

Sur la représentation et le traitement des connaissances descriptives : une application au domaine des éponges du genre *Hyalonema*

par Jacques LE RENARD, Claude LÉVI, Noël CONRUYT & Michel MANAGO

Abstract

Sponges of the genus *Hyalonema* were used as an application domain for conceiving, developing and evaluating softwares that cover a great part of the taxonomist needs: acquiring and managing descriptions, processing characters for classification purpose, building expert systems based on induction or case-based reasoning technologies for identification purpose. All these tools are relying on a new representation formalism of descriptions, which is naturally structured. The structure is elaborated by representing all what is observable in a descriptive model. The multiple interests of this methodological approach are discussed, in particular its affinity to high quality information exchanges between researchers, through networks. Some already available software tools are briefly presented.

Key-words: knowledge representation, description, classification, identification, induction, case-based reasoning, Porifera.

Résumé

Les Éponges du genre *Hyalonema* ont servi de domaine d'application pour la conception, le développement et l'évaluation de logiciels couvrant une grande partie des besoins des systématiciens: saisie et gestion de descriptions, manipulation des caractères à des fins de classification, systèmes experts de détermination fondés sur l'induction ou le raisonnement sur cas. Tous ces outils s'appuient sur un nouveau mode de représentation des descriptions, sous forme naturellement structurée. Cette structuration est élaborée par la représentation de tout l'observable sous forme d'un modèle descriptif. Les intérêts multiples de l'approche méthodologique proposée sont discutés, en particulier son adéquation à des échanges d'information de haute qualité entre chercheurs, via les réseaux. Les outils logiciels déjà disponibles sont rapidement présentés.

Mots-clefs : représentation des connaissances, description, classification, détermination, induction, raisonnement sur cas, Porifera.

Introduction

Toutes les tâches ayant trait aux sciences d'observation ont recours aux descriptions. La connaissance ne peut en effet s'accommoder d'observations non formalisées, et implique une

abstraction plus ou moins consciente de la réalité, sous forme de descriptions. S'il peut paraître trivial d'affirmer ainsi le rôle fondamental des descriptions dans tous domaines des sciences naturelles, il l'est moins de remarquer que leur réalisation même est toujours restée dépendante du "savoir-faire" individuel du biologiste. On apprend à lire, à compter, mais pas à décrire ; probablement est-ce considéré comme inné. Pourtant, dès lors qu'on prend les descriptions comme un sujet d'étude, on doit bien constater qu'elles ne constituent des supports de connaissances qu'empiriques et imparfaits.

Il n'y a là aucune fatalité, et nous nous emploierons ici à démontrer que, à l'ère de l'informatique, des solutions peuvent être trouvées pour rendre aux descriptions toute leur puissance, en particulier pour les adapter au mieux aux problèmes de classification et d'identification.

Certaines de ces solutions ont été testées sur un petit domaine réel, celui des Spongiaires du genre *Hyalonema* GRAY, 1832 (Fig. 1), choisi pour être bien maîtrisé par l'un de nous et dont la taille et la complexité sont représentatives de nombreux autres domaines tant zoologiques que botaniques. Afin de faire ressortir les difficultés que rencontre le systématicien dans son travail de tous les jours, nous partirons d'une analyse de la pratique concrète de la détermination ; les solutions logicielles appropriées pourront alors être présentées.

LA PRATIQUE DE DETERMINATION

Toute détermination d'un objet vivant ou inanimé commence par une phase essentielle de description de ce spécimen. Ensuite les éléments d'observation obtenus sont comparés aux descriptions préexistantes, souvent synthétisées en diagnoses ou présentées sous forme de clefs. La comparaison peut aussi porter sur l'examen direct de spécimens déjà déterminés en collection. L'objectif est l'identification du groupe taxinomique auquel appartient le spécimen, sous forme d'un nom scientifique. Le système binominal dû à Linné permet à tous de parler de



Figure 1. - Un individu du genre *Hyalonema* GRAY, 1832.

la même chose, dans la mesure où chaque binôme est correctement employé.

Or la taxinomie des différents groupes d'organismes se trouve à des niveaux d'élaboration très variable, en fonction de la complexité du groupe et du nombre de ses spécialistes. Le problème de la qualité de la taxinomie est maximal pour les taxons les plus mal connus, dont les Spongiaires constituent un bon exemple ; c'est dans ce cas qu'il apparaît le plus important de disposer d'une base d'analyse des caractères qui soit acceptée par tous les spécialistes concernés.

On conçoit que la taxinomie s'affine grâce à l'évolution des techniques d'observation. Ainsi, la liste des caractères à prendre en compte lors des déterminations se complète progressivement : caractères externes, puis internes macroscopiques puis microscopiques, biologiques puis chimiques, adultes puis larvaires, géographiques puis écologiques, etc. Combien de caractères encore ignorés seront-ils utilisés à l'avenir ?

Car la détermination exige la connaissance d'une diversité de plus en plus grande de caractères indépendants, et celle de la variabilité de chaque caractère, dont on tente de préciser les états. Le nombre d'opérations variées à accomplir augmente, et le temps nécessaire à l'identification s'accroît corrélativement : collecte de spécimens vivants, préparations de spicules, histologie, mesures et calculs statistiques, dessins, photographies, microscopie électronique, chimie des produits majeurs, séquences ...

Des états partiels de la connaissance sont de temps en temps dressés par le spécialiste d'un genre ou d'une famille. Ces monographies peuvent donner l'impression d'un savoir sûr et définitif ; cependant, du fait de la constante évolution des caractères pris en compte et de leur variabilité,

toute connaissance instantanée est appelée à être réexaminée ou complétée. Par voie de conséquence, les noms scientifiques eux-mêmes voient leur sens glisser avec le temps : les spongiologues percevront tous l'évolution qui est advenue à des noms comme *Spongia* ou *Halichondria*. Dans ces conditions, il y a lieu de se méfier du confort qu'apporte l'utilisation des binômes linnéens, qui, pour pratiques qu'ils soient, peuvent néanmoins masquer de réelles ambiguïtés, voire des erreurs de détermination.

La pratique idéale exigerait une connaissance exhaustive des spécimens eux-mêmes, mais l'extrême dispersion géographique (et linguistique) des musées et des bibliothèques rend ce travail impossible. A défaut, il faudrait pouvoir disposer d'une base complète de descriptions détaillées de tous les spécimens connus. Cet objectif n'est pas irréaliste, mais il exige un important travail collectif entre spécialistes, avec un minimum de méthode qu'il convient de définir.

La première tâche à accomplir, avant même de procéder aux descriptions, est de dresser un état des lieux, afin d'en dégager une base commune de vocabulaire (thésaurus), une sélection des caractères à prendre en compte, enfin la reconnaissance des liens divers entre ces caractères.

En effet, les connaissances en biologie ont pour spécificité de porter sur des objets naturels, à la fois complexes, variables et plus ou moins partiellement inaccessibles (fréquence des "données manquantes"). Tout mode de représentation doit d'abord avoir la capacité de tenir compte de cette "richesse" du vivant, sans la tronquer ni la biaiser. Dans l'idéal, une description scientifique doit être parfaitement objective, donc à la fois complète et exacte ; bref, une abstraction

équivalente à la réalité. Dans la pratique, on peut se satisfaire d'une représentation "épistémologiquement valable" (conforme à la vérité scientifique), ce qui permet de s'arrêter au degré de finesse qu'on juge adéquat, en fonction de son objectif et des techniques d'investigation dont on dispose.

C'est donc au groupe d'experts taxinomistes spécialiste d'un domaine biologique particulier d'énoncer quelles sont les caractéristiques qui doivent être prises en compte dans les descriptions. Actuellement, cette tâche préalable n'est que très rarement identifiée. Dans le meilleur des cas, elle est explicitée sous forme d'une liste de caractères illustrée. Une partie des connaissances disponibles concernant l'observable est généralement ignorée, en particulier les relations naturelles entre les caractères.

REPRESENTATION INFORMATIQUE DES DESCRIPTIONS

L'utilisation de l'informatique pour aider le travail du systématicien n'est pas nouvelle (voir l'état de l'art dans PANKHURST, 1991). Au cours du temps, elle a évolué vers des outils de plus en plus finement adaptés aux exigences scientifiques. L'approche que nous présentons ici s'est imposé de préserver à tous les niveaux de traitement le même degré de détail. Elle respecte aussi les divers types de savoir-faire mis en jeu lors de l'activité de description. L'objectif technique est de parvenir à des outils qui n'imposent qu'un minimum de contraintes d'emploi, tout en aidant l'utilisateur systématicien à améliorer sa rigueur et à formaliser son savoir.

La représentation des descriptions sous forme informatique s'est ainsi appuyée sur une analyse des différents mécanismes descriptifs, sur la nature des objets à décrire, et surtout sur la reconnaissance explicite de tout ce qui peut être observé donc décrit.

Pour fixer les idées, il est nécessaire de limiter la réflexion à un domaine bien circonscrit. Les éponges du genre *Hyalonema*, qui nous ont servi de support à des développements plus génériques, illustrent bien la possibilité de scinder la connaissance en domaines "élémentaires" plus aisément appréhendables par les spécialistes. C'est l'étude collective de la conformation de ce domaine qui va permettre d'en établir le modèle descriptif, clef de voûte de toute notre approche.

Il est en effet fondamental, pour des spécialistes isolés dans l'espace et dans le temps, de pouvoir s'appuyer sur une base descriptive commune, acceptée par tous mais modifiable. C'est une condition nécessaire pour parvenir à une taxinomie cohérente, au moins pour les groupes d'organismes encore imparfaitement connus.

LE MODELE DESCRIPTIF

Nous avons introduit la notion de modèle

descriptif (LE RENARD & CONRUYT, 1994) pour adapter aux connaissances descriptives ce qui était désigné comme "théorie du domaine" ou comme "background knowledge" dans des domaines plus généraux.

Il ne faut pas confondre le modèle descriptif, qui est une représentation des connaissances (abstraites) sur l'observable, et les descriptions, qui sont des connaissances (concrètes) sur l'observé.

Le terme de modèle désigne ici une représentation abstraite, organisée (structurée sous forme d'un arbre), mais sans connotation fonctionnelle. On peut considérer le modèle descriptif d'un domaine comme le plan d'organisation des descriptions portant sur l'ensemble des individus couverts par ce domaine.

Il existe bien sûr un lien réciproque entre l'observable et l'observé : l'observable n'est accessible que par une étude de l'observé connu ; et l'observé correspond, pour un individu donné, à une instanciation ou restriction de l'observable connu. Cette relation circulaire implique d'ailleurs que la progression des connaissances descriptives ne peut résulter que d'une confrontation réciproque du monde observé et du monde observable.

L'expérience de la réalisation d'un modèle descriptif valable pour le genre *Hyalonema*, montre qu'il s'agit d'un travail essentiel d'expertise. Il a exigé une relecture exhaustive de toute la littérature, la recherche de tous les caractères utilisés peu ou prou par les auteurs, et le réexamen critique de ces caractères, à la lumière de leurs dépendances réciproques, considérées non pas comme un biais mais comme une richesse informative.

La prise en compte de la structuration des connaissances (voir par exemple DALLWITZ, 1980, ALLKIN, 1984) a constitué un progrès très comparable à celui réalisé lors du passage des fichiers de données à des bases de données (hiérarchiques, puis relationnelles, puis orientées-objets). La structure prend en charge une quantité de connaissances de fond, souvent qualifiées "de bon sens", et qui se révèlent d'une grande puissance dans les algorithmes de saisie, de gestion ou de traitement des connaissances descriptives (à condition, bien sûr, que ceux-ci soient suffisamment "intelligents" pour prendre en compte la structure).

L'importance de disposer d'un modèle descriptif étant établie, notre problème courant devient : étant donné un domaine particulier dont je suis spécialiste, comment en établir un bon modèle ?

LES LOGIQUES DESCRIPTIVES

De même que pour la conception des bases de données relationnelles il existe un certain nombre de guides appelés "règles des formes normales", on peut formaliser au travers des modèles descriptifs un ensemble de façons naturelles

d'observer. Nous les avons appelées "logiques descriptives" (LERENARD & CONRUYT, 1994 : 310-313).

Le point commun de toutes ces logiques est de faire passer du plus général au plus particulier ; on les met en œuvre récursivement jusqu'à ce qu'on ait atteint le niveau de finesse désiré, appelé granularité de l'information ou atomicité de la connaissance.

Les logiques descriptives sont surtout perceptibles quand on analyse la méthode employée par le descripteur (la personne qui effectue une description). En effet, à chaque logique, abstraite donc appartenant au monde de l'observable, s'attache un mécanisme de description, concret donc plus facile à saisir.

Le point de départ du processus de description est l'entité observable du domaine ; il constitue la racine de l'arbre représentant la structure. Par exemple, dans le cas des Éponges, l'entité observable est "éponge", c'est-à-dire l'éponge universelle, toute éponge, éponge. Chaque fois que le modèle descriptif sera utilisé pour générer une description, l'entité observable sera instanciée en une entité observée, une éponge en tant qu'individu, 1 éponge. Nous verrons plus loin ce qu'il faut entendre par individu, mais il faut retenir que, par le caractère objectif que nous lui imposons, une description ne s'applique qu'à un individu (spécimen) unique et non pas à un groupe.

L'exposé qui suit s'appuiera sur le domaine réel des *Hyalonema*, bien que cela limite un peu sa compréhension aux spécialistes des Spongiaires, du fait du vocabulaire spécialisé employé.

Reconnaissance de traits ("features").

L'observation consiste essentiellement en l'évaluation d'un certain nombre de traits morphologiques ou autres, selon une simple logique analytique.

Nous emploierons le terme d'objet pour désigner toute partie d'individu séparément observable, dès lors qu'elle peut être l'objet d'une description. Chaque objet a son propre modèle descriptif, et peut être représenté formellement par la liste de ses attributs (ou propriétés) et par la liste de ses sous-objets (de ses "descendants") ; chaque attribut, par exemple la forme, la taille ou la couleur de l'objet, peut prendre un certain nombre de valeurs (on parle aussi de caractères et d'états de caractère).

La représentation sous forme de "frames" (terme sans équivalent français satisfaisant), où chaque trait observé est traduit par un triplet "objet attribut valeur", semble bien adaptée aux descriptions. Elle permet de passer assez naturellement du langage de l'informaticien à celui du biologiste.

Trait : L'attribut A de l'objet O a pour valeur V, formalisé en :

AOV : [Attribut(Objet)=Valeur]

Pour faciliter la distinction grammaticale entre les trois niveaux O, A et V, on pourra se guider sur les règles empiriques suivantes :

- Objet : Entité, ou partie séparément observable. Présente, absente ou inconnue.
- Attribut : Nom (ex. : forme, couleur, taille, etc.). Peut s'appliquer à plusieurs objets.
- Valeur : Qualificatif (ex. : allongé, blanchâtre, de 10 mm, etc.). Mesure d'un attribut.

Par exemple, pour représenter le trait "pédoncule torsadé", on utilisera :

AOV : [forme(pédoncule)=torsadée]

On aurait certes pu écrire :

[torsadé(pédoncule)=oui], mais, comme torsadé est une valeur et non un attribut, on se serait fermé la possibilité ultérieure d'admettre d'autres valeurs pour la forme du pédoncule, comme cela paraît souhaitable.

L'opérateur d'affectation = employé ci-dessus représente le mécanisme dit d'attribution.

Mécanisme de décomposition progressive en sous-parties

La logique de composition, ou "part-whole logics", ou "meronymic relationships" (CRUSE, 1986), exprime simplement le fait que l'individu est un tout (holonyme : cf. p. 418) décomposable en différentes parties (méronymes), chacune de ces parties étant un tout décomposable à son tour.

"An oriole is a type of bird (class inclusion), has wings (meronymy), and is brightly colored (attribution)". "Meronymic relationships structure semantic space in a hierarchical fashion" (WINSTON *et al.*, 1987 : 418, 420).

Par exemple, n'importe quel individu de *Hyalonema* est composé d'un corps et d'un pédoncule (Fig. 1) ; le pédoncule n'est pas décomposable, alors qu'on peut distinguer dans le corps une face latérale et une face exhalante, cette dernière pouvant comporter (ou non) un cône central, et ainsi de suite.

On conviendra que cet agencement des différentes parties et sous-parties correspond à un plan d'organisation naturel et facilement perceptible. Il peut être représenté par un arbre de parties, très simple dans le cas des *Hyalonema* (Fig. 2). On notera par exemple l'absence de la partie "face latérale", considérée *a priori* comme dépourvue de pouvoir discriminant par l'expert du domaine (Cl. LEVI).

L'aspect récursif de la décomposition apparaît clairement dans la représentation sous forme d'arbre. On pourra ainsi se représenter le modèle descriptif des *Hyalonema* comme un agencement de sous-modèles descriptifs, chacun portant sur un "individu" réduit à une sous-partie. En transposant à une description réalisée en suivant le

modèle descriptif, on pourra dire qu'elle sera une conjonction de descriptions locales (des différentes sous-parties) ; à condition de ne pas oublier d'y adjoindre l'information portée par l'arbre lui-même, à savoir les liens qui existent entre les sous-parties :

Description = AOV_1 AOV_2 ... AOV_n
relations structurelles entre les objets

Sous sa forme la plus simple, un modèle descriptif revêt donc l'aspect d'un arbre dont chaque nœud est un objet. Ce mode de représentation est comparable aux façons classiques de présenter les caractères observables et leurs états, que ce soit sous forme d'une liste textuelle ou sous forme d'un tableau (par exemple : forme du corps = {arrondie, aplatie, allongée} ; diamètre maximal du corps = {20 - 200} mm ; etc.). Le contenu informatif semble équivalent dans les deux cas.

Pourtant, le fait de prendre en compte la structure naturelle des données comporte un avantage. En effet, les liens de la logique de composition permettent des raisonnements de bon sens très utiles. Par exemple, si l'on affirme lors d'une description qu'il n'y a pas d'amphidisque (constituants microscopiques du corps de certaines éponges siliceuses), il n'y a plus lieu de s'intéresser aux sous-parties de l'objet amphidisque : tout l'arbre descriptif correspondant à cet objet est simplement tronqué (objet A absent pas de description de A). Il s'agit d'un syllogisme évident (si objet A absent, et si objet B = sous-partie de A, alors B est absent), appelé "merological syllogism" par WINSTON *et al.* (1987 : 417). Cette simplification de la formulation obéit au "irrelevance principle" de SUBRAMANIAN (1990). Par opposition, dans les tableaux de données, on serait contraint de saturer toutes les cases correspondant à la description des amphidisques par une pseudo-valeur "NA" (non applicable), qui ne fait que refléter l'inadéquation de ce mode de représentation.

De même, lorsqu'un objet n'est pas observable, on ne peut le décrire. Par exemple, si le pédoncule de l'éponge a été arraché lors de la récolte, on se bornera à indiquer le statut "inconnu" pour l'objet pédoncule, ce qui tronquera *ipso facto* son sous-arbre descriptif.

Ces avantages procurés par la structuration des données tiennent au fait qu'à chaque objet est associé un état trivalué : présent, absent, inconnu. Seule la présence d'un objet permet d'en faire la description. Et une distinction très nette est faite entre l'absence d'un objet (il s'agit d'une propriété qui peut être distinctive et discriminante), et le fait que l'objet ne puisse pas être observé dans les conditions présentes.

Il est aussi possible que la présence / absence d'un objet n'ait aucune signification classificatoire. Ainsi, dans le cas du pédoncule du genre

Hyalonema, l'état "absent" a été interdit par l'expert du domaine : une *Hyalonema* a toujours un pédoncule ; si on ne l'observe pas, c'est qu'il a été arraché, et il n'est donc jamais absent mais inconnu. Cette connaissance experte, utilement incluse dans le modèle descriptif, évite que deux éponges soient artificiellement séparées parce que l'une a son pédoncule arraché et l'autre non.

Mécanisme d'expression de l'imprécision

Il suffit de parcourir quelques descriptions, parmi les millions qui ont été publiées, pour se convaincre de l'omni-présence de l'imprécision. Pour rester le plus objectif possible, le descripteur est amené à apporter de multiples nuances ou pondérations. Le respect de cette finesse d'expression est aussi nécessaire que compliqué à formaliser.

Pour le moment, nous avons renoncé à traiter séparément les différentes sources d'imprécision dans les observations, qu'elles soient dues à des erreurs de mesure, à des situations mal définies, à des difficultés d'appréciation ou même à la prise en compte de variations locales.

Dans le cas des valeurs numériques (taille, nombre, etc.), la possibilité est donnée de fournir, à la place d'une valeur simple, un intervalle plus ou moins large. Par exemple, si, pour l'objet amphidisque, on répond [taille(amphidisque)={100,120}], cela signifie que la taille des amphidisques est comprise entre 100 et 120 (microns). Notons que, pour traduire la variabilité, le modèle descriptif peut fort bien comprendre des attributs statistiques (moyenne, écart-type, etc.).

Dans le cas des valeurs symboliques (couleur, forme, répartition, etc.), il est convenu que le choix simultané de plusieurs valeurs correspond à une disjonction de ces valeurs. Par exemple, la réponse [couleur(fleurs)={brun, jaune}] signifie que la couleur est située quelque part entre le brun et le jaune. La virgule entre brun et jaune signifie "ou", et jamais "et".

Mécanisme des objets multiples

Ce mécanisme s'applique dans deux situations, en fait très voisines.

La première n'a pas été rencontrée dans le domaine des *Hyalonema*. Aussi l'illustrerons-nous par la fleur de tournesol. Comment décrire la couleur de cette fleur, jaune à la périphérie et brune au centre? C'est au modèle descriptif d'offrir la seule façon naturelle de résoudre ce problème. En effet l'expert créateur du modèle ne peut pas ignorer que la fleur de tournesol est en fait une inflorescence, avec des fleurs d'un certain type à la périphérie, et des fleurs d'un autre type (et d'une autre couleur) au centre. Si le modèle descriptif est bien conçu, il doit permettre de décrire successivement les deux types de fleurs. Pour cela le modèle descriptif doit prévoir que l'arbre de description de l'objet fleur puisse, au moins dans le cas du tournesol, être parcouru non

pas une mais deux fois. Le problème du “brun *et* jaune” se pose au niveau de l'observable, non de l'observé, et sa solution n'est pas à rechercher au moment de la description.

La seconde situation se rencontre lorsqu'il est évident, d'emblée, qu'il existe plusieurs sortes différentes d'un même objet. L'expert connaît d'ailleurs les noms de ces différentes sortes, et est capable de les identifier. Par exemple, si l'expert a employé dans ses descriptions d'amphidiskues de *Hyalonema* les termes de macramphidiskues, mésamphidiskues et micramphidiskues, c'est qu'il les connaît et les reconnaît. La solution qui consiste à distinguer dans ce cas trois objets différents, de noms différents, est la seule qui soit offerte par la plupart des logiciels d'analyse de données. Cette représentation “développée à plat” présente cependant de graves défauts.

- Alors que, lors de la description d'une éponge donnée, le constat de l'absence de mésamphidiskues peut parfaitement se traduire au travers du modèle descriptif, on se trouve par contre coincé le jour où une éponge se révèle comporter quatre sortes d'amphidiskues. Le cas a été rencontré.

- On ne peut pas assurer le respect de l'homologie des caractères, c'est-à-dire que, lors des comparaisons ultérieures entre descriptions, on mettra bien en regard les mésamphidiskues d'une éponge avec les mésamphidiskues d'une autre, etc. En effet, rien ne ressemble plus à un mésamphidisque d'un individu de grande taille qu'un macramphidisque d'un individu plus petit, puisque la classification des amphidiskues repose sur leur dimension, donc sur un critère hautement méristique. L'expert avoue lui-même qu'il lui arrive d'hésiter. Il apparaît de plus que deux experts différents peuvent fort bien avoir des classifications décalées, et donc ne pas parler exactement de la même chose.

- Le respect de l'homologie implique celui de l'homonymie, c'est-à-dire que tout descripteur devra maîtriser la terminologie scientifique spécialisée du domaine, ce qui constitue un “pré requis” hasardeux. Pourtant, sous l'angle purement descriptif qui nous concerne, ce savoir théorique n'a aucune raison d'être imposé au descripteur. Nous connaissons ce problème sous le nom “du morse et de l'éléphant” : n'importe qui est capable de faire une description correcte des défenses de ces animaux sans savoir que dans un cas il s'agit de canines et dans l'autre cas d'incisives!

Dans ces deux situations, le modèle descriptif doit ouvrir la possibilité de multiplier les descriptions d'un même objet (fleur ou amphidiskues dans nos exemples). C'est la logique de multi-instanciation qui permet de remplacer l'arbre descriptif d'un objet par l'empilement de n arbres descriptifs analogues, chacun étant instanciable à son tour en une description. La valeur de la multiplicité n de chaque objet peut être bornée dans le modèle

descriptif ; par exemple, l'expert des *Hyalonema* a contraint n à ne pas dépasser 4 pour l'objet amphidiskues, ce qui laisse au descripteur, selon son degré de familiarité avec le domaine, la possibilité de décrire jusqu'à 4 sortes différentes d'amphidiskues, sans avoir à se soucier du nom à leur donner (ce qui n'a rien à voir avec les véritables difficultés de la description).

Mécanismes de particularisation

Nous regroupons sous ce nom trois mécanismes élémentaires, qui ont en commun d'exiger la modification dynamique de l'arbre descriptif en fonction de certaines situations particulières de description.

- Prise en compte de contraintes inter-objets. Il existe fréquemment des relations, dites de concomitance ou d'exclusion, entre les valeurs observées de certains attributs d'objets et les valeurs observables de ces objets (ou d'autres). Par exemple, la face exhalante des *Hyalonema* comporte selon les cas un ou plusieurs orifices exhalants (Fig. 2) ; dans le cas où l'orifice est unique (OAV : [nombre(orifices)=1]), il n'y a pas lieu de s'intéresser à leur répartition sur la face exhalante, ni à leurs dimensions, ni à la répartition de leur taille, etc. L'élagage de l'arbre descriptif doit se faire automatiquement ; c'est un cas de remplacement d'un sous-arbre descriptif par un autre, plus élagué donc plus particulier.

- Spécialisation d'un objet. Ici, c'est une information conceptuelle qui va permettre de particulariser l'arbre descriptif d'un objet. Par exemple, supposons que nous sachions pertinemment que la sorte d'amphidisque que nous sommes en train de décrire est un micramphidisque. L'expert a indiqué que, dans ce cas précis, la mesure de l'allongement de l'amphidisque est difficile et n'offre de plus aucun intérêt. La logique de spécialisation va permettre de substituer à l'objet amphidiskues et à son arbre générique un objet spécialisé micramphidisque avec son arbre spécifique (qui ne peut être qu'un sous-arbre de l'arbre générique).

- Traitement des cas exceptionnels. Nous n'avons pas rencontré chez les *Hyalonema* de description “hors norme”. Supposons néanmoins que le cas se présente, par exemple un individu de 90 cm de long, alors que la taille de tous les autres individus jusque-là connus n'atteignait pas 70 cm. Il est indiscutable que notre connaissance sur l'observable est modifiée. Mais est-ce une raison suffisante pour altérer le modèle descriptif ? Si l'expert décide que non, il doit être donné la possibilité au descripteur d'effectuer néanmoins sa description, sans blocage (du genre : “valeur maximale autorisée : 70 cm ; entrez une nouvelle valeur”) ; la description produite portera la mention “exception” et sera stockée à part, de façon à ne pas perturber les traitements sur l'ensemble des cas normaux. On notera que les exceptions sont les seuls cas où la description n'est

pas totalement supervisée par le modèle descriptif. La logique sous-jacente est : à situation exceptionnelle, traitement exceptionnel.

Mécanisme des points de vue

Lorsque la description d'un objet ou sous-partie peut se faire selon plusieurs points de vue, il est intéressant de pouvoir expliciter cette connaissance structurante dans le modèle descriptif.

Ainsi, l'objet corps des *Hyalonema* est habituellement décrit à deux échelles : celle des macro-constituants (les sous-parties macroscopiques), et celle des micro-éléments, observables seulement à partir de préparations microscopiques (Fig. 2). Dans d'autres domaines, d'autres points de vue pourraient intervenir, comme la biochimie, les séquences génétiques, etc. De façon générale, ces points de vue constituent des plans d'observation superposés, mais qui suivent le schéma structurel morphologique de base. Bien que cela n'ait pas été fait dans le cas des *Hyalonema*, il aurait été possible de structurer les

descriptions microscopiques organe par organe, donc en suivant le schéma des différentes parties macroscopiques.

A noter que la logique des points de vue est moins fondamentale que celle de composition, car elle n'autorise pas des inférences aussi fortes. En particulier, la présence / absence d'un point de vue ne peut revêtir aucune importance classificatoire. Néanmoins, la distinction des différents points de vue peut apporter souplesse et clarté dans les descriptions très riches, que le foisonnement d'informations tendrait à rendre inextricables.

DE L'UNICITE DU MODELE DESCRIPTIF

La synthèse des différents éléments élaborés précédemment forme un modèle descriptif. La figure 2 montre le modèle morpho-anatomique que nous proposons pour le genre *Hyalonema*. On notera que ce modèle englobe un certain nombre d'informat(ions permettant d'établir le contexte de chaque description (localisation et

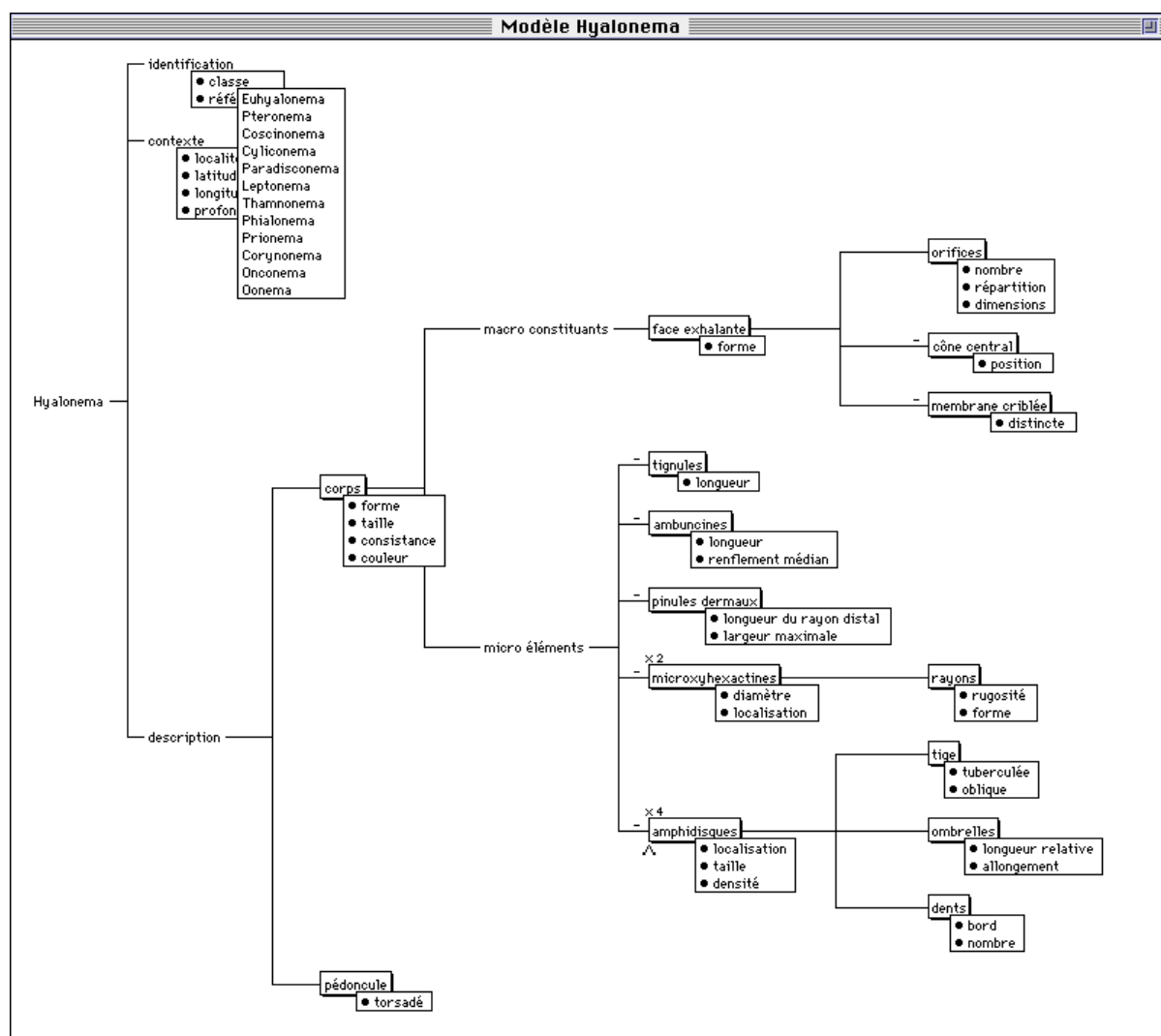


Figure 2. - Vue globale du modèle descriptif du genre *Hyalonema*.

origine de l'échantillon, etc.) ; ces informations ne revêtent aucune signification classificatoire.

Avant même de savoir ce qu'on pourra en faire, on peut se demander si un modèle différent et tout aussi valable n'aurait pas pu être obtenu. Nous pensons, sans pouvoir en apporter la preuve, que l'application raisonnée des différentes logiques, en commençant par la logique de composition, limite considérablement le nombre de modèles descriptifs fondamentalement différents qui peuvent être produits.

L'analyse de l'observable ne permet en effet de faire émerger qu'une organisation des connaissances correspondant à une réalité de la nature. Au moins dans leurs grandes lignes, les modèles descriptifs obtenus peuvent faire l'objet d'un consensus entre les différents experts concernés. Inversement, on voit mal comment deux experts travaillant indépendamment sur un même domaine pourraient parvenir à deux modèles incompatibles, la matière étant la même.

Bien entendu, il est capital que le modèle descriptif soit modifiable, notamment pour permettre l'introduction de caractères ou d'attributs de caractères nouveaux. Il doit rester opérationnel à l'échelle des taxons inférieurs, de l'espèce à la famille par exemple.

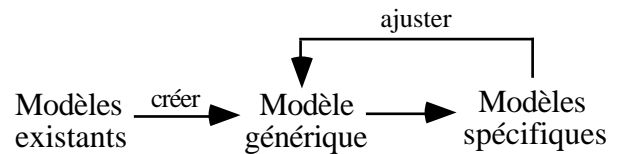
Créer un modèle descriptif assurant une représentation satisfaisante de l'observable d'un domaine, et permettant de guider l'acquisition et les traitements de descriptions de qualité, comme nous le verrons plus loin, devrait donc s'imposer comme une méthode de travail efficace et valorisante.

Mais il serait souhaitable que cet apport soit également manifeste quand il s'agit non plus d'un domaine isolé, mais de plusieurs domaines proches constituant un super-domaine. Par exemple, comment faire pour développer des modèles descriptifs pour les différents taxons (genres ou familles) de Spongiaires ? Nous manquons certes de recul sur ce sujet, mais un certain nombre de réflexions peuvent être avancées.

Idéalement, le modèle descriptif d'un domaine doit être un sous-modèle (une restriction ou spécialisation) du modèle descriptif du super-domaine. Selon cette constatation, il faut donc commencer par s'attaquer au super-domaine, puis le restreindre à son domaine d'intérêt. Or, dans la pratique, la seule solution viable serait un développement modulaire des différents domaines par différents experts, si possible en indépendance les uns des autres. Concilier deux approches aussi diamétralement opposées nécessite un compromis, qui préserve à la fois l'unicité de la solution globale (toutes les parties communes aux différents modèles doivent être héritées du modèle générique) et la liberté de chacun pour les parties qui lui sont spécifiques. Il faudra donc un travail de concertation préalable, d'autant plus facile à mener à bien qu'il pourra se fonder sur la

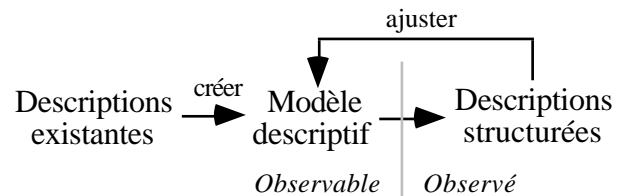
généralisation critique d'un ou plusieurs modèles préexistants.

La démarche méthodologique ci-dessus esquissée peut se traduire par le schéma suivant :



DESCRIPTIONS STRUCTUREES

Lors de l'explication du paradigme de modèle descriptif, nous avons mis en évidence le va-et-vient entre l'observable et l'observé nécessaire à la progression des connaissances descriptives. Le schéma correspondant est tout à fait analogue à celui ci-dessus ; tous deux traduisent en effet la même démarche hypothético-déductive.



Ce schéma introduit le concept de description structurée, que nous limitons aux descriptions qui sont obtenues en conformité avec un modèle descriptif, c'est-à-dire explicitement structurées.

Avant d'en venir aux descriptions structurées, à leur conception et aux applications qui en découlent, il nous paraît important de bien cerner ce sur quoi portent les descriptions.

L'INDIVU, OBJET DE DESCRIPTION

Toute description, si elle veut rester objective, se doit de ne pas comporter de généralisations. Sinon elle incorpore à l'observé une partie d'observable (plausible, mais non effectivement observé). De la sorte, une description ne peut porter que sur un individu considéré isolément, et non sur un ensemble plus ou moins hétérogène d'individus. Par exemple, si nous faisons la description synthétique d'un blond aux yeux bruns et d'un albinos aux yeux bleus, ceci couvrirait les blonds aux yeux bleus ainsi que les improbables albinos aux yeux bruns : la description serait devenue une définition (l'intention) d'un ensemble d'individus plus large et pas forcément réel (une classe, au sens mathématique du terme). De même, la description ne saurait s'appliquer à une entité abstraite ; par exemple, la description de la licorne n'a pas de sens : la classe des licornes est vide à jamais, même s'il en existe des définitions,.

On ne peut donc, en théorie, décrire qu'un individu qu'on a sous les yeux.

Cependant, dans le cas des *Hyalonema*, dont nous avons affirmé qu'il était représentatif d'autres domaines, les situations de description revêtent une bien plus grande diversité. D'une part, les individus réels des différentes espèces ne sont pas toujours disponibles, soit qu'ils se trouvent éparpillés aux quatre coins du globe, parfois on ne sait où, soit que leur préservation les ait plus ou moins altérés ou détruits au cours des ans. Et, quand on dispose de matériel frais, il est souvent composé de spécimens plus ou moins incomplets, si bien que leur description est parfois composite (emprunte ses éléments à plusieurs individus). Dans d'autres groupes se posent d'autres problèmes, comme le dimorphisme sexuel, l'existence de stades ontogéniques, les formes coloniales, etc. Pour pouvoir travailler, le biologiste devra donc se contenter d'une notion "extensible" de l'individu.

De plus, les descriptions ne peuvent se limiter aux individus concrets. Ce n'est pas parce que tous les individus qui en étaient connus ont disparu qu'une espèce n'a plus d'existence. Seulement, dans ce cas, il n'en reste plus que des descriptions, faites par des auteurs divers selon des formes et avec une précision dans les détails diverses. Il est alors nécessaire de ré-interpréter ces données, en s'appuyant aussi bien sur le texte que sur l'illustration, afin de les ré-adapter à la structure définie par le modèle descriptif. C'est ce travail de spécialiste qui a dû être fait pour la plupart des espèces de *Hyalonema*. Cela a été pour nous l'occasion de comparer le savoir-faire de différents spécialistes ayant décrit, indépendamment, des spécimens de la même espèce ; il est très inégal.

Ainsi donc, en pratique, ce sont souvent des pseudo-individus qui sont décrits dans la littérature ; on fait simplement en sorte qu'il ne soit pas abusif de les traiter comme de véritables entités individuelles, en veillant en particulier à ne pas introduire de généralisations. Dans tous les cas, le contexte de la description doit clairement être indiqué, au titre de données d'accompagnement.

UTILISATION DES CONNAISSANCES DESCRIPTIVES STRUCTUREES

Les descriptions en tant qu'éléments de connaissance. Bases de connaissances

Chaque description, accompagnée des illustrations auxquelles elle se réfère, constitue une connaissance complète. Nous qualifions cette connaissance d'élémentaire, car, considérée isolément, elle n'offre que peu d'intérêt en biologie. De même que, sur le plan concret, les échantillons ne sont utiles que lorsqu'ils sont rassemblés en collections, les descriptions ont pour vocation d'être regroupées en bases de connaissances. En effet, c'est de la comparaison des descriptions entre elles que jaillit la lumière sur les rapports et différences entre les individus, considérés dans toute leur diversité.

Le volume des données informatiques est de moins en moins un obstacle à la réalisation de bases de connaissances regroupant un grand nombre de descriptions individuelles. Cette tendance irréversible rend possible ce qui paraissait déraisonnable il n'y a pas si longtemps, à savoir représenter la diversité inter-individuelle non plus par des résumés statistiques mais par la mise en mémoire des caractéristiques individuelles (données "brutes").

Les descriptions en tant que vecteurs d'informations

La formalisation de l'observation sous forme de descriptions n'est nécessaire qu'à l'instant où il faut la transmettre ; sinon une simple mémorisation suffit. Inversement, toute amélioration dans la formalisation du savoir issu de l'observation est un élément favorable à sa transmission.

C'est sans doute l'avantage majeur de la mise en œuvre de descriptions structurées que d'offrir une base assurée aux échanges d'information entre les différents spécialistes (là où les qualités d'objectivité et d'exactitude sont primordiales). La transmission du savoir descriptif sous forme structurée assure tout à la fois une information riche, conforme à la réalité, auto-documentée, dépourvue d'ambiguïté et pouvant être mobilisée sans interprétation préalable. C'est donc une voie vers la pérennisation du savoir, qui préoccupe tant les systématiciens à l'approche de la retraite.

A une époque où les réseaux informatiques scientifiques prennent une part de plus en plus importante dans la vie des chercheurs, il est temps de réaliser les limites des échanges informels, et de se soucier de mettre le savoir individuel à la disposition de la collectivité, sous la forme la plus efficace et la plus nette possible.

Analyse des descriptions à des fins de classification, d'études de phylogénie, de biogéographie, etc.

Disposer d'une base de descriptions permet d'étudier la répartition des caractères en fonction des individus. De plus, si chaque description peut être attribuée par l'expert à un taxon déterminé, elle devient un exemple (ou "cas") de ce taxon. Cela permet de caractériser les différents taxons, en extrayant le ou les caractères qui les discriminent le mieux et en les synthétisant en définitions ; cette approche "ascendante" renoue avec une tradition aristotélicienne trop rapidement condamnée. De plus, si la valeur évolutive des caractères est connue, les "sauts" entre taxons peuvent être évalués sous l'angle phylogénétique. Et si le contexte géographique ou écologique des descriptions individuelles a été relevé, les liaisons entre les caractères et les répartitions thématiques peuvent être approfondies.

Visualisation des descriptions sous forme explicitement structurée

De par la conformité des descriptions au modèle descriptif, chaque cas se superpose topographiquement au canevas du modèle : l'instanciation se traduit par la particularisation de l'arbre de structure et par l'adjonction des valeurs prises par chaque attribut de ses objets (comparer les Figs. 2 et 3). Cette présentation visuelle est plus confortable que l'affichage sous forme de liste ou de tableau ; elle permet en outre, en faisant défiler les cas page par page, de comparer directement les descriptions entre elles.

La Figure 3 correspond par exemple à la visualisation d'une description d'un individu

appartenant à *Oonema* (l'un des sous-genres de *Hyalonema*) ; on peut y remarquer :

- les objets à statut "absent" barrés d'une croix (ex. : tignules) ;
- les objets à statut "inconnu" apparaissant en fantôme (ex. : pédoncule, en bas de la Fig. 3) ;
- les objets multi-instanciés (ex. : amphidisque) : une seule instance est affichée, avec son sous-arbre propre ; en cliquant dans la flèche (signe ">") de la quatrième sorte d'amphidisque, on fait apparaître les autres descriptions d'instances de cet objet.

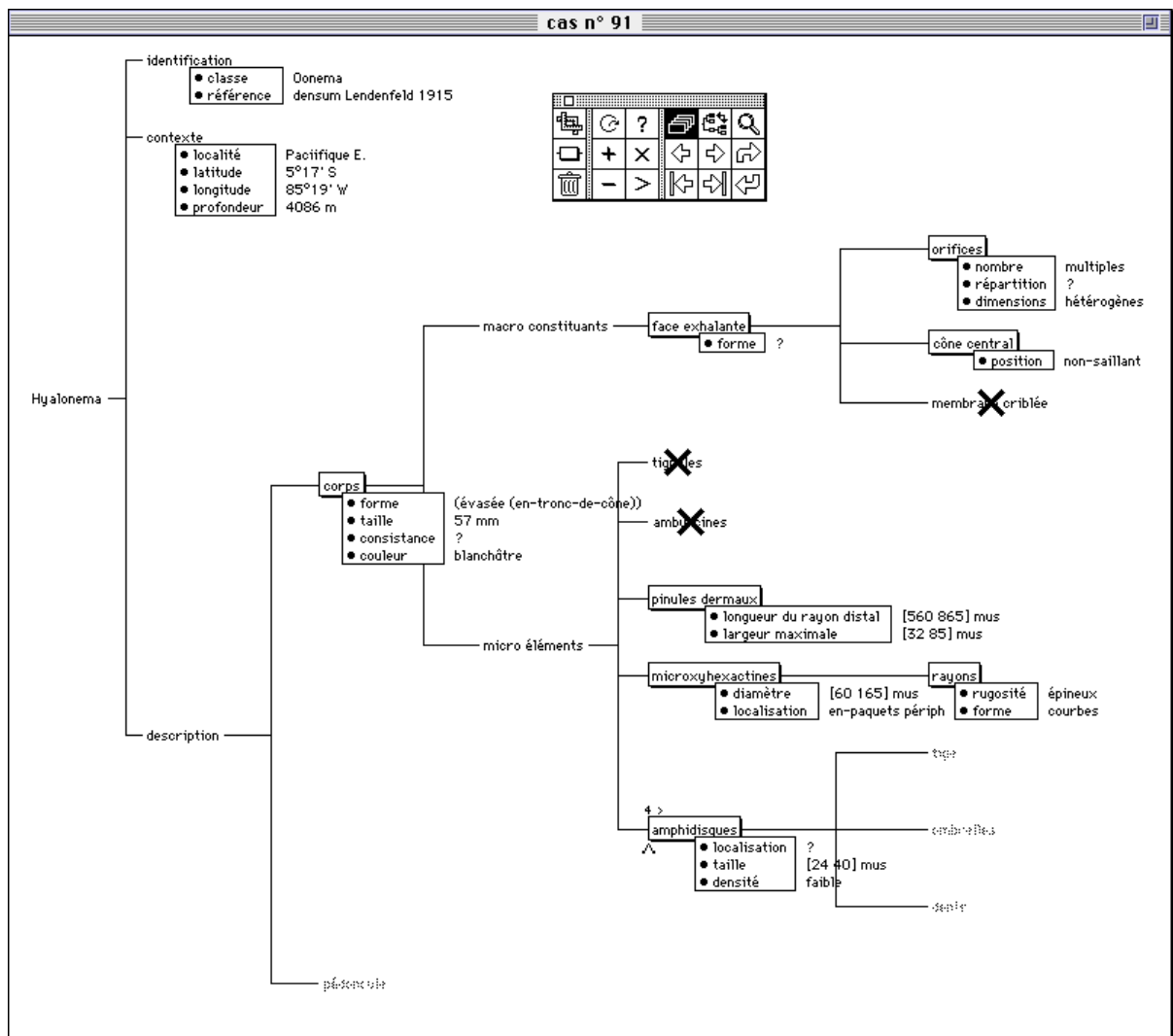


Figure 3. - Présentation d'une description de *Hyalonema* sous forme structurée.

Comparaison des descriptions

L'intérêt le plus évident de disposer de descriptions structurées est de rendre possible entre elles des comparaisons fiables. Dès l'instant où la position de chaque caractère dans une description est prédéterminée, il est simple de faire ressortir immédiatement les différences entre

descriptions ou au contraire les similitudes (points communs).

L'établissement des diagnoses (descriptions limitées aux caractères discriminants par rapports aux autres taxons), de "rapports et différences", de clefs dichotomiques, ou de véritables systèmes experts de détermination, peut alors être facilité

par des logiciels d'aide, voire même complètement automatisé.

OUTILS EXISTANTS ET A VENIR

Les développements logiciels qui ont accompagné le projet "Éponges" ont abouti, au travers d'un travail de thèse (CONRUYT, 1994), à la mise au point d'une chaîne complète d'outils permettant de créer et manipuler les connaissances structurées tant observables qu'observées, et de les utiliser à des fins de classification (pour le spécialiste) ou de détermination (par n'importe qui).

Tout d'abord, un générateur interactif, à forte composante graphique, permet de créer puis de faire évoluer le modèle descriptif d'un domaine. Il n'était pas disponible au moment de la création du modèle descriptif des *Hyalonema*, et a donc pu bénéficier des observations faites à cette occasion. Ensuite, un questionnaire hypermédia, fabriqué automatiquement à partir du modèle descriptif et en traduisant les moindres détails, permet de saisir des descriptions, soit *de novo*, soit par recopie différentielle d'une description préexistante. La navigation dans le questionnaire peut se faire localement par renvoi d'écran à écran (chaque écran étant relatif à un objet), ou bien par référence directe à l'arbre de structure du modèle descriptif (à l'aide de la vue globale, Fig. 3). Le processus de description apparaît bien ainsi comme une instanciation progressive, dans un ordre logique, de la variable-description que constitue le modèle descriptif. Du fait de sa conception même, l'outil autorise l'incorporation à la description (de chaque objet du modèle descriptif) d'une ou plusieurs illustrations saisies au scanner par exemple.

Les descriptions sont stockées dans une base de cas, sous une forme interne auto-explicite reflétant la structure du modèle. En ce qui concerne les *Hyalonema*, les 12 classes d'identification (Fig. 2) correspondent aux sous-genres reconnus par l'expert. De la sorte, chaque description est un exemple de sa classe d'identification ; à terme, l'ensemble des exemples d'une classe (ici un sous-genre, mais cela pourrait être aussi bien une espèce, ou une autre catégorie taxinomique) illustre la variété individuelle propre à cette classe, sans recours à la statistique ou aux définitions généralisantes.

Il n'existe pas encore de gestionnaire de base de cas, sa nécessité ne s'étant pas fait sentir dans le cadre de ce premier travail. L'extraction des cas est déportée dans chaque outil d'utilisation des descriptions.

Pour l'aide à la classification, on utilise une technique dite d'apprentissage par induction, qui permet d'ordonner les différents caractères observables en fonction de leur uniformité au sein de chaque classe (critère d'entropie de SHANNON). L'un des intérêts de cette évaluation de la qualité des caractères (chargée de faire ressortir les

"bons" caractères classificatoires du point de vue de leur efficacité à discriminer) est de substituer à la notion fort critiquable de distance entre individus, celle plus naturelle d'ordre ou de régularité dans les descriptions. Un autre avantage est le caractère clairement explicatif du raisonnement inductif, qui travaille caractère par caractère (il est dit monothétique ; voir SUTCLIFFE, 1994) plutôt que sur une composition contre nature de plusieurs caractères. Cette caractéristique monothétique permet par ailleurs à l'expert de tester manuellement les effets de permutations dans l'ordre d'importance des caractères. L'expert peut aussi modifier l'arbre d'induction produit, de façon à rapprocher le plus possible cet arbre "artificiel" d'un arbre taxinomique biologiquement plus satisfaisant. Notons pour l'anecdote que l'induction sur les exemples de *Hyalonema* a fait ressortir comme le plus discriminant la forme du corps de l'éponge, caractère considéré universellement jusqu'alors comme inutilisable (du fait de sa variabilité trop insaisissable).

Pour l'aide à la détermination, on peut opérer de deux manières.

La première met en œuvre un système expert automatiquement dérivé de l'arbre d'induction. Son avantage est la performance. Son inconvénient est de n'utiliser qu'une partie du corpus de connaissances, à l'instar des clefs dichotomiques, et donc de ne pas savoir se comporter correctement face aux observations incomplètes (CONRUYT *et al.*, 1992).

La seconde mobilise l'intégralité des connaissances implicitement contenues dans les descriptions stockées dans la base de cas. Le raisonnement sur cas, ou Case-Based Reasoning (CBR), permet en effet d'opérer à une sorte de recherche multicritères dans la base de cas. Cette recherche, au lieu de procéder par élimination de tous les cas qui ne sont pas parfaitement compatibles avec l'inconnu à identifier, permet de se focaliser progressivement sur les cas les plus similaires. L'avantage de cette approche, plus sophistiquée, est de pouvoir intégrer à la mesure de similarité les données inconnues, sans perte de performance notable dans la qualité du résultat (tant bien sûr que l'inconnu reste dans des proportions raisonnables, ce qui est heureusement généralement le cas ; voir AURIOL 1995, p. 145).

L'autre avantage réside dans l'aspect "dynamique" du processus : tout nouvel exemple rajouté à la base de cas est pris en compte immédiatement, sans qu'il soit nécessaire de "recompiler" un arbre de détermination.

Tous ces logiciels ont été conçus sous forme d'outils génériques, adaptables aux besoins de chacun, et développés selon des normes professionnelles. Les solutions retenues ont un caractère résolument novateur. Elles mettent en jeu des techniques dérivées de l'intelligence artificielle, jusqu'ici réservées à

d'autres fins, comme la représentation d'objets sous forme de "frames", l'apprentissage par induction ou le raisonnement sur cas. L'intégration de l'application "Éponges" dans un projet européen de recherche en informatique (MANAGO *et al.*, 1993) a permis, au stade même de leur conception, de s'assurer de l'adéquation des logiciels aux besoins précis du systématicien.

On dispose donc déjà d'une panoplie d'outils de cette nouvelle génération capable de tirer profit de la structuration des données descriptives. Les années à venir accroîtront encore leur diversité et leurs performances.

Il est prévu de développer un traducteur en langage naturel, de façon à pouvoir restituer, sous une forme écrite habituelle, aussi bien les descriptions que les "rapports et différences", que les diagnoses, etc. Il est envisageable à ce niveau d'offrir le choix entre différentes langues-cibles, si bien que chacun pourra disposer de l'ensemble des connaissances d'observation sans avoir à les traduire au préalable. L'équivalence des termes scientifiques entre les différentes langues n'aura à être établie qu'une fois pour toutes ; elle pourra s'appuyer, dans le cas des Spongiaires, sur le thesaurus déjà préparé. Cette capacité multilingue sera un avantage appréciable au niveau des échanges d'information internationaux.

Conclusions

La prise en compte explicite des données structurelles lors de la réalisation des descriptions permet un net progrès lors de leurs divers traitements. La comparaison entre différentes descriptions est en particulier rendue beaucoup plus efficace.

Les conséquences méthodologiques sont de deux ordres. D'abord, il vaut mieux multiplier les descriptions individuelles que de généraliser l'information dans des "descriptions" de groupes qui font nécessairement appel à la subjectivité. Ensuite, il est primordial d'effectuer, préalablement aux descriptions elles-mêmes, une étude détaillée de l'observable, et de le formaliser sous une forme qui préserve sa richesse (modèle descriptif).

Le recueil des descriptions sous forme structurée permet de plus aux utilisateurs systématiciens de visualiser les descriptions sous une forme plus claire, tout en ouvrant la possibilité de traductions automatiques sous forme de textes (multilingues), et en pouvant constituer le support d'échanges entre utilisateurs, sous une forme non ambiguë.

L'expérience de l'application aux Éponges du genre *Hyalonema* indique clairement que ce type d'approche est très favorable à une meilleure formalisation des connaissances descriptives et à leur pérennisation dans des bases de connaissances spécialisées et intégralement mobilisables (sur

place comme à distance). De l'avis même de l'expert du domaine, le travail de départ, nécessaire pour formaliser l'observable et créer le modèle descriptif, conduit à la production d'un outil qui améliore sensiblement la perception des différents caractères taxinomiques.

L'intérêt des modèles descriptifs, fondés sur une analyse de l'observable morpho-anatomique, pourrait être généralisé à des modèles fondés sur des considérations fonctionnelles ou taxinomiques. On pourrait ainsi imaginer de structurer l'information en disséquant l'individu en ses principales fonctions (alimentation, reproduction, défense), et de regrouper les observations élémentaires sous l'angle de ces fonctions. De tels modèles, possédant une "valeur explicative" supérieure, seraient potentiellement plus intéressants pour le spécialiste. Mais ils seraient aussi plus délicats à mettre au point, car la nature produit souvent des caractères dont le rôle est mixte (par exemple les dents du chien : alimentation et défense ; ou les canines du morse : défense et reproduction) ou au contraire difficile à appréhender (à quoi sert la corne du narval?). Des difficultés supplémentaires peuvent aussi surgir du fait de la part importante de l'interprétation, qui ne permet plus de préserver l'unicité du modèle.

Remerciements

Nous remercions le Pr. J. LEBBE pour ses précieux conseils et pour sa longue et constructive critique du manuscrit.

Références

ALLKIN, R., 1984. Handling taxonomic descriptions by computer. In ALLKIN R. & BISBY F.A. eds., Databases in Systematics. Systematics Association Special Volume, No. 26, London (Academic Press), pp. 263-278.

ARISTOTE. Les parties des Animaux. pp. v-xviii. Ed. Les Belles Lettres, Paris, 1990.

ARISTOTE (345-342 ff., B.C.) *Historia Animalium*. Translated by A. L. Peck. Cambridge : Harvard University Press (Loeb Classical Library), I(1-3) : civ + 240 + 8 pp. 1965. II(4-6) : viii + 414 + 8 pp., 2 endpaper graphs 1970. III(7-9).

AURIOL, E., 1995. Intégration d'approches symboliques pour le raisonnement à partir d'exemples. L'induction et le raisonnement par cas dans le diagnostic technique. Thèse, Paris (Univ. Paris IX - Dauphine), 217 pp.

CONRUYT, N., 1994 Amélioration de la robustesse des systèmes d'aide à la description, à la classification et à la détermination des objets biologiques. Thèse, Paris (Univ. Paris IX - Dauphine), 281 pp.

CONRUYT, N., MANAGO, M. & LE RENARD, J., 1992. Modélisation, formalisation et analyse d'objets biologiques en vue de leur identification ; application au domaine des Eponges marines. *Actes des 3èmes Journées "Symbolique-Numérique"*, Paris, pp. 121-124.

CRUSE, D.A., 1986. *Lexical semantics*. Cambridge University Press, Cambridge, England.

DALLWITZ, M.J., 1980. A general system for coding taxonomic descriptions. *Taxon*, 29 (1), pp. 41-46.

LE RENARD, J. & CONRUYT, N., 1994. On the representation of observational data used for classification and identification of natural objects, IFCS'93, Lecture Notes in Artificial Intelligence, Springer Verlag, pp. 308-315.

MANAGO, M., ALTHOFF, K.D., AURIOL, E., TRAPHÖNER, R., WESS, S., CONRUYT, N., & MAURER, F., 1993. Induction and reasoning from cases, Proceedings of the First European Workshop on Case-Based Reasoning (EWCBR-93), RICHTER, M.M., WESS, S., ALTHOFF, K.D. & MAURER, F. (Eds.), Vol II, Kaiserslautern, Springer Verlag, 1994.

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

SUBRAMANIAN, D., 1990. A theory of justified reformulation. Change of Representation and inductive Bias. BENJAMIN, D. (Ed.), Kluwer Academic Publisher, Boston, pp. 147-167.

SUTCLIFFE, J.P., 1994. On the logical necessity and priority of a monothetic conception of class, and on the consequent inadequacy of polythetic accounts of category and categorization. IFCS'93, Lecture Notes in Artificial Intelligence, Springer Verlag, pp. 55-63.

WINSTON, M. E., CHAFFIN, R. & HERRMANN, D., 1987. A taxonomy of Part-Whole relations. *Cognitive Science*, 11 : 417-440.

Jacques LE RENARD & Claude LEVI
Muséum National d'Histoire Naturelle
Laboratoire de Biologie des Invertébrés Marins
et Malacologie, CNRS URA 699
55, rue de Buffon, F - 75005 Paris, France

Noël CONRUYT
Muséum National d'Histoire Naturelle
Laboratoire de Biologie des Invertébrés Marins
et Malacologie, CNRS URA 699,
55, rue de Buffon, F - 75005 Paris, France

&
IREMIA, Université de la Réunion
15, av. René Cassin, BP 7151
F - 97715 Saint-Denis

Messag. Cedex 9, France
&
Société AcknoSoft
58A, rue du Dessous des Berges
F - 75013 Paris, France

Michel MANAGO
Société AcknoSoft
58A, rue du Dessous des Berges
F - 75013 Paris, France

