, elles alimentent la réflexion de l'expert en apportant une quantité de renseignements sur les spécimens.

Les nouvelles technologies hypermédias offrent un support idéal pour organiser ces différents types d'information sous forme de liens hypertextes [VAN 96].

Le système d'identification peut faire référence à ces informations pour aider l'utilisateur à interpréter correctement les questions posées. Par le biais de commentaires, d'illustrations ou de définitions de termes spécifiques au domaine (lexique), certaines ambiguïtés et imprécisions peuvent être levées.

Nous considérons donc qu'elles sont partie intégrante de la base de connaissances.

3.2. L'expertise

Dans le but de mettre en œuvre la méthode scientifique en biologie (conjecturer et tester) [POP 73], notre méthodologie suit le processus naturel d'apprentissage des connaissances par un expert. Celle-ci est divisée en trois étapes :

- Acquisition des connaissances,
- Traitement des connaissances,
- Validation.

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*).

3.2.1. Acquisition du modèle descriptif

Le modèle descriptif représente tout ce qui est *observable* pour notre domaine d'étude. Sa définition 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*. (*Pocillopora*). 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, itération (multi-instanciation), conditions contextuelles, etc. [CON 94], [LE R 96]. D'autres approches d'acquisition de connaissances plus "cogniticiennes" [AGU 89], [AIM 94] représentent la description du domaine par des schémas ou des graphes conceptuels. Pour notre part, nous ne souhaitons pas mélanger les relations dans un même plan visuel, car nous estimons qu'il existe un ordre sémantique entre les relations (temps, composition, spécialisation, instanciation). Cet ordre servira de guide à l'observation des spécimens dans le questionnaire.

Les nœuds de l'arbre correspondent aux objets (appelés également parties ou encore composants observables). Chaque objet est décrit à l'aide de caractères ou attributs typés, pouvant admettre des valeurs symboliques, numériques, uniques, multiples (variables), ordonnées, structurées ou imprécises. Il faut offrir une certaine souplesse d'expression à l'expert avec un langage de représentation des connaissances suffisamment puissant : logique multi-valuée, avec variables (ordre 1), taxonomies de valeurs, démons entre objets du modèle, etc. (facteur d'**exhaustivité**).

Par exemple, certains objets peuvent être absents (signalé par un signe négatif) comme l'objet "calices des verrues" (Fig. 2). La présence d'autres objets comme les septes, la muraille dépend de la présence du premier : cette connaissance de fond (règle d'inapplicabilité) doit être explicitement représentée afin de pouvoir assurer la cohérence de la phase de description (facteur de **cohérence**) (Fig. 3).

La structuration est un moyen efficace de prendre en compte des connaissances de fond de bon sens (dépendance entre objets) qui peuvent être utiles pour stocker, gérer et traiter les descriptions [ALL 84].

Nous avons montré que l'acquisition du modèle descriptif est la phase la plus importante de la méthode [CON 94] : la robustesse des résultats de classification et d'identification résulte de la qualité des descriptions traitées, et donc d'un modèle descriptif bien conçu. Par exemple, la notion d'**homologie** est essentielle en biologie : elle permet de s'assurer que l'on ne compare que des objets comparables (facteur de **précision**). C'est pourquoi nous permettrons de décrire différentes sortes de calices au sein de la colonie, chaque sorte étant spécialisée en fonction de son emplacement (verrues, apex, branches).

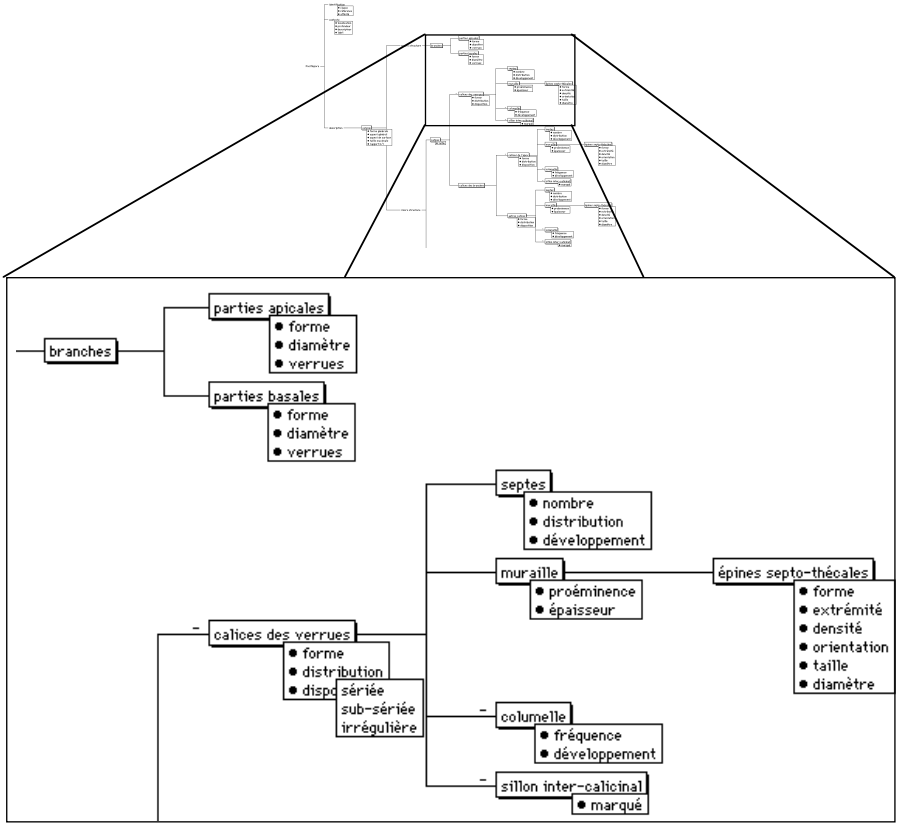

Figure 2. *Une partie du modèle descriptif des Pocillopora*

3.2.2. Génération automatique d'un questionnaire

Partant de la définition du modèle descriptif dans un formalisme de représentation utilisant la syntaxe BNF, un programme génère automatiquement un questionnaire. Des mots-clés permettent de repérer les objets et les relations. Ils sont ensuite interprétés en entités hypertextes (piles, cartes, boutons, champs) au moment de la création du questionnaire.

3.2.3. Acquisition des exemples (cas)

Le questionnaire généré permet à l'expert et aux autres biologistes d'acquérir une base de cas. Le questionnaire est personnalisable par l'expert. Chaque cas renseigné associe une description à une identification et constitue une *instance* de l'arbre de description (Fig. 3).

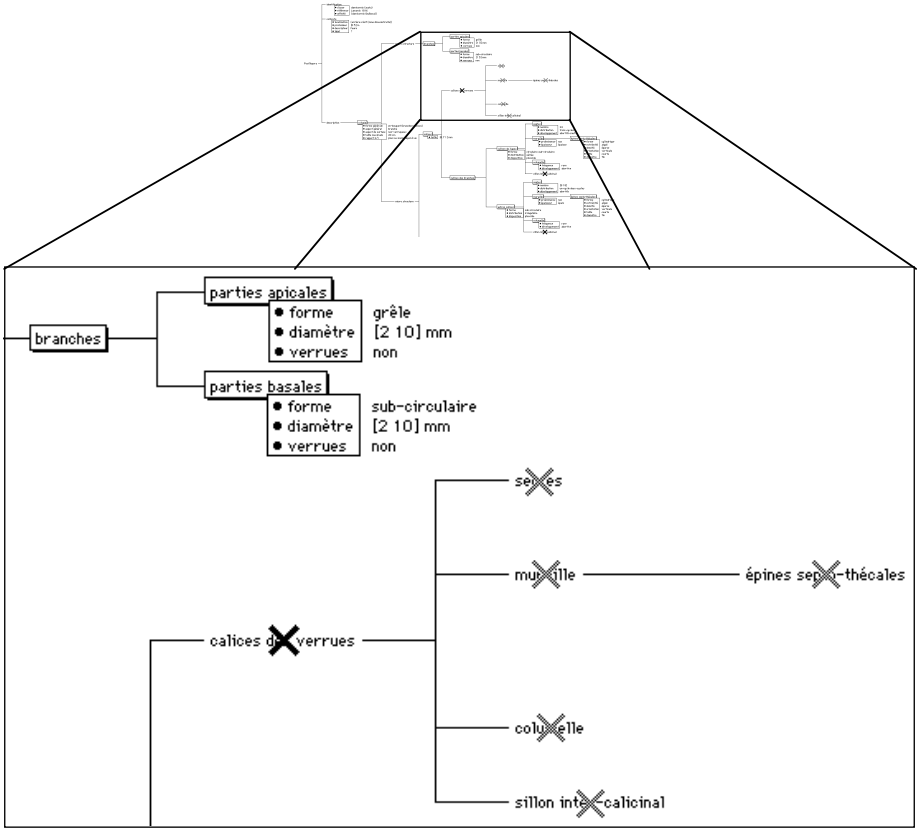

Figure 3. *Une partie instanciée de l'arbre de description de P. damicornis acuta*

Afin de représenter la diversité de l'observé, nous souhaiterons acquérir un grand nombre d'exemples par classe, de manière à privilégier les descriptions de spécimens (facteur de **redondance**) plutôt que les "descriptions" de concepts [VIG 91].

Le questionnaire suit la structure de l'arbre de description : il y a autant de cartes qu'il y a d'objets et d'attributs. Chaque carte est une vue locale d'un objet (Fig. 4). L'utilisateur peut naviguer entre les cartes en suivant le chemin des descendants et des parents étape par étape. Il peut passer d'une description locale (caractéristiques) à l'autre (composants). Il peut également sauter à un autre objet en passant par la vue globale (Fig. 3).

Les cartes des attributs sont les feuilles de l'arbre de description. Des commentaires et illustrations peuvent être associés par l'expert aux différentes valeurs afin d'aider l'utilisateur à interpréter correctement les questions posées (Fig. 5) (facteur d'**ergonomie**).

Lorsque l'utilisateur indique qu'il a fini sa description, une vérification des oublis éventuels a lieu en confrontant l'observé à l'observable afin d'assurer la cohérence de l'utilisateur vis-à-vis de ses réponses (différence entre l'oubli et l'inconnu) (facteur de **cohérence**).

Les descriptions observées résultantes sont des sous-arbres du modèle descriptif (Fig. 2 et Fig. 3). Ainsi, elles peuvent être directement comparées en faisant circuler les cartes de chaque description. Elles se superposent à celle du modèle descriptif, ce qui est visuellement plus facile que de comparer des listes de couples attribut-valeur dans un tableau de données.

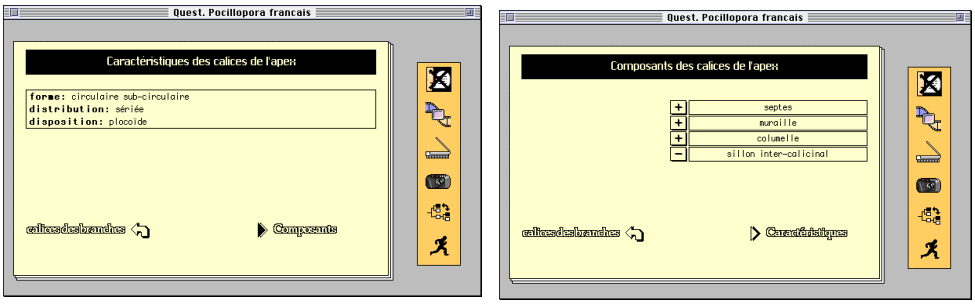

Figure 4. *Une vue locale de l'objet "calices de l'apex" chez P. damicornis acuta. Les caractéristiques de l'objet sont à gauche et les composants à droite.*

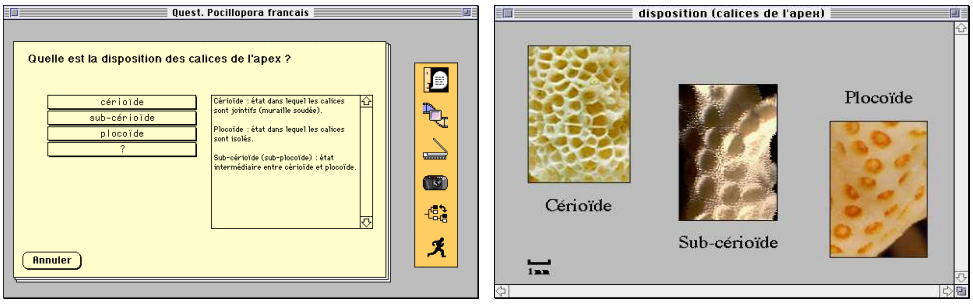

Figure 5. *Un exemple d'attribut commenté et illustré.*

Ces facilités d'utilisation (guide d'observation interactif, simplicité d'emploi pour des non-informaticiens) sont essentielles pour la constitution d'une base de cas robustes, ou bien encore la consultation du système expert résultant. En effet, nous avons pu observer de grandes différences d'interprétation de l'observation des utilisateurs par rapport aux experts. Nous devons donc mettre au point des outils pédagogiques de formation à l'expertise pour récupérer des descriptions les plus "proches" possibles de celles de l'expert, afin d'acquérir des résultats plus fiables encore (facteur de **fiabilité**).

3.2.4. Traitement des connaissances

En fonction du but à atteindre, deux types de méthodes sont utilisées : induction pour la classification, raisonnement à partir de cas pour l'identification.

Pour la **classification**, un arbre de décision est construit. A partir des descriptions (représentation en extension) une méthode inductive fondée sur la mesure d'entropie et du gain d'information [SHA 49], [QUI 86] fabrique une caractérisation de ces classes par un ensemble de règles. Chaque chemin depuis la racine vers les feuilles de l'arbre de décision est une règle de classification (également appelée *diagnose*).

Pour les *Pocillopora* , nous donnons à titre indicatif l'arbre de décision suivant (Fig. 6) qui classifie les 9 descriptions des types (espèces et écomorphes). Cet arbre de classification peut être utilisé pour déterminer une nouvelle observation. Néanmoins, lorsque l'utilisateur répond inconnu, la consultation de cet arbre est inadaptée [MAN 93].

Pour **l'identification**, le raisonnement à partir de cas (CBR) est utilisé [BAR 89]. 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 réponses de l'utilisateur. 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.

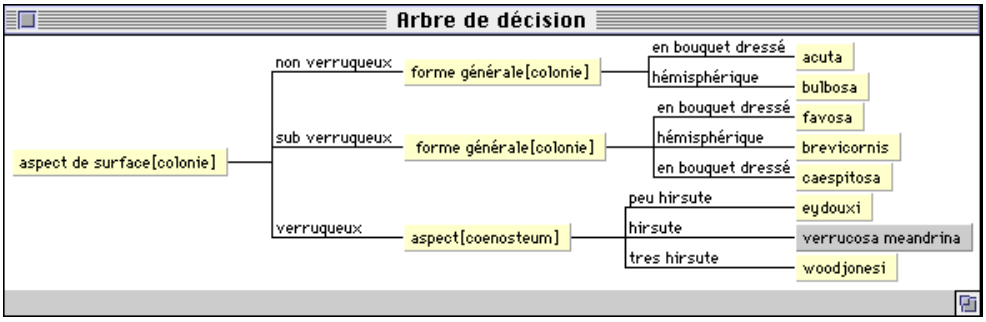

Figure 6. *Un arbre de décision pour classifier le genre Pocillopora*

Néanmoins, cette méthode monothétique d'identification ne permet pas de faire face aux erreurs de description. D'autres stratégies de CBR sont polythétiques (utilisent une combinaison de critères) et tolèrent les erreurs de description (facteur de **tolérance aux bruits**). Elles sont dérivées de la méthode *des k-plus-proches-voisins* en analyse de

données. Il s'agit d'une procédure de comparaison qui implique l'ensemble des attributs. Un score entre 0 et 1 donne un pourcentage de ressemblance entre deux cas. En ce qui nous concerne, nous sommes en train de développer une mesure de similarité pour la comparaison des descriptions structurées en biologie afin de tenir compte des dépendances entre caractères.

Pour la consultation, il existe un intérêt de combiner ces deux méthodes (induction et CBR) : différentes stratégies d'intégration ont permis d'obtenir de meilleurs résultats [AUR 94] : coopération, atelier, intégrée. Elles concourent à améliorer la qualité des consultations.

Certains auteurs en Sciences de la vie [PAN 91], [LEB 91], [DAL 93] ont aussi développé des systèmes d'identification tolérants aux erreurs de description, permettant d'utiliser les caractères dans n'importe quel ordre, prenant en compte les valeurs multiples pour gérer l'incertitude, etc.. Ces programmes s'appuient sur le format DELTA, langage de description taxinomique utilisé mondialement par de nombreux biologistes (botanistes).

L'aspect innovant que nous voulons développer est la mise en place d'une validation suivie par une phase d'itération sur le modèle descriptif et la mise à jour des anciens cas.

3.2.5. Validation

Avec l'aide de ces outils, l'expert peut évaluer les résultats de la classification et de l'identification, en fonction de la qualité de ses propres descriptions et de la définition du modèle descriptif. L'apprentissage inductif, tout comme l'utilisation répétée du questionnaire lui permettent de détecter certaines incohérences dans la base de cas et donc d'améliorer la base de connaissances (facteur de **mise à jour**).

Par cette expérimentation, le cogniticien peut aussi être amené à améliorer les algorithmes d'apprentissage pour ajuster le système aux besoins des utilisateurs : par exemple, dans l'arbre de décision de la figure 6, deux espèces très proches (*P. verrucosa* et *P. meandrina*) ne sont pas discriminées. Ceci peut être interprété par l'expert comme un résultat de classification intéressant car les auteurs ne s'accordent pas pour savoir si ces deux espèces n'en forment en fait qu'une. Néanmoins, les descriptions de ces espèces font ressortir quelques différences au niveau de deux attributs multi-valués (distribution et nombre de septes). Or, la présence d'états multiples a été interprétée par le cogniticien dans son algorithme comme une disjonction de valeurs due à l'imprécision et non pas à une conjonction de valeurs due à la variation intra-spécifique. Donc, un objet représentant un ensemble peut partager différents états simultanément (voir fig. 2 la forme des calices de l'apex : circulaire *et* subcirculaire). Cette connaissance de fond doit être traitée différemment : en cas de doute, nous devons nous retenir de discriminer, alors qu'en cas de variation, nous devons poursuivre la séparation des cas (c'est-à-dire trouver un autre critère après l'aspect hirsute du cœnosteum).

Ainsi, nous souhaiterons représenter la différence entre disjonction d'imprécision et conjonction de variation dans le modèle descriptif, ou encore paramétrer la mesure de ressemblance pour qu'elle tienne compte de la structure des descriptions (dépendance entre objets).

4. Discussion

Les outils proposés pour mettre en œuvre cette méthodologie sont les briques d'un atelier logiciel que nous somme en train de développer à l'IREMIA : IKBS (Iterative Knowledge Base System). L'aspect itératif est primordial car on ne peut exiger de l'expert une analyse exhaustive de l'observable dès le départ. Pour l'informaticien, cela suppose la conception d'outils de maintien de la cohérence de l'ancienne base de cas par rapport aux changements effectués dans un nouveau modèle descriptif (élimination-rajout d'objets, d'attributs ou de valeurs possibles, modification de la structure de description, etc.).

Nous souhaitons que cet atelier, qui s'appuie sur la démarche naturelle de l'expert, leur permette de mieux **stocker**, **gérer** et **transmettre** leurs connaissances :

4.1 Stocker : représenter et conserver les connaissances

L'étude de la diversité des êtres vivants est une source de difficultés à la fois au niveau de la représentation et du traitement de ces connaissances. L'informaticien est confronté à la complexité des individus à représenter et à leur nature. Une description synthétise-t-elle un ensemble d'individus (description d'espèce) ou bien est-elle le codage d'un spécimen ? Dans le premier cas, les valeurs multiples d'un caractère peuvent exprimer une disjonction de variation (l'espèce se trouve dans tel biotope ou tel biotope), ceci n'étant pas possible pour les valeurs multiples du deuxième cas qui expriment des états présents simultanément.

Les variations intra-coloniales possibles nous obligent à multiplier les sous-arbres descriptifs des calices, en fonction de leur localisation dans la colonie (apex, branches, verrues). Cette représentation va permettre de pouvoir comparer des objets homologues.

4.2 Gérer : manipuler, comparer, traiter les connaissances

Par l'utilisation de mesures de ressemblance sur ces descriptions, nous mettons en évidence les possibles variations intra-spécifiques (au sein des espèces) plus importantes que les variations inter-spécifiques (entre espèces). Par exemple, deux échantillons appartenant à deux espèces différentes présentent une convergence de forme induite par un facteur écologique ou environnemental dominant (adaptation morphologique). Inversement, une même espèce peut développer des formes différentes en réponse aux facteurs du milieu (écomorphes).

Un comptage précis des égalités et des différences de valeurs sur la centaine de caractères macro et microscopiques des descriptions du genre *Pocillopora* a permis de révéler un certain niveau d'affinité entre deux spécimens appartenant à deux espèces différentes (*favosa* et *eydouxi*). Bien que ces spécimens ne se trouvaient pas dans les mêmes milieux biotiques (par exemple en mode calme et battu) et donc extériorisaient des faciès très différents, ils montrent néanmoins une similitude importante au niveau des caractères microscopiques. Cette information nouvelle induite par la mesure de ressemblance utilisée pour comparer les exemples interpelle l'expert sur la possibilité des deux spécimens d'appartenir en fait à la même espèce.

4.3 Transmettre : faire "passer" l'expertise et la distribuer.

L'un des soucis majeur des systématiciens est à la fois de transmettre ses connaissances et de former de nouveaux systématiciens [DUR 96]. Dans ce sens, nous essayons de proposer des outils conviviaux d'aide à la formation à l'expertise en utilisant des interfaces qui correspondent aux besoins et aux habitudes des biologistes.

Par exemple, ceux-ci travaillent couramment sur des structures arborescentes (arbres phylogénétiques, clés d'identification). Ils n'ont donc pas de difficultés à appréhender un arbre de description ou de décision, ce qui ne serait certainement pas le cas avec d'autres méthodes de représentation (graphes, treillis, etc.).

De plus, nous nous sommes aperçus du problème de la construction d'un modèle descriptif par un expert isolé. Nous nous heurtons à une approche personnelle et parfois subjective du traitement des échantillons. La manière de décrire de l'intervenant est influencée par la littérature utilisée (qui est parfois contradictoire) et par l'interprétation personnelle des spécimens rencontrés. C'est pourquoi nous voulons introduire une dimension nouvelle dans notre méthodologie, la **Télésystématique** : il s'agit d'une démarche multi-experte de recherche collaborative à distance utilisant la visioconférence, et qui permet de trouver un consensus au niveau du choix du vocabulaire (thesaurus) et des illustrations (images, dessins) entre experts répartis dans le monde entier.

L'attrait de ce travail collaboratif à distance, c'est de permettre la concertation et une définition plus objective des caractères descriptifs des espèces (facteur d'**objectivité**), ceci afin d'améliorer la robustesse globale de la base de connaissances multi-experte.

5. Conclusion

Nous avons mis en œuvre une méthodologie et des outils permettant de modéliser, décrire, classifier, identifier et valider les connaissances des systématiciens.

Les outils accompagnent la démarche naturelle de l'expert et sont très bien accueillis. Ils permettent de rendre explicite son savoir-faire à l'aide de descriptions codées, de matérialiser l'acquisition inconsciente de ses connaissances, de retracer son cheminement conceptuel tout en gardant une mémoire de son travail passé. Parallèlement, l'expert évalue l'état de ses connaissances en les confrontant à celles produites par le système.

La robustesse intervient donc pendant toutes les phases du cycle de stockage, de gestion et de transmission des connaissances.

Il est fondamental que le cogniticien puisse travailler en étroite collaboration avec l'expert afin qu'il intègre toutes les contraintes sémantiques liées à la biologie. En retour, ce dernier ne doit pas voir le module de traitement comme une boite noire : il doit pouvoir recevoir à tout instant des explications pour pouvoir interpréter correctement un résultat.

Nous sommes en phase de généralisation du modèle *Pocillopora* (1 genre) à celui de la famille des *Pocilloporidæ* (4 genres). Dans cette phase itérative généralisante sur le modèle descriptif, nous introduisons des nouvelles valeurs, attributs et objets ainsi qu'une

modification de la structure de description. Cela nous oblige à refondre le modèle initial pour rester cohérent avec les nouvelles observations.

L'étude en cours de genres appartenant à d'autres familles (*Poritidæ*, *Fungiidæ*, *Agariciidæ*) et faisant intervenir quatre experts permet d'entrevoir l'extrême complexité du domaine à représenter. Nous essayons d'y répondre par notre approche ascendante de la modélisation en partant des spécimens. Ceci justifie la nécessité d'une collaboration pluridisciplinaire et la création d'un groupe d'experts au niveau international, reliés entre eux par des moyens de communication à haut débit.

6. Remerciements

Nous remercions chaleureusement N. Aussenac-Gilles pour ses précieux conseils.

7. Bibliographie

[AGU 89] AGUIRRE CERVANTES J.L., *Construction automatique de taxonomies à partir d'exemples dans un modèle de connaissances par objets*. Thèse de doctorat de l'I.N.P.G., 1989.

[AIM 94] AIMEUR E., *METIS : un système et une Méthode d'Explicitation de Taxonomies destinés à l'Identification de Structures conceptuelles*. Thèse de doctorat de l'Univ. Paris-VI., 1994.

[ALL 84] ALLKIN R., "Handling taxonomic descriptions by computer", In; Allkin R and Bisby FA (eds.), Databases in systematics. Systematics Association London, Academic Press, (26) pp 263-278, 1984.

[ALT 95] ALTHOFF K.D., AURIOL E., BARLETTA R., MANAGO M., *A Review of Industrial Case-Based Reasoning Tools*, AI Intelligence, 1995.

[AUR 94] AURIOL E., MANAGO M., ALTHOFF K.D., WESS S., DITTRICH S., "Integrating induction and case-based reasoning: methodological approach and first evaluations" EWCBR-94 - Second European workshop on case-based reasoning. M Keane, JP Haton & M Manago (Eds.), AcknoSoft Press, pp 145-155, 1994.

[BAR 89] BAREISS R., *Exemplar-based knowledge acquisition: a unified approach to concept representation, classification and learning*, London, Academic Press inc, 1989.

[CON 94] CONRUYT N., *Amélioration de la robustesse des systèmes d'aide à la description, à la classification et à la détermination des objets biologiques*. Thèse de doctorat en informatique, Univ. Paris-IX Dauphine, pp 1-281, 1994.

[CON 97] CONRUYT N., FAURE G., ANCEL G., LE RENARD J., GUILLAUME M., NAIM O., GRAVIER-BONNET N., "A Knowledge Base for corals of the Mascarene Archipelago: Genus Pocillopora" Proc 8th Int Coral Reef Symp Panama, 1997.

[DAL 93] DALLWITZ M.J., PAINE T.A., ZURCHER E.J., "User's guide to the DELTA System. A general system for processing taxonomic descriptions" Canberra: CSIRO, Div. Entomol., 4th ed., 1993.

[DUR 96] DURRIEU G., "L'informatique, un outil pédagogique pour enseigner la systématique ?" Biosystema 14, Informatique et Systématique, pp. 109-116, 1996.

[FAU 82] FAURE G., *Recherche sur les peuplements de scléractiniaires des récifs coralliens des Mascareignes*, Thèse es sciences, Univ Aix-Marseille II, (2) pp 1-206, 1982.

[LEB 91] LEBBE J., *Présentation des concepts en biologie et en médecine. Introduction à l'analyse des connaissances et à l'identification assistée par ordinateur.* Thèse de doctorat, Univ. Paris-VI, pp 1-282, 1991.

[LEB 96] LEBBE J. "Quelques réflexions sur l'informatique appliquée à la systématique en france", Biosystema 14, Informatique et Systématique, pp. 5-10, 1996.

[LE R 96] LE RENARD J., LEVI C., CONRUYT N., MANAGO M., "Sur la représentation et le traitement des connaissances descriptives : une application au domaine des éponges du genre *Hyalonema*", vol. 66 suppl., Biologie, Recent advances in sponge biodiversity and documentation, P. Willenz (Ed), Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, pp. 37-48, 1996.

[MAN 93] MANAGO M., ALTHOFF K.D., AURIOL E., TRAPHÖNER R., WESS S., CONRUYT N., MAURER F., "Induction and reasoning from cases", First European workshop on case-based reasoning (EWCBR-93), MM Richter, S Wess, KD Althoff and F Maurer (Eds.), Springer Verlag, (2), 1993.

[PAN 91] PANKHURST R.J., *Practical taxonomic computing*. Cambridge University Press, Cambridge, pp 1-202, 1991.

[POP 91] POPPER K.R., *La logique de la découverte scientifique*. Payot (Eds.) Press, Paris, 1973.

[QUI 86] QUINLAN J.R., *Induction of decision trees*, Machine Learning 1 : 81-106, 1986.

[SHA 49] SHANNON C.E., *The mathematical theory of communication*, University of Illinois Press, Urbana, 1949.

[VAN 96] VAN SOEST R.W.M., SCHALK P.H., SMITH K., PICTON B.E., BRUGMAN M., DIAZ M., SANDERS M.L., DE WEERDT W.H., RÜTZLER K., "POR-LINNAEUS: The application of interactive multimedia software for species data storage and computer assisted identification of Porifera", vol. 66 suppl., Biologie, Recent advances in sponge biodiversity and documentation, P. Willenz (Ed), Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, 1996.

[VER 76] VERON J.E.N., PICHON M., *Scleractinia of eastern australia*, vol. I, Part I, Australian Institute of Marine Science Monograph Series, 1976.

[VIG 91] VIGNES R., *Caractérisation automatique de groupes biologiques*. Thèse de doctorat, Univ. Paris-VI, pp 1-260, 1991.

[WIE 92] WIELENGA B., SCHREIBER A., BREUKER J., "KADS : a modelling approach to knowledge engineering", in Knowledge Acquisition, vol. 4, 1992.