

## **IV LE ROLE CENTRAL DES DESCRIPTIONS EN SCIENCES NATURELLES**

Les sciences dites d'observation reposent sur la capacité d'analyser la réalité des choses, en bref de décrire. L'activité de description va tellement de soi que son analyse apparaît presque comme superflue. Pourtant chacun sait qu'il y a de bonnes descriptions et de moins bonnes, et que leur emploi indispensable en tant qu'outils pose de nombreux problèmes.

Pourquoi les descriptions ? Quelles qualités doivent-elles avoir ? Telles sont les questions auxquelles il convient d'apporter des réponses claires, avant même de rechercher quelles solutions informatiques sont de nature à faciliter et à améliorer les activités de description, telles qu'elles interviennent par exemple en classification ou en détermination des êtres vivants et autres objets naturels.

### ***4.1 Objectifs de la description***

---

La description des différentes entités qui composent notre monde est apparue dès la plus haute antiquité comme le moyen fondamental d'accroissement des connaissances. Pour “apprendre” ce qu'est un animal, une plante, une roche, etc., il faut l'observer certes, mais aussi s'en faire une représentation mentale (pour soi-même) ou écrite (pour les autres). La transmission du savoir implique la notion de description.

Une description scientifique est une abstraction objective. Abstraction parce qu'elle permet de s'affranchir de l'observation réelle qui lui a donné matière ; et objective parce qu'elle ne laisse pas place à l'interprétation. Dans l'idéal, il n'y a pas de déformation mais une simple transcription “à l'identique” des traits concrets de l'entité observée en des caractères, ou traits, représentés. La représentation se fait traditionnellement sous forme textuelle, souvent agrémentée de dessins ; notre époque voit la banalisation de supports permettant plus de richesse et de souplesse, qu'il importe de savoir adapter à la complexité et à la variabilité du monde réel.

---

<sup>1</sup> Les paragraphes 4.1 à 4.5 (inclu) correspondent à la version française de l'article [Le Renard & Conruyt, 1994].

Décrire peut constituer un objectif en lui-même. L'intérêt de passer, avec une perte d'information minimale, du niveau concret à un niveau plus "manipulable" se fait sentir dès que le savoir doit être diffusé. On peut multiplier les descriptions, pas les originaux. Pour un domaine donné, correspondant à un certain groupe naturel particulier, les "échantillons" sont en général disséminés dans de multiples collections, en des lieux divers, ce qui rend leur consultation longue et difficile ; s'il existe des recueils de leurs descriptions, cela revient en quelque sorte à démultiplier leur disponibilité. De plus, certains échantillons sont périssables, ou certains de leurs caractères s'altèrent avec le temps ; les descriptions et les figurations peuvent permettre dans une certaine mesure de pallier les problèmes de conservation.

Mais surtout les descriptions constituent l'un des ingrédients de nombreuses activités scientifiques plus "nobles". Elles constituent en effet des sources d'information, de données ou de connaissances, irremplaçables. L'étude des analogies et des différences entre descriptions est à la base de tous les processus de classification ; et la comparaison d'un objet naturel (concret ou abstrait) avec des descriptions préexistantes sous-tend toute activité de détermination.

On peut dire ainsi que l'on décrit pour, dans un premier temps, accroître les connaissances ponctuelles, et ensuite apprendre la nature à un niveau plus général et mieux la comprendre. De ce double objectif découlent les qualités que l'on doit attendre des descriptions.

## **4.2 Qualités de la description**

Nous avons vu que la qualité primordiale d'une description est son objectivité, c'est-à-dire qu'elle doit être à la fois vraie et complète. Toute méthode visant à faciliter la description doit donc permettre de couvrir tout ce qui est observable et de l'exprimer exactement, sans ambiguïté ; cette condition reflète le souci de maximiser le contenu informatif de la description. Dans l'absolu, une description parfaite devrait permettre de reconstituer exactement l'objet primitif ; dans la pratique, on se contente de pouvoir se faire une "bonne idée" de cet objet, surtout en ce qui concerne ses particularités. Notons que ceci implique que soient pris en compte non seulement les caractères descriptifs, mais aussi les différents liens (topologiques, relationnels, de dépendance, etc.) qui peuvent exister entre ces caractères ; car ces liens sont eux-mêmes porteurs d'information.

D'autres qualités annexes peuvent être citées, comme la clarté et la concision, valables pour tout écrit scientifique. Certains auteurs s'attachent à l'élégance du texte ; une bonne forme et un bon style ne sauraient en effet nuire à un bon contenu. Il est rarement fait cas de l'intelligibilité, qualité qui rendrait la compréhension de la description aisée même par un non spécialiste ; elle implique le recours à un vocabulaire moins technique, avec en contrepartie une

perte de concision et de précision : il y a donc là un compromis à trouver, dans l'attente d'une solution qui permettrait d'adapter le "niveau" de la description à celui de l'utilisateur. Mais il ne suffit pas à une description d'être excellente en elle-même : il lui faut de plus être comparable.

### **4.3 Qualités des descriptions**

---

Aux fins de classification ou de détermination, l'essentiel est de pouvoir comparer des descriptions entre elles. Quand ces descriptions ont été rédigées par un même auteur, elles suivent généralement un même plan d'ensemble, ce qui facilite les comparaisons en localisant les caractères homologues à des parties correspondantes des textes. Mais, quand les auteurs sont différents, ils ont pu suivre des "méthodes" d'observation hétérogènes ; les comparaisons en sont rendues bien plus délicates.

La notion d'**homologie** est essentielle ; elle permet de s'assurer que l'on ne compare que des caractères comparables et découle du fait que tous les objets biologiques possèdent un plan d'organisation, qui se retrouve identique chez leurs semblables. La reconnaissance et la prise en compte de ce plan de constitution général autorise une structuration naturelle des descriptions, suivant ce que nous appellerons un modèle descriptif.

Remarque : Les considérations précédentes concernent virtuellement toutes les descriptions d'objets naturels. Cependant, aussi bien en classification qu'en détermination, chaque spécialiste limite ses études à un domaine plus particulier, comme un groupe zoologique ou botanique, et/ou une zone géographique, et/ou un écosystème, etc.. Dans cette thèse, c'est un tel domaine bien délimité que nous aurons en tête, plutôt qu'un "système universel" encore actuellement inaccessible.

### **4.4 La représentation des données descriptives**

---

Partant du principe que l'on ne comprend que ce que l'on modélise, et qu'il vaut mieux adapter le modèle à la réalité que l'inverse, nous allons étudier plus en détail quels sont les éléments constitutifs d'une description, et comment ils sont agencés entre eux par le descripteur. (Nous appelons ainsi la personne, généralement un spécialiste, qui fait la description, et non un caractère décrit qui devrait s'appeler un ... décrit ou *descriptum*). Nous en déduisons comment doivent être conçus les modèles descriptifs, en tenant compte des contraintes de qualité énoncées précédemment.

### 4.4.1 Structuration naturelle

Plaçons-nous dans le cadre d'un domaine quelconque ; par exemple les animaux de la ferme, domaine pour lequel chacun peut revendiquer le titre de “spécialiste”. Et voyons comment le spécialiste va s'y prendre pour classer et déterminer ces animaux.

Première observation : tous ces animaux ont quatre membres, deux à l'avant, deux à l'arrière. Les membres antérieurs sont soit des pattes, soit des ailes chez les volailles. On apprend ainsi qu'il existe deux catégories principales, que le spécialiste dénommera aussitôt Mammifères et Oiseaux (avec des majuscules : on est dans un domaine scientifique). Ensuite, parmi les Mammifères, le chat et le chien de la ferme se distinguent en mangeant de la viande. Voilà deux autres catégories : Carnivores pour eux, Herbivores pour les autres. Parmi les Carnivores, il y a le Chat qui ressemble au Tigre, et le Chien qui ressemble au Loup : Félines et Canins. Etc. On voit s'établir de la sorte une véritable hiérarchie de catégories (ou classes, au sens large du terme), de la plus générale (Animaux) aux plus particulières (les Espèces, comme Chat, Chien, Cheval etc.). Cette “systématique” s'appuie sur une hiérarchie de caractères discriminants (la nature des membres antérieurs, le régime alimentaire, etc.) plus ou moins faciles à mettre en évidence ; par exemple, sur quoi s'appuie la distinction entre Félines et Canins ?

C'est le propre du spécialiste de savoir reconnaître du premier coup d'œil : il est expert de son domaine. Mais pour comprendre, pour connaître, il a besoin d'analyser la réalité de plus près. Ce n'est qu'après avoir fait (ou avoir lu) les descriptions des divers Félines que lui-même (ou un autre spécialiste avant lui) pourra préciser la définition de la classe appelée Famille des Félinidés et dire qu'elle se différencie, entre autres, par la présence de dents postérieures tranchantes dites carnassières, là où ne se montrent chez les Canidés donc chez le Chien que des molaires peu spécialisées.

On peut noter dès ici que la “distinction entre Félines et Canins” relève en fait de deux approches duales. Dans un sens, celui suivi en classification, on apprend, par un processus de généralisation, que le caractère “présence de carnassières” synthétise (on dit : subsume) tout ce qui a pu être observé chez les différents Félines en matière de dents postérieures. Dans l'autre sens, celui suivi en détermination, on déduit, par l'analyse du fait qu'il a des carnassières, que Minou est un Félin et non un Canin. Mais dans les deux cas, il a fallu recourir à la description des dents postérieures ; c'est ce que nous appellerons une description locale.

#### **4.4.2 La logique de composition**

La description d'une entité (disons d'un chat, ou d'un chien) est une composition de descriptions locales, correspondant à tout ce qui est observable (donc descriptible).

L'ordre dans lequel progresse la description n'est pas quelconque, mais répond à une certaine logique qu'il convient de reconnaître. Alors que le chat et le chien sont tous deux composés d'un corps, d'une tête, de quatre pattes et d'une queue (ils héritent tout cela de leur appartenance à la classe des Mammifères), ce serait de l'anticonformisme que de faire commencer leur description par la queue ; commencer par les pattes serait curieux, à moins que le descripteur ne soit une fourmi ; quant au choix entre le corps et la tête, il reste ouvert. En fait, cette logique est affaire de spécialistes, qui seuls sont capables de se mettre d'accord pour cerner l'ordre le plus "naturel" à suivre lors des descriptions. Si c'est la tête qui vient en ... tête, selon une logique toute trouvée, alors la description s'intéressera d'abord à ses caractères propres, comme sa forme, ses dimensions, sa couleur, etc., à ses rapports avec les autres parties, et ensuite, ici aussi selon un ordre non arbitraire, on passera à la description de ses sous-parties (yeux, bouche, nez, oreilles, etc.). Et ainsi de suite.

Ce mécanisme de base est celui de la décomposition en sous-parties ; il est répété autant de fois qu'il est nécessaire pour atteindre la finesse de description souhaitée (qui, rappelons-le, dépend de l'utilisation visée). On peut convenir ainsi d'un "arbre d'exploration", à chaque nœud duquel on doit effectuer une description locale, et dont chaque branche traduit une relation de partie à sous-partie. Pour que cet arbre n'introduise pas de limitation arbitraire, il doit prévoir toutes les situations susceptibles de se rencontrer, y compris les cas particuliers et les exceptions. De ce fait, il est généralement plus touffu qu'il n'est nécessaire pour chaque situation de description considérée individuellement, où certaines branches peuvent s'avérer non pertinentes ou sans objet.

En particulier, lors des descriptions, il est fait appel à un mécanisme d'élagage automatique, de pur bon sens. Ainsi, le fait de constater l'absence d'une certaine partie rend sans objet la description de ses sous-parties ; de même, si par exemple je décris Médor, le chien de garde de la ferme, et qu'il ne veut pas ouvrir la gueule, je préférerais qu'on me dispense de décrire ses dents ou sa langue. Ceci illustre une situation couramment rencontrée lors des descriptions d'objets naturels, où certaines descriptions locales ne sont pas possibles du fait de la situation d'observation (partie cachée ou présentement indistincte) ou du fait de l'état incomplet du spécimen décrit. On peut considérer que, par convention, l'absence d'une description locale signifie que la partie correspondante est inconnue ; par opposition, le fait que cette partie soit constatée absente constitue une information qui doit figurer explicitement dans la description.

Cette dernière distinction est importante. Quand je décris un Chat, si je déclare qu'il n'a pas de queue, je m'oriente vers le fait qu'il appartient à la race Manx (chats sans queue de l'Ile de Man), à moins qu'il ne s'agisse d'un accident. Par contre, si je reste muet sur sa queue, je n'apporte aucune information ; la "valeur «inconnu»" souvent invoquée dans ce cas est un non-sens, ou pire un artifice tendant à donner la forme d'une information à ce qui n'en est pas.

### **4.4.3 La logique des points de vue**

Il arrive souvent que la description d'un objet naturel puisse se faire à différents niveaux. Par exemple, elle s'intéressera à la morphologie, ou à l'anatomie, ou à la cytologie, voire à la biochimie ou la carte génétique. Ceci est vrai d'ailleurs pour chacun des points d'observation ou parties concernées. Il faut donc pouvoir répartir l'information selon ces différents points de vue, et tenir compte des relations structurelles existant entre ces différents plans d'observation.

Pratiquement, la logique des points de vue est très semblable à la logique de composition. Cependant elle ne possède pas une sémantique aussi riche ; le fait que, dans une description d'une partie donnée, un niveau d'analyse ne soit pas accessible n'implique pas que ce niveau demeure inaccessible lors de la description des sous-parties. Une autre différence réside dans le fait que, lors d'une classification, l'absence avérée d'une sous-partie sera prise en compte comme une information, alors que l'absence d'un point de vue n'a aucune signification classificatoire.

L'intérêt de cette logique est de permettre l'analyse des descriptions de plusieurs entités selon des points de vue analogues. Alors que l'on s'attendrait à ce qu'une classification établie selon un certain point de vue (disons la morphologie externe) corresponde à la classification établie selon un autre point de vue (disons le génome), force est de constater généralement des divergences de résultats. Pour expliquer ces écarts, il est nécessaire de pouvoir disposer de "passerelles" entre les différents points de vue, donc de les appliquer à une structure commune ; c'est là l'un des gros intérêts d'établir un modèle descriptif, qui permet de préserver l'homologie même à des échelles d'observation différentes. Le modèle descriptif permet en quelque sorte d'indexer les connaissances, de les positionner les unes par rapport aux autres ; c'est un peu l'équivalent des structures relationnelles et/ou hiérarchiques dans les banques de données.

### **4.4.4 La logique de spécialisation**

Revenons aux animaux de la ferme, et supposons que nous disposions d'une classification des différents types de ferme d'élevage. Alors que, si nous ne savons rien de "notre" ferme, le modèle général des animaux qui y sont élevés comporte quatre membres, s'il s'agit d'une ferme spécialisée dans l'aviculture, nous pouvons partir d'un modèle plus précis, d'animaux pourvus de deux ailes,

de deux pattes, d'un bec, de plumes, ou au contraire dépourvus de cornes ou de dents, etc..

Le fait de disposer, à un niveau déjà abstrait, d'un concept plus précis de notre ferme nous permet de restreindre l'étendue du domaine de connaissances, et de renseigner par avance (sans observation) un certain nombre de descriptions locales. Ce mécanisme, dit de spécialisation, est d'une généralité telle qu'il peut être découvert dans de très nombreuses descriptions rédigées par les naturalistes, en lieu et place de véritables descriptions locales. Ainsi, la simple indication que notre ferme élève des oiseaux aquatiques (des canards par exemple) supplée en partie à une description des pattes (qui sont toujours palmées) ou du plumage (qui est toujours étanche à l'eau).

La spécialisation constitue ainsi un raccourci commode, permettant de remplir "par défaut" tout ou partie d'une description locale réelle par une description conceptuelle. Ceci comporte bien sûr un risque d'imprécision, voire d'inexactitude ; il est donc nécessaire de compléter "manuellement" l'information ainsi déduite.

#### **4.4.5 La logique de particularisation**

Alors que la spécialisation procède par restriction du domaine observable, la particularisation permet au contraire d'élargir le domaine courant pour tenir compte d'un cas particulier. Supposons que nous apprenions que notre ferme fait de l'aquaculture ; donc plus d'animaux à quatre membres, mais des poissons (pisciculture) ou même des huîtres (ostréiculture). Les descriptions vont devoir tenir compte de caractères concernant des écailles, des nageoires, ou des coquilles ; si ces caractères n'étaient pas présents dans le modèle général des animaux de ferme, il va falloir, pour ce cas particulier, procéder à un élargissement du modèle.

Cette démarche, complémentaire de celle de spécialisation, apparaît comme une complication (un peu comme les "verrues" dans un logiciel). Il semble préférable de ne la suivre que dans des situations réellement exceptionnelles, telles qu'il soit justifié de les traiter à part plutôt que de les intégrer dans le moule général.

#### **4.4.6 La logique d'itération**

L'étude des mécanismes précédents s'est appuyée implicitement sur une mise en correspondance d'une part d'une description composée de sous-descriptions (ou de descriptions locales), d'autre part d'un modèle descriptif composé de sous-modèles descriptifs. La description concerne l'observé, le modèle concerne l'observable.

Il arrive fréquemment que, dans une description, plusieurs caractères, bien que non rigoureusement identiques, soient du même "type" et suivent un même sous-modèle descriptif. Prenons l'exemple des dents des Mammifères. Si nous avons

à décrire la denture d'un homme (si nous avons peur de celle du chien), nous voyons bien qu'il existe plusieurs sortes de dents, disons 3 ou 4 sortes selon notre perspicacité. Les plus savants les désigneront d'emblée : incisives, canines, prémolaires et molaires ; mais il n'est nullement nécessaire de connaître leur nom pour les décrire correctement. Il suffit de suivre un sous-modèle commun de description des dents, et de l'appliquer de façon itérative autant de fois que nécessaire (ici 3 ou 4 selon que le descripteur percevra ou non la différence de nature entre les prémolaires et les molaires).

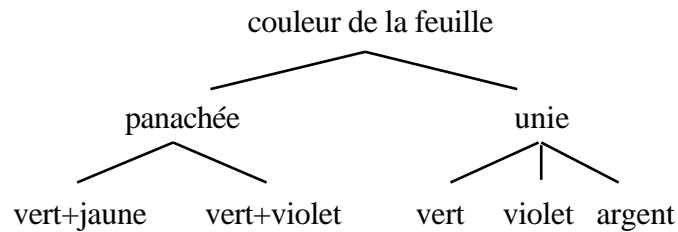
Nous avons en effet vu l'importance du respect du principe d'homologie. Si nous devons comparer dans le détail les dentures du chien et du chat, il faut s'assurer que nous comparons bien les canines (ou "crocs") de l'un avec les canines de l'autre ; sinon, on s'égare. Il faut être conscient du risque d'interprétation (donc de subjectivité) qu'il peut y avoir à s'aventurer dans des "déterminations locales" ; le descripteur, non averti des limites de son savoir, qui désignerait comme des canines les défenses du Morse et celles de l'Eléphant commettrait une erreur, qui par suite le conduirait à comparer des objets non véritablement homologues : les défenses de l'Eléphant sont des incisives modifiées, contrairement à celles du Morse qui sont bien des canines, quoique d'une taille exceptionnelle. Il est vrai qu'il est difficile de se borner à décrire, sans rechercher à comprendre et à apprendre ; mais, paradoxalement, une bonne description ne devrait pas faire appel à l'intelligence, au risque de se trouver biaisée par un modèle mental hélas faillible. En boutade, cela ne désigne-t'il pas les activités de description comme d'excellents clients pour l'intelligence artificielle ?

Une autre situation peut se rencontrer lors des descriptions. Supposons que nous procédions à la description locale de l'inflorescence d'une plante, et que le sous-modèle descriptif correspondant donne comme liste des couleurs possibles blanc, jaune et rouge, et autorise plusieurs réponses (choix multiple). Si nous répondons à la fois blanc et jaune, cela signifie que la couleur est blanc ou jaune, et traduit une imprécision (peut-être une nuance intermédiaire comme un blanc-jaunâtre ?). Pour exprimer que nous observons effectivement les couleurs blanc et jaune simultanément, il faut procéder à deux descriptions locales successives, l'une de fleurs de couleur blanche uniquement, l'autre de fleurs de couleur jaune ; il est en effet fort probable que ces deux types de fleurs ont d'autres caractéristiques qui les distinguent, comme leur localisation au sein de l'inflorescence voire même leur sexe, et qu'elles n'ont pas la même signification organique.

Remarque : Il faut bien distinguer de ce dernier cas la description des associations (de couleur par exemple) bien répertoriées sous le nom de panachures, de mosaïques, etc.. Le fait qu'une feuille soit panachée vert et jaune ne doit pas se traduire par le choix simultané de vert et de jaune, mais par le



choix unique de l'association vert+jaune dûment répertoriée. Ceci peut se traduire dans le modèle descriptif par une hiérarchie de valeurs du type :



Le processus d'itération est donc celui qui doit être mis en œuvre chaque fois qu'il faut exprimer des faits conjoints, c'est-à-dire qui sont constatés simultanément.

#### 4.4.7 Les conditions contextuelles

Les caractères sont généralement dépendants les uns des autres. Plutôt que de déformer la réalité en émettant une quelconque hypothèse d'indépendance (très rarement vérifiée), il convient de tirer parti au maximum de toute l'information véhiculée par ces relations.

Particulièrement fréquentes sont les relations de *concomitance* et d'*exclusion*. Elles traduisent respectivement une condition de présence ou d'absence d'un caractère en fonction du "contexte" formé par d'autres caractères. Par exemple, il intervient dans la classification des Mammifères le fait qu'ils soient ou non dotés d'un placenta (distinction entre les Placentaires et les Aplacentaires) ; il est bien évident que cela ne saurait être observé chez les individus mâles ; si le contexte de la description correspond à un taureau, il est "non pertinent" de s'intéresser à savoir s'il est gravide, ou au nombre des pis portés par ses mamelles. Notons que, comme pour la "valeur «inconnu»", cela n'a pas de sens de parler de "valeur «n. p»" sauf s'il s'agit de combler des cases vides dans un tableau de caractères : l'information liée à la "non pertinence" du caractère gravide est entièrement portée par le fait que le taureau est de sexe mâle, et traduit le fait général de l'exclusion entre masculinité et gestation. La nature est ainsi faite.

On imagine sans mal l'existence de relations de concomitance, quand la présence d'un caractère découle "automatiquement" du contexte. De telles relations ne sont parfois sensibles qu'au spécialiste, et lui permettent de constituer son expertise. Nous prendrons un exemple réel tiré du diagnostic en maladie des plantes, où l'expert, constatant un dessèchement de l'extrémité des feuilles, va regarder à l'endroit le plus inattendu, à savoir le "collet" (base de la tige), s'il n'y a pas là un "chancre" qui entraverait la circulation de la sève; il exploite de la sorte une relation de concomitance, et plus précisément une relation de cause à effet.

Du fait de la variété de la nature, les dépendances entre caractères ne sont que rarement absolues ; par exemple, il existe des dessèchements non dûs à un

chancre du collet, et la nature a horreur des “règles” et “lois” dépourvues d'exceptions : c'est le mâle de l'Hippocampe qui est enceint ! Il importe donc, non seulement de tenir compte des relations de dépendance, mais de pouvoir préciser leurs limites d'application, donc les exceptions et les “contextes” afférents.

Dans de nombreuses situations d'observation, une partie de ce qui devrait théoriquement être observable ne l'est pas ; ou au contraire, une description locale n'est possible que dans certaines conditions. Ceci peut ici encore se traduire par des règles contextuelles, par exemple : Si le chien est méchant, ne pas observer ses dents ; ou : Si l'oiseau est en vol, décrire les marques du dessous de ses ailes. Ces conditions traduisent des “connaissances” de bons sens, et peuvent être utilement exploitées pour guider “intelligemment” l'observation.

## **4.5 Représentation structurée selon un modèle descriptif**

---

Pour un domaine donné, le modèle descriptif est créé par l'expert. Il doit représenter sous une forme structurée tout ce qui est observable.

De plus, le modèle descriptif a pour objectif majeur de constituer un guide d'observation lors des descriptions. Il doit être à même de traduire sans les contraindre l'ensemble des mécanismes ou logiques d'observation mises en évidence précédemment. Il s'agit donc d'une représentation de l'ensemble des connaissances sur l'observable, adaptée à l'acquisition des connaissances sur l'observé.

Pour un domaine donné, le modèle descriptif peut revêtir, de façons équivalentes, plusieurs formes, selon l'utilisateur visé.

De façon profonde, il est représenté sous une forme informatique, adaptée à une base de connaissances observables ; on y trouvera des objets, des “frames”, des listes, des tableaux, des règles, des illustrations, etc., rédigées selon une syntaxe traduisant aussi exactement que possible les différents mécanismes d'observation et les “connaissances de fond” sur le domaine. Sous cette forme, il n'est pas destiné à être lu par le naturaliste ; c'est seulement une représentation technique, servant d'entrée et/ou de sortie aux différents modules de traitement des descriptions. Le modèle informatique se doit bien sûr de suivre un formalisme immédiatement transposable au plan mathématique, afin de permettre l'exploitation des connaissances par les logiciels d'analyse de données symboliques, d'induction et autres.

Au naturaliste spécialiste du domaine, qui travaille à son élaboration et à sa mise au point, le modèle descriptif doit se présenter de façon beaucoup plus pratique et synthétique. Son caractère structuré justifie une présentation sous forme d'un graphe d'organisation du domaine, dans lequel apparaissent au premier aspect les différentes parties ou sous-parties, avec leurs relations et leurs caractéristiques propres. La manipulation de ces "objets" (au sens informatique du terme), pour les créer, les modifier, les déplacer, leur associer des illustrations etc, gagne à se faire par voie graphique, grâce à des outils interactifs, simples d'emploi pour les mettre à la portée des biologistes non informaticiens.

Un dernier habillage, peut-être le plus important pratiquement, permet de présenter le modèle descriptif comme un véritable guide d'observation ; nous l'avons appelé "questionnaire" dans les applications que nous avons développées, car c'est lui qui est mis entre les mains du descripteur, et donne la possibilité de naviguer de manière souple mais logique entre différents écrans de saisie. Chaque écran ("carte" selon l'outil HyperCard ici employé) correspond à l'acquisition d'une description locale, en exacte conformité avec la partie correspondante du modèle descriptif. Notons ici que le modèle descriptif peut prévoir une gradation dans la précision des réponses, voire autoriser la fameuse réponse "?" qui traduit l'incertitude totale ; ceci est absolument essentiel lors de descriptions réelles, où le contexte ou les circonstances n'autorisent pas toujours des descriptions complètes (*Cave canem !*).

Les descriptions obtenues, dont la cohérence est assurée à chaque étape et dont la complétude est vérifiée en fin de saisie, peuvent être présentées à leur tour sous plusieurs formes. Leur forme initiale, qui est celle du questionnaire une fois rempli, peut être ré-importée pour apporter des corrections ou des compléments de description. Mais il est parfois utile de pouvoir les visualiser sous forme d'un sous-graphe instancié du modèle descriptif ; cette présentation permet de faire ressortir la structure sous-jacente à la description, qui est quelque peu perdue de vue lors de la navigation dans le questionnaire. Ces deux formes sont en fait très complémentaires, et la possibilité doit être donnée de basculer aisément de l'une à l'autre.

D'autre part, il est presque nécessaire de pouvoir présenter à l'utilisateur les descriptions sous forme d'un texte en langage naturel, ainsi qu'il y est habitué depuis toujours ; il n'est pas difficile à ce niveau d'offrir le choix entre plusieurs langues-cibles. Enfin, les descriptions saisies sont archivées avec la même représentation informatique que les modèles descriptifs, par souci d'efficacité et d'homogénéité ; de la sorte, les connaissances sur l'observable et celles sur l'observé bénéficient du même formalisme particulièrement adapté, ce qui permet de les utiliser de façon conjointe et de rendre plus cohérents et plus efficaces les programmes qui les traitent.

Nous détaillerons au chapitre 6 les différentes solutions techniques qui nous ont permis de formaliser au plan informatique les mécanismes observationnels décrits ci-dessus. Il est fait appel à des “frames” comme base de la structure. Les frames sont des “objets” possédant des attributs propres. Chaque attribut peut prendre une ou plusieurs valeurs possibles (dans une liste, éventuellement hiérarchisée, s'il s'agit de valeurs symboliques ; dans un intervalle pour les valeurs quantifiées) ; une fois valué, l'attribut exprime un caractère décrit ou trait. Quand les objets correspondent à des sous-parties (mais pas à des points de vue), leur absence constatée est enregistrée comme significative. Les mécanismes de spécialisation et de particularisation sont classiquement traduits par l'instanciation de “classes” (au sens informatique du terme) avec héritage. Le mécanisme d'itération implique quant à lui la mise en œuvre d'une logique du premier ordre, avec l'utilisation de “variables”. Enfin les conditions contextuelles sont exprimées sous forme de règles ou de démons.

Il est ainsi possible, grâce au recours à des méthodes de représentation des connaissances issues de l'intelligence artificielle, de formaliser des descriptions aussi complexes que l'exige la “vérité” de la Nature, sans biais de transposition, sans avoir recours à la subjectivité, et limitant la perte d'information autant que voulu.

Il existe un bon moyen de s'assurer de la qualité des descriptions ainsi obtenues. Il suffit en effet de comparer ces descriptions, produites sous leur forme rédigée en langage naturel, avec celles directement réalisées par les spécialistes du domaine. Il est très facile d'évaluer alors les défauts des unes et des autres ; ceci indépendamment du fait que les descriptions “conformes” (au modèle descriptif) possèdent l'énorme avantage d'être comparables entre elles et aisément mobilisables.

## **4.6 Illustration sur les données des éponges marines**

### **4.6.1 Acquérir l'observable**

Représenter le modèle descriptif, c'est-à-dire acquérir l'observable, est l'étape la plus importante de la conception des systèmes de détermination. De la qualité du modèle descriptif dépendra la qualité des descriptions et, par voie de conséquence, la qualité des classifications et des déterminations futures.

Le modèle descriptif regroupe toutes les connaissances observables d'un individu du domaine des *Hyalonema*. Les objets du modèle correspondent aux différents composants descriptifs de cet individu. L'élaboration de ce modèle est un processus mettant en jeu les connaissances de l'expert du domaine et le cognicien chargé de représenter ces connaissances dans la syntaxe du langage basé sur les frames (voir § 6.4.2). Ce processus peut être itératif dans la mesure

où des corrections sur le modèle se révéleront souvent nécessaires après la phase d'exploitation de celui-ci.

Cette première étape s'appuie sur l'expérience de l'expert, c'est-à-dire sur ses observations initiales : la genèse de l'observable se fait donc forcément à partir de l'observé préexistant. Il est très important que le spécialiste focalise son attention à ce niveau car là se trouve reproduite une grande partie de son expertise qui sera ensuite utilisée par les outils. L'expert doit faire une analyse exhaustive de la variabilité de l'observé, afin d'être en mesure de synthétiser un modèle complet de l'observable. Cette phase est bien évidemment contraignante mais elle permet d'éviter des retouches futures du modèle, qui ne devront être que locales et ne pas affecter sa structure globale.

Les logiques de description que nous avons introduites précédemment montrent les différents aspects théoriques de la modélisation d'un domaine biologique. Plus pratiquement, ces aspects se retrouvent à différents niveaux dans la conception d'un modèle descriptif tel que celui sur le Genre *Hyalonema*.

Les concepts descriptifs n'ont pas le même niveau sémantique, certains se plaçant au niveau de la décomposition d'un domaine, d'autres au niveau d'une description locale d'un composant du domaine. C'est pourquoi il est très important de les différencier dans l'optique de construire un questionnaire ayant la qualité d'un bon guide d'observation.

Nous avons conçu le modèle observable à trois niveaux de description : objet-attribut-valeur que nous explicitons dans les trois paragraphes suivants. Ensuite, nous caractérisons chacun de ces trois niveaux plus précisément, ce qui constitue notre méthode d'élaboration du modèle descriptif.

#### **4.6.2 Décomposition de l'entité globale en objets**

Ce procédé correspond à la **logique de composition** (§ 4.4.2). Cette logique naturelle introduit la relation de dépendance entre les objets. En effet, le mécanisme le plus évident consiste à partitionner la description en sous-descriptions plus ou moins indépendantes les unes des autres et que l'on appelle des *descriptions locales*. Ceci est obtenu en décomposant l'entité physique observable en sous-parties, chacune d'elles donnant accès à ses propres sous-descriptions. Nous avons déjà mentionné au chapitre 3 qu'à l'échelle d'une description individuelle, l'objet équivalait à un composant de l'individu.

Par exemple, une éponge est une entité constituée de parties physiques avec certaines relations entre elles, dont principalement la relation de sous-partie. On décomposera donc l'éponge en différentes sous-parties décrites localement par des caractéristiques propres. Cette décomposition suit le principe logique qui veut que l'on décrive les objets en allant du plus général au plus particulier.

Ex : l'éponge possède un corps et un pédoncule, le corps est constitué de...

### 4.6.3 Recherche des attributs de chaque objet

Parmi les faits observables d'un domaine, il y a les objets observables liés entre eux par des relations, ainsi que leurs caractères observables (caractéristiques, propriétés, variables ou attributs) et les différents états possibles de ces caractères (valeurs d'attributs).

Un objet observable est un élément de description qui possède des caractéristiques propres. Il est défini par son existence en tant que composant au sein de l'entité ou individu à analyser. Par opposition, les attributs n'existent pas par eux-mêmes sans référence à un objet préexistant. Chaque objet forme donc un tableau avec la liste de ses attributs propres dans autant de colonnes.

Considérons l'attribut "taille" de l'objet "corps" de l'éponge. Cette "taille" n'existe pas sans faire référence à l'objet "corps", alors que le "corps" existe par lui-même dès lors que l'on considère le domaine des *Hyalonema*. Pour différencier les objets des attributs pour un domaine particulier, il suffit de se poser les questions suivantes : "Y a t-il un... objet ?" et "Y a t-il un objet ayant tel ... attribut ?".

De même, nous aurions pu tout aussi bien considérer la taille de l'objet "amphidisque" qui est un élément microscopique de l'éponge. Bien qu'ayant le même nom de propriété, le domaine de définition des valeurs prises par l'attribut est différent. La liste des valeurs possibles de l'attribut dépend donc de l'objet considéré.

### 4.6.4 Recherche des valeurs de chaque attribut

Une valeur observable correspond à un état possible de l'attribut de l'objet pour l'individu à observer. L'ensemble des états possibles de l'attribut pour l'objet observable forme l'**espace d'observation** noté  $O$ . Cet espace relatif à l'objet étudié ne doit pas être confondu avec le **référentiel** (ou domaine de définition) qui est l'ensemble de tous les symboles relatifs à l'attribut pris isolément. Par exemple, le référentiel de l'âge en années est l'ensemble des réels positifs alors que l'espace d'observation de l'âge d'un humain est un sous-ensemble de  $\mathbb{R}^+$  tel que  $[0, 120?]$ .

Les valeurs d'un attribut devraient avoir comme caractéristique d'être mutuellement exclusives (sans recouvrement). Cette propriété est nécessaire pour permettre de comparer sans ambiguïté des descriptions [Vignes, 1991]. Ainsi pour tout attribut, un objet ne possède normalement qu'un seul état. Si au moment de son observation, on lui associe plusieurs valeurs, ou un intervalle de

valeurs, cela ne peut traduire qu'une imprécision et non pas une gamme de variation (voir § 4.6.9).

La valeur est reconnaissable en répondant à la question "Y a-t-il un objet avec un attribut qui vaut... valeur ?"

### Résumé

Pour créer un objet, un attribut ou une valeur, on doit donc se poser les trois questions suivantes :

l'objet est-il un composant descriptible de l'objet ... ?

l'attribut est-il une caractéristique substantive propre de l'objet ... ?

la valeur est-elle un qualificatif de l'attribut ... ?

Certains auteurs comme Lebbe [Lebbe, 1991] appellent nos valeurs des attributs (avec une acception linguistique), nos attributs des qualités, les objets des sujets et la composition "qualité[sujet(s)]" est appelé un descripteur (connotatif au sens de [Colless, 1967]). Nous préférons employer la terminologie objet-attribut-valeur des langages de frames car nous pensons qu'elle permet de mieux mettre en valeur les trois niveaux grammaticaux de description de l'observable.

## 4.6.5 Les propriétés d'un objet observable

### 4.6.5.1 Statuts

Tout objet observable créé est par nature potentiellement présent : par la suite, il peut donc être décrit (la présence d'un objet est la condition nécessaire et suffisante à sa description). De plus, au moment de la création d'un objet observable, on peut définir ses alternatives d'observation possibles (1, 2 et 3) ainsi que sa sémantique (4) lors des futures descriptions :

- 1) l'absence de l'objet peut ou non revêtir une signification classificatoire,
- 2) lorsque le contexte de l'observation ne permet pas de le décrire, la réponse "inconnu" est autorisée ou non,
- 3) on pourra ou non en observer différentes sortes,
- 4) l'objet a une signification classificatoire, ou au contraire il n'intervient que pour structurer le domaine.

1) L'absence d'un objet est une information à prendre en compte pour deux raisons : la première est qu'elle permet d'inférer des règles implicites très fortes : les sous-parties d'un objet absent sont nécessairement absentes. La seconde est qu'elle possède un sens pour la classification.

Considérons l'objet "amphidisque", ce micro élément peut être absent d'une éponge de manière naturelle, ce qui peut être une information discriminante.

Inversement, le pédoncule de l'éponge est un objet qui ne peut être naturellement absent d'un individu (dans le cas des *Hyalonema*) : le statut "absent" n'est pas admissible. Il arrive néanmoins que cet objet ne soit pas présent pour un spécimen à observer (du fait que le pédoncule a été arraché du corps lors de la récolte par exemple). L'absence porte alors la même signification que l'inconnu dans ce contexte car l'expert sait qu'il a existé un pédoncule au spécimen mais on ne peut pas le décrire (voir statut 2).

2) La possibilité pour un objet d'être inconnu est en principe valable pour tous les objets dans notre application ; c'est pourquoi cette information n'apparaît pas explicitement dans le modèle descriptif. La réponse "inconnu" traduit le fait qu'il est impossible de décrire l'objet à cause du contexte d'observation : nous décrivons un spécimen à partir d'une photo où l'objet à observer est masqué ou dans l'ombre, le grossissement du microscope est insuffisant pour voir les détails de l'objet, etc..

Cependant, il pourrait avoir un sens que la réponse "inconnu" ne soit pas autorisée pour certains objets, ce qui serait une manière de forcer la description locale d'un objet. Par exemple pour le corps de l'éponge, il ne serait pas autorisé de répondre "inconnu" pour cet objet car l'expert n'accepterait pas une description limitée au pédoncule d'une éponge, alors que la plupart des caractères majeurs sont ceux du corps (cette situation ne s'est d'ailleurs jamais observée concrètement dans les descriptions effectuées sur les *Hyalonema*).

Comme pour l'absence d'un objet, la reconnaissance du fait qu'une partie est inconnue permet d'inférer que ses sous-parties sont inconnues.

3) Un objet multiple est un objet dont plusieurs sortes peuvent voisiner dans une même description. La multiplicité d'un objet ne doit pas être confondue avec sa cardinalité. Elle signifie le *nombre de sortes possibles* d'objets du domaine alors que la cardinalité veut dire le *nombre possible* de cet objet. La multiplicité d'un objet peut s'exprimer à l'aide de deux chiffres : le premier indique la multiplicité minimale de l'objet, le second sa multiplicité maximale.

Par exemple, si l'on voulait décrire les membres d'un humain, on pourrait dire que la cardinalité de cet objet est comprise entre 0 et 4 : il s'agit du nombre possible de membres (1, 2, 3 ou 4) en tenant compte des accidents. A l'opposé, la multiplicité de [2, 2] indiquerait le nombre de sortes possibles de membres à décrire chez cet humain : les pieds, les mains. Par contre, pour les singes, la multiplicité pourrait être 1 et 2 selon qu'il s'agit d'un quadrumane (quatre mains) ou d'un bipède (deux mains et deux pieds), et nous la noterions [1, 2].



La multiplicité d'un objet correspond à la **logique d'itération** (§ 4.4.6). Dans une même description, il peut être utile de faire apparaître plusieurs instances ("sortes") d'un même objet sans avoir à les nommer explicitement. On a constaté en effet que le nom d'un objet observable n'est pas toujours maîtrisé par un observateur du domaine autre que l'expert (voir § 4.6.7). C'est pourquoi l'utilisateur doit avoir la possibilité de décrire autant de sortes d'objet qu'il le souhaite dans la limite de la multiplicité maximale imposée par l'expert.

Lorsque chaque type d'objet est faiblement différencié par rapport aux autres, c'est-à-dire avec des différences minimales liées au positionnement, à la symétrie (par exemple les pouces des mains), au nombre, on parle alors de **variation méristique** [Perinet-Marquet, 1993] [Encyclopédie Quillet, 1993]. Cette variation n'affectant pas la nature même de l'objet, la logique d'itération est la plus appropriée pour décrire différentes sortes sans les nommer. Inversement, si la variation correspond à une différenciation<sup>2</sup> suffisamment forte pour être reconnue par la grande majorité des observateurs, il est plus approprié d'utiliser la logique de spécialisation (par exemple différencier et décrire séparément les pieds et les mains chez le bipède).

4) Nous dirons qu'un objet est *fictif* quand il ne correspond pas à une caractéristique discriminante du domaine, mais que sa vocation est d'aider à la structuration du domaine. Dans le cas contraire, on dit qu'il est *significatif* (l'objet correspond à une réalité physique intéressante pour la discrimination). Les objets fictifs sont utiles à la structuration du domaine et à la fabrication d'un bon guide d'observation. Prenons l'exemple des objets fictifs "macro constituants" et "micro éléments" : il est certain que ces concepts ne sont pas importants pour la classification des espèces de *Hyalonema*, ils n'ont pas de signification classificatoire<sup>3</sup>. Par contre, pour structurer le domaine, ces objets sont intéressants car ils permettent de ne pas présenter au même niveau d'observation des points de vue macro et microscopiques, ce qui correspond à un contexte différent de description des objets.

Au statut fictif de l'objet correspond la **logique des points de vue** (§ 4.4.3).

#### 4.6.5.2 Attributs

Les attributs sont des propriétés propres à l'objet, ils sont caractérisés par :

##### ① un type

<sup>2</sup> Morphologique et/ou fonctionnelle.

<sup>3</sup> On ne doit pas discriminer deux groupes d'éponges sous prétexte que les préparations microscopiques étaient disponibles pour l'un et pas pour l'autre.

Un attribut de type numérique peut être un **entier naturel** ou un **réel** et sa valeur peut être soit simple soit un intervalle. La fourchette de valeurs que l'utilisateur indique pour un intervalle représente l'imprécision globale attachée à la mesure de l'attribut numérique.

Un attribut de type qualitatif peut être **nominal** ou **classifié** en référence à la manière dont sont structurées les valeurs possibles (voir le paragraphe suivant).

Un attribut **texte** est une information supplémentaire sur le cas à décrire mais qui ne devra pas être prise en compte lors du traitement des cas observés. Il s'agit simplement d'une chaîne de caractères utile à titre documentaire pour la maintenance de la base de cas. Par exemple, la longitude est un attribut commentaire (129° 57' E) de l'objet fictif contexte qui n'est pas utilisé lors de la discrimination (figure 4.1) :

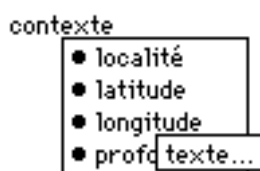


Fig. 4.1 : Exemple de l'attribut "longitude" de type "texte"

Parmi les attributs de type texte, on mentionnera l'attribut "image" qui permet de référencer un numéro correspondant à une image numérisée illustrant tel ou tel objet décrit d'un cas particulier.

Un attribut de type **booléen** est un attribut dont la valeur est soit oui, soit non : par exemple, le pédoncule de l'éponge peut être soit torsadé, soit non torsadé (figure 4.2) :

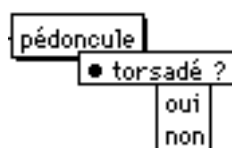


Fig. 4.2 : Exemple d'attribut de type "booléen"

Remarque : le point d'interrogation permet de montrer que l'adjectif torsadé peut être transformé en un nom d'attribut à qualifier par la réponse alternative oui-non. Il se trouve que l'expert ne perçoit pour l'objet pédoncule que l'utilité de décrire le fait qu'il peut être torsadé ou non. Nous aurions pu tout aussi bien choisir un véritable attribut comme l'aspect du pédoncule avec un type nominal et les valeurs "torsadé", "non torsadé".

Mais si plus tard, l'expert découvre qu'il existe des spécimens avec un pédoncule d'aspect tressé, ou s'il s'intéresse à la résistance ou à la longueur du

pédoncule, il faudra homogénéiser la description de cet objet en introduisant l'aspect, la résistance et la longueur comme autant d'attributs.

### ② des valeurs possibles

Il s'agit de l'ensemble parmi lequel un attribut prend sa valeur. Ces valeurs sont définies par :

une unité de mesure et un intervalle (minimum-maximum) lorsque l'attribut est numérique,

exemple : l'attribut profondeur de l'objet contexte (indiquant la profondeur à laquelle a été récoltée l'éponge) est compris entre 100 et 5000 m (figure 4.3) :



Fig. 4.3 : Exemple d'attribut de type "numérique"

une liste (ordonnée ou pas) lorsqu'il s'agit d'un attribut nominal, des valeurs numériques en nombre restreint pouvant aussi apparaître comme valeurs d'un attribut nominal,

exemple : l'attribut couleur du corps de l'éponge peut prendre les valeurs blanchâtre, grisâtre, autre (figure 4.4) :

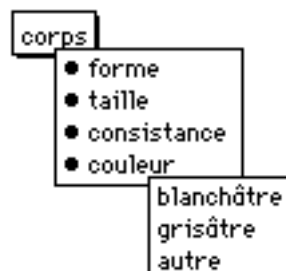


Fig. 4.4 : Exemple d'attribut de type "nominal"

Remarque : La réponse "autre" renvoie à un commentaire en texte libre (cf. § 4.6.5.3)

par une hiérarchie de classification dans le cas d'un attribut classifié.

Exemple : l'attribut "forme" du corps de l'éponge peut être classifié de la manière suivante (figure 4.5) :



Fig. 4.5 : Exemple d'attribut de type "classifié"

On aurait pu représenter cette hiérarchie sous forme d'arbre (cf. § 4.4.6).

### ③ une valeur par défaut

Une valeur par défaut correspond éventuellement à la valeur la plus fréquemment observée, elle est mentionnée par l'expert dans le modèle mais dans le questionnaire, l'utilisateur peut lui substituer une autre valeur pour un cas observé.

Exemple : la couleur du corps des *Hyalonema* est blanchâtre par défaut.

En fait, toutes les *Hyalonema* sont blanchâtres. Il a cependant été jugé utile de rentrer cette information, non discriminante, pour le cas où le domaine serait étendu à d'autres Genres d'éponges marines, où la couleur est connue comme un caractère distinctif.

#### ④ un choix

Il faut préciser si l'utilisateur a droit à une seule réponse (choix unique) ou à plusieurs (choix multiple). S'il a droit à plusieurs réponses, cela lui permet de traduire une incertitude : l'expert reconnaît alors qu'il n'est pas toujours possible à l'utilisateur de prendre une décision univoque entre les valeurs possibles proposées. Ceci est typique pour les couleurs par exemple (quelle est la différence entre brun et marron ?). Le choix multiple permet à l'utilisateur de prendre une décision partielle (exprimer un doute entre des éléments d'un sous-ensemble de valeurs) :

ex : la forme du corps de l'éponge prend deux valeurs : "en-cône, en-trompette".

Il s'agit bien là d'une imprécision sur la forme d'un objet particulier (la virgule possède la sémantique du "ou") et non pas de la traduction d'une présence simultanée de deux valeurs. Pour que la présence simultanée soit valide, il faut qu'elle soit dûment répertoriée dans la liste des valeurs possibles : "en-cône+en-trompette" par exemple. Nous rappelons que toutes les valeurs sont mutuellement exclusives et que toutes les combinaisons valides doivent être introduites dans la liste des valeurs possibles. Si aucun choix ne convient à l'utilisateur, il doit répondre "inconnu" et non pas biaiser la description en associant des valeurs avec la sémantique du "et" (présence simultanée de deux valeurs non ambiguës) alors qu'elles seront interprétées au niveau du traitement comme une imprécision sur le choix d'une valeur unique.

Cette remarque n'est pas anodine. Elle oblige l'expert à concevoir un modèle robuste pour éviter les descriptions ambiguës. Cela est d'autant plus vrai quand la description doit porter sur un concept représentant en fait un ensemble d'objets d'une même sorte et non pas un ensemble singleton (voir § 4.6.9). Par exemple, dans une application sur le diagnostic des maladies de la tomate, au niveau de la localisation de la tache sur foliole, le concept de tache représente une population de taches homogènes (de la même sorte). Si ces taches sont positionnées à l'extrémité de la foliole et entre les nervures, il faut représenter la valeur "extrémité+entre-nervures" explicitement. De plus, l'expert doit prévoir que l'attribut ne soit pas sémantiquement équivoque : par exemple, la localisation des amphidiskues dans le corps de l'éponge donnait droit à un choix multiple avec comme valeurs possibles : éparses, périphériques, à-l'intérieur. Nous nous sommes aperçu de l'ambiguïté du terme "éparses" qui représentait plutôt un attribut de densité (quantitative) à la place d'une véritable localisation topographique (qualitative). Le choix multiple était donc un biais pour donner deux valeurs simultanément au niveau d'un seul attribut polysémique. Nous avons remplacé le terme "éparses" par "partout" et nous avons remplacé le choix multiple par un choix unique, ce qui fournit une interprétation bien meilleure et sans confusion (figure 4.6) :



Fig. 4.6 : Exemple d'attribut non polymorphe

Selon notre approche, pour qu'un attribut soit cohérent, il faut que ses états soient des réponses alternatives à une question [Lebbe, 1991]. La cohérence du choix des valeurs possibles d'un attribut est un facteur essentiel de la robustesse du système de détermination.

#### 4.6.5.3 Autres propriétés

Ces dernières caractéristiques sont plutôt de nature ergonomique, elles permettent de structurer le dialogue avec l'observateur dans le questionnaire pour chaque objet et chaque attribut :

##### ① genre de l'objet

La connaissance du genre de l'objet dans le modèle descriptif permet d'éviter à l'expert de fournir explicitement la question associée à chaque objet. Ces questions sont de nature répétitive et gagnent à être générées automatiquement et de manière lisible à partir de la connaissance des genres de chaque objet. Il s'agit du genre, au sens grammatical, du nom pris par un objet, à savoir : masculin singulier, masculin pluriel, féminin singulier et féminin pluriel.

Ex : Au lieu de poser la question "Caractéristiques de l'objet corps de l'objet Hyalonema ?" qui est indépendante du genre des deux objets cités, on pourra construire automatiquement deux questions relatives aux caractéristiques et aux composants de l'objet "corps" : "Caractéristiques du corps de la Hyalonema ?" et "Composants du corps de la Hyalonema ?" relativement à la description locale du corps et à celle de ses sous-parties.

##### ② question associée à l'attribut

Il s'agit de la question posée dans le questionnaire qui permet de comprendre le sens des mots associés (attribut et valeurs possibles) et d'y répondre précisément par le choix d'une valeur appropriée. L'écriture de cette question n'est pas automatisable du fait justement qu'elle doit être précise :

Par exemple, pour la localisation des amphidiskues de l'éponge, la question pourrait être : "Où ces amphidiskues sont-ils localisés ?"

### ③ Commentaires et remarques

Il peut néanmoins se trouver que dans le questionnaire, au moment où l'utilisateur désire décrire un objet, il ne comprenne pas le sens de la question ou encore le sens du vocabulaire utilisé par les attributs et les valeurs. Il peut se trouver aussi en face d'un spécimen à décrire dont aucun des états proposés pour l'un de ses objets ne correspond à son observation. Au lieu de répondre "inconnu", il pourrait manifester son incompréhension en rédigeant un texte libre dont l'expert pourra par la suite tenir compte. Cette faculté est importante pour instaurer un dialogue entre l'expert et les utilisateurs de son modèle descriptif car cela peut lui permettre une rentrée d'informations pertinentes auxquelles il n'avait pas pensé plus tôt. Par exemple, la valeur "autre" de la figure 4.4 pourrait renvoyer à un commentaire et l'observateur y indiquerait la couleur du spécimen qu'il décrit.

## 4.6.6 Niveau de précision d'un objet

### 4.6.6.1 Spécialisation

Lorsque l'on veut faire intervenir de la connaissance supplémentaire sur les objets, les attributs et sur les valeurs sous forme d'une plus grande précision, on introduit le lien de spécialisation. Cela permet de construire des regroupements utiles afin de faire hériter les propriétés des objets les plus généraux vers les objets les plus particuliers : un objet se spécialise en un "sous objet" et toutes les caractéristiques communes aux deux objets sont "remontées" au niveau de l'objet le plus général. On aboutit ainsi à former des taxonomies d'objets selon un certain degré de généralité (ou de précision) (figure 4.7) :

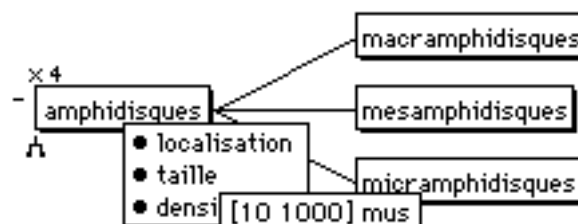


Fig. 4.7 : Exemple de spécialisations d'un objet

Pratiquement, les objets les plus bas dans la taxonomie héritent des propriétés (attributs et valeurs possibles) des objets situés plus hauts. Il en est de même des composants de l'objet plus général. Dans le cadre de cette logique de spécialisation (§ 4.4.4), il s'agit de "masquer" les attributs situés plus haut dans la hiérarchie en restreignant le domaine des valeurs possibles de chacun de ces attributs.

Dans l'exemple ci-dessus, les macramphidiques, les mésamphidiques et les micramphidiques sont des spécialisations d'amphidiques, c'est-à-dire que l'intervalle des tailles possibles pour chaque spécialisation est théoriquement restreint par son type : macro- (500 à 1000 mus<sup>4</sup>), méso- (100 à 500 mus) et micro- (10 à 100 mus). Nous verrons que les spécialisations imposent des contraintes sur la description, ce qui peut engendrer des descriptions erronées (cf. § 4.6.7).

#### 4.6.6.2 Particularisation

Au contraire, dans la logique de particularisation (§ 4.4.5), on élargit le domaine en définissant de nouveaux attributs et de nouvelles valeurs particulières définies plus bas et qui n'apparaissent pas aux niveaux supérieurs. Cette logique est adaptée au traitement des exceptions.

Comme exemple de particularisation, prenons la description de l'objet "Mammifère" dans une application sur la reconnaissance de différents types de Mammifères. L'objet "Ours" est une spécialisation de l'objet "Mammifère" alors que l'objet "Ornithorynque" est une particularisation de l'objet "Mammifère". En effet, en plus des propriétés qu'il aura héritées de l'objet "Mammifère", à savoir de posséder des mamelles, des poils, d'avoir quatre membres, une queue..., l'objet "Ornithorynque" aura ses attributs particuliers (avoir un bec, des pieds palmés, pondre des œufs, etc.). Ces attributs ne sont pas introduits plus haut dans la hiérarchie au niveau de l'objet "Mammifère" car ils fausseraient la juste vision que l'on peut avoir des Mammifères en général, qui n'ont ni bec, ni pieds palmés, etc..

Les processus de spécialisation et de particularisation des objets sont des caractéristiques des langages à objets qui permettent de tenir compte du contexte d'observation. Ainsi s'assure-t-on que les caractéristiques examinées pour un contexte donné sont toujours pertinentes ; or la pertinence est une composante importante de la robustesse du modèle descriptif.

---

<sup>4</sup> Micromètres.



#### 4.6.7 Possibilité de multi-instanciation de chaque objet

Dans une même description, il peut être utile de faire apparaître plusieurs instances (“sortes”) de la même partie sans avoir à les nommer explicitement. Ce constat provient du fait que le nom d'un objet observable n'est pas toujours maîtrisé par un observateur autre que l'expert. Le vocabulaire qu'il utilise pour le désigner est sujet à un certain arbitraire dont l'expert a conscience : il ne souhaite pas contraindre l'observateur avec son propre vocabulaire spécialisé (“charabia” scientifique). L'exemple pour comprendre cette idée est illustré sur la figure 4.8 suivante :

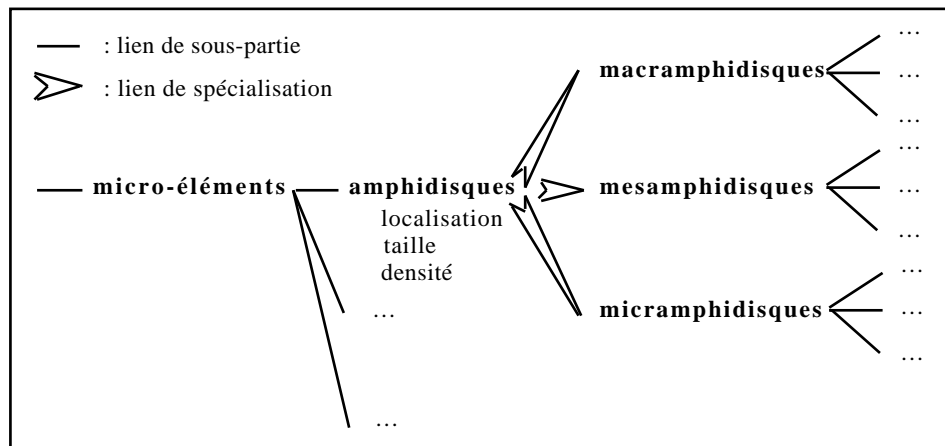


Fig. 4.8 : Exemple de hiérarchie d'objets non instanciés

Ici, les termes de macramphidisque, mésamphidisque et micramphidisque sont des dénominations de spécialiste qui désignent des sortes d'amphidisque dont le type dépend principalement de la taille. Or, il s'agit d'une spécialisation qui n'a pas de sens dans certains contextes d'observation. Il suffit d'imaginer deux types d'éponges à décrire dont le premier est de taille cinq fois plus grande que le second. Premièrement, un utilisateur décrivant l'échantillon de grosse taille pourrait très bien nommer macramphidisque un amphidisque qui serait en fait un micramphidisque pour l'expert. Ce problème est bien connu des systématiciens sous le terme du respect de l'**homologie** entre objets. De plus, lors de l'induction, le système serait incapable de distinguer les différents types d'amphidisque selon leur taille, mais procéderait à l'unification en acceptant le caractère isomorphe des objets dont on n'est pourtant pas certain de la correspondance structurelle. La subjectivité du nom de ces spécialisations ou le fait que ce nom ne soit pas maîtrisé par l'utilisateur final du système nous oblige à définir différentes sortes d'un même objet “amphidisque” (figure 4.9). Le non spécialiste peut ainsi décrire autant de sortes d'amphidisque qu'il le souhaite, de manière plus souple. Le nombre maximal d'instances ou de sortes possible pour un objet est néanmoins fixé préalablement par l'expert dans le modèle descriptif, il s'agit de la *multiplicité* de cet objet (§ 4.6.4.1). On sera donc amené à décrire plusieurs instances en nombre limité d'une même partie au sein d'une description

d'éponge marine, chaque "sorte" suivant un schéma analogue de description et ne se distinguant que par des différences méristiques (§ 4.6.5.1.3) :

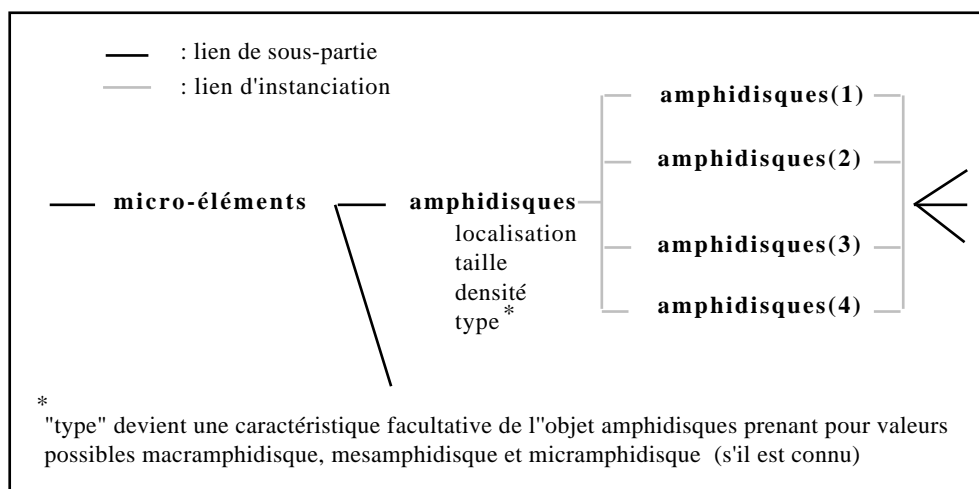


Fig. 4.9 : Exemple de hiérarchie d'objets instanciés

Remarque : il pourra être intéressant de comparer *a posteriori* les réponses des utilisateurs sur les tailles des amphidisques et leur type avec la catégorisation effectuée par l'expert *a priori*. Une classification locale au niveau de l'objet pourrait de même permettre de connaître l'acception de termes selon la majorité des utilisateurs biologistes.

La multi-instanciation d'objets correspond à la **logique d'itération** (§ 4.4.6).

#### 4.6.8 Définition des dépendances inter / intra objets

Les différentes parties peuvent être plus ou moins dépendantes les unes des autres :

- 1 - On a vu précédemment que l'absence d'une partie ou l'ignorance sur son existence implique l'absence ou l'ignorance sur l'existence de ses sous-parties,
- 2 - Certaines valeurs d'une caractéristique d'une partie peuvent aussi rendre inobservable une ou plusieurs de ses sous-parties,
- 3 - Certaines valeurs d'une caractéristique d'une partie peuvent conditionner la ou les autres valeurs d'une autre caractéristique d'une autre partie (ou de la même),
- 4 - Certaines valeurs d'une caractéristique d'une partie peuvent masquer l'accès à une ou plusieurs caractéristiques de cette partie ou de ses sous-parties.

Cela va se traduire par l'écriture de règles ou contraintes déclenchées localement sur certaines parties et qui vont agir sur l'existence d'autres parties ou de

certaines de ses caractéristiques ou bien encore sur la mise à jour de certaines valeurs attachées aux caractéristiques de ces parties. Ces règles expriment les **conditions contextuelles** (§ 4.4.7) d'accès aux objets et attributs pour pouvoir les décrire et permettent le maintien de la cohérence dans le questionnaire ou guide d'observation.

#### 4.6.9 Niveau de généralité d'un objet

Nous avons déjà mentionné au chapitre 3 qu'un objet biologique prenait deux sens différents selon le point de vue d'observation de l'utilisateur : soit il s'agit d'un **individu** d'une population, soit c'est un **composant** d'un individu (voir l'exemple des Mammifères au § 3.6).

Parmi les objets, on peut distinguer aussi ceux dits "observables" et ceux dits "observés". Les premiers correspondent à l'étape d'acquisition du modèle descriptif alors que les seconds concernent la phase de description. Chaque modèle descriptif est constitué d'objets observables alors que chaque description est constituée d'objets observés.

Dans le chapitre 5, nous allons formaliser les descriptions d'objets observés pour les mathématiciens. Traditionnellement, les objets observés à analyser forment une ligne dans un tableau de données classique. Cette disposition ne permet pas facilement de faire la distinction entre l'objet au sens de l'individu et l'objet au sens du composant. Du fait de l'approche par structuration du domaine que nous avons conçue, chaque objet observé forme un tableau avec ses propriétés propres. L'individu est alors décrit par une suite de tableaux reliés entre eux par la relation de composition : ces objets correspondent aux descriptions locales et sont appelés *objets composites*.

Pour chaque description locale d'objet, on retrouve la liste des attributs propres ainsi que les valeurs affectées par l'utilisateur du questionnaire. La valeur vient s'inscrire dans la case du tableau de l'objet à l'intersection entre la colonne de l'attribut et la ligne de l'individu étudié. La valeur doit véhiculer une information pertinente mais non nécessairement discriminante. C'est pourquoi toute réponse inconnue pour un attribut donné dans le questionnaire n'est pas répertoriée dans la description de l'individu : la réponse "inconnu" n'est pas une information en biologie !<sup>5</sup>

---

<sup>5</sup> Dans d'autres domaines, elle pourrait avoir un sens. Par exemple, et par boutade, le fait de répondre "inconnu" à la question "Quel est votre âge ?" n'est elle pas réputée caractéristique de la population féminine ?

Il est encore important de préciser pour chaque objet sur lequel porte une description s'il est :

- 1) unique et concret,
- 2) synthétique, résumant de manière plus ou moins statistique un ensemble d'objets individuels.

#### 4.6.9.1 Représentant unique

Dans le premier cas, nous avons affaire à une *description d'un objet particulier* avec les conséquences suivantes :

il s'agit d'une véritable description, l'objet ayant comme valeur d'attribut un seul état observé de l'espace d'observation qui résume toute la connaissance,

l'extension en retour de cette description est l'objet lui-même, la description étant suffisamment exhaustive et précise pour ne représenter que l'objet lui-même (ou ses clones),

chaque variable de l'objet a une valeur nécessairement unique mais plus ou moins précise<sup>6</sup>. Si la valeur est précise, cela n'empêche pas qu'elle puisse être erronée, c'est-à-dire que l'observateur puisse attribuer un mauvais état parmi ceux proposés,

L'imprécision se traduit par une distribution de **possibilité** au niveau des valeurs prises par les objets [Haton *et al.*, 1991]. Ces possibilités sont des hypothèses (un ensemble d'états) émises par l'observateur qui traduisent son opinion pour "encadrer" le véritable état qu'il n'est pas en mesure de donner. A chaque hypothèse peut être associée une vraisemblance (sous forme de degré de croyance par exemple) pour renforcer le poids de certaines d'entre elles [Dubois & Prade, 1987],

L'objet peut aussi bien être l'individu lui-même qu'un composant de cet individu. Telle valeur de la taille du spécimen d'éponge ainsi que celle de la longueur du pédoncule de l'éponge sont deux descriptions d'objets uniques,

---

<sup>6</sup> Si la valeur n'est pas unique (représentée par un intervalle numérique ou bien par un choix multiple), c'est parce que l'observateur doute ou bien n'a pas la possibilité de prendre une décision univoque entre les valeurs possibles proposées. L'incertitude sur la valeur peut être due soit à l'utilisation d'appareils de mesure pour les attributs numériques, soit laissée à l'appréciation subjective de l'observateur. Cette imprécision est néanmoins une décision partielle : il faut en tenir compte puisqu'elle exclut les autres états possibles. C'est une information comparée à l'indécision totale de la réponse «inconnu» qui exprime la disjonction de toutes les valeurs possibles ou encore une autre valeur non représentée dans l'espace d'observation (non exhaustivité du modèle descriptif).

#### 4.6.9.2 Représentant multiple

Dans le second cas, nous sommes en présence d'une "description" de classe d'objets avec les autres conséquences suivantes :

il s'agit d'un regroupement de descriptions d'objets d'une même sorte qui correspond en fait à une définition (§ 3.2.1.2), dont les valeurs d'attribut traduisent une distribution d'états (la variabilité intra-classe),

l'extension en retour est l'ensemble des objets de la sorte couverts par la "description" de la classe : elle est entachée d'une certaine généralisation,

Chaque valeur prise par les attributs de la classe est fondamentalement multiple<sup>7</sup> (sauf pour une classe de clones) et exprimée par un intervalle de variation pour les attributs numériques (ou par une disjonction de valeurs pour les attributs nominaux),

La logique des **probabilités** permet d'apprécier quantitativement la distribution de fréquence des objets au sein de la classe, ce qui demande l'observation d'échantillons représentatifs de la classe ne variant pas dans l'espace et dans le temps, ou bien une estimation subjective de l'expert plus facile à mettre en œuvre,

L'objet peut être aussi bien une classe d'individus qu'une classe de composants de cet individu. Dans le premier cas, il s'agit par exemple de la description synthétique du sous-genre *Oonema* (appartenant à *Hyalonema*) regroupant une population de spécimens. Dans le second cas, c'est la description de l'ensemble des amphidisque du corps d'un unique représentant de ce Sous-Genre.

#### 4.6.10 L'imprécision d'observation

On trouve de l'imprécision dans les observations aux deux niveaux de l'acquisition des connaissances en amont de la phase de traitement des descriptions : le premier niveau concerne l'expert et l'acquisition de l'observable, le second niveau concerne le descripteur et l'acquisition de l'observé.

---

<sup>7</sup> Les réponses multiples expriment ici une conjonction d'états simultanés : les objets sont distribués simultanément sur un sous ensemble de l'espace d'observation pour chaque attribut.

#### 4.6.10.1 Au niveau de l'observable

Nous avons déjà eu l'occasion de montrer l'attention qu'il faut apporter au choix des valeurs en liaison avec la sémantique de l'attribut (§ 4.6.5.2.4 sur le choix unique ou multiple d'un attribut).

De plus, l'expert doit prendre garde d'associer les attributs au niveau des bons objets. Considérons les deux descriptions d'objets suivantes :

- couleur de la robe du zèbre = blanc    noir
- couleur d'un feu de signalisation routière = orange    rouge

Les associations (objet = robe, attribut = couleur, valeurs = blanc, noir) et (objet = feu, attribut = couleur, valeurs = orange, rouge) sont inappropriées car elles permettent d'obtenir des conjonctions de valeurs qui indiquent la présence simultanée de deux états, ce qui est contraire au principe d'états mutuellement exclusifs. En fait, il ne peut pas y avoir d'ambiguïté sur la couleur si l'on fait porter l'attribut sur l'objet adéquat, à savoir les rayures (ou zèbrures) du zèbre et les spots du feu de signalisation.

On s'aperçoit alors qu'il existe deux sortes de rayures qui ont des couleurs bien définies, que ces rayures ne sont pas spécialisables autrement que par les noms de "rayures blanches" et "rayures noires", alors que pour les feux de signalisation routière, il existe trois sortes bien définies à la fois par la couleur et par la position (en-haut, au-milieu, en-bas).

Le modèle descriptif reflète un état de connaissance à un moment donné, il est censé alors être complet, ce qui permet de le définir comme une **monographie de l'observable** d'un domaine. La complétude est une qualité attendue du modèle alors que son imprécision est un défaut à éviter. Si le modèle est complet par rapport à un état de connaissances, cela n'empêche pas les connaissances sur les spécimens d'évoluer sous l'effet des nouvelles techniques d'observation. Ces techniques peuvent faire apparaître des erreurs de perception sur les descriptions antérieures qui ont été jugées pourtant précises à l'époque.

Même si l'on a pu constater des périodes de stabilité dans l'histoire des descriptions, les apports de nouvelles techniques sont aujourd'hui considérables : par exemple, l'apparition du microscope électronique à balayage a remis en cause la perception des observateurs utilisant le microscope optique, les techniques génétiques apportent un point de vue nouveau sur les spécimens. Le renouvellement de plus en plus accéléré des techniques met en lumière la nécessité de mise à jour du modèle descriptif et des descriptions. Cela n'ôte rien à l'obligation d'obtenir un modèle de l'observable complet et précis, à l'origine de descriptions robustes.

#### 4.6.10.2 Au niveau de l'observé

Une fois trouvée la manière de concevoir un modèle de l'observable robuste, il reste néanmoins la difficulté d'interpréter les observations elles-mêmes. L'imprécision est là aussi la source de faiblesse des descriptions. Elle dépend de plusieurs facteurs :

- 1) la nature des objets observés (représentant unique ou multiple cf. § 4.6.9),
- 2) la qualité du matériel observé (fraicheur de l'échantillon, conservation),
- 3) la qualité du mesurage des appareils,
- 4) la qualité de perception de l'observateur,
- 5) la qualité des outils de description (modèle descriptif, questionnaire),

1) En décrivant un objet qui représente en fait un ensemble, on introduit une imprécision due à la variation à l'intérieur de cette classe.

Ce que l'on décrit est en fait l'objet "moyen" sélectionné parmi tous les objets observables de l'ensemble. Il s'agit par exemple du plant de tomate (individu) appartenant à la culture et dont les symptômes sont représentatifs de la maladie ou bien encore d'un spicule d'éponge (un composant) qui est un objet modal de la classe des spicules observée sur lamelle au microscope. Il est de toute manière évident que l'on ne va pas décrire toute la population matérielle de l'ensemble. Ce que l'on cherche à décrire, c'est un représentant **prototypique** d'une classe jugée homogène (à l'intérieur de laquelle la variation n'est pas importante). Ce représentant est choisi pour faire une pseudo-description qui va constituer en fait une définition de sa classe (§ 3.2.1.2). Le représentant prototypique permet de définir les propriétés les plus fréquemment rencontrées chez les membres de la classe : c'est un objet représentatif de la classe. L'observateur décrit ainsi ce représentant. Il conserve néanmoins la liberté plus ou moins consciente d'élargir l'espace d'observation de certains caractères pour couvrir la description d'autres objets.

Inversement, lorsqu'un objet est un représentant unique d'une classe (par exemple un seul représentant d'une maladie dans la culture ou encore le collet d'un plant de tomate), le problème du choix de l'objet ne se pose pas, il s'impose.

La situation au Muséum National d'Histoire Naturelle est différente : objets et individus sont bien différenciés : les objets ne sont pas des individus, ce sont les composants d'un individu. Face aux collections de spécimens et en fonction de l'objectif de classification de ces individus, deux types de descriptions sont préconisées :

Pour chaque objet, s'il y a beaucoup de représentants qui peuvent être décrits localement, on choisit le représentant prototypique de l'ensemble, que l'on définit (ce n'est pas une véritable description). Le biologiste aura

donc tendance à synthétiser les descriptions d'objets, ce qui est normal compte tenu de l'objectif de classification des individus (et non pas des objets).

Pour les individus à classer au contraire, le biologiste va multiplier les descriptions de spécimens qu'il juge appartenir à la même classe afin d'éviter les synthèses de descriptions : chaque description de spécimen sera affectée de l'étiquette du nom de la classe. Regrouper ces spécimens en une seule description conduirait au risque de décrire sous un même nom d'Espèce des spécimens appartenant en fait à des classes différentes. Ce regroupement généralise et fait perdre de l'information. Par exemple, si on décrit un spécimen dont la taille des spicules est [5, 12] mus et un autre spécimen dont la taille des spicules est [10, 15] mus et que dans un premier temps, on considère que les deux spécimens appartiennent à la même classe, on serait amenés à regrouper deux descriptions en une seule avec la taille des spicules égale à [5, 15] mus. Ce faisant, si on découvre plus tard que les deux spécimens appartiennent en fait à des classes différentes, soit on devra refaire des descriptions conformes aux spécimens (il faut donc bien conserver les échantillons) ou alors si les spécimens ne sont plus disponibles, on aura perdu de l'information (car on a généralisé en regroupant deux descriptions).

**2)** La qualité de l'échantillon est un facteur important pour être en mesure de faire de bonnes descriptions. L'idéal est de pouvoir récupérer des renseignements sur le spécimen vivant dans son milieu d'origine, puis de le décrire en laboratoire avec les techniques appropriées. On peut préserver ainsi certaines nuances de description (les couleurs par exemple) ainsi que des informations complémentaires utiles pour pouvoir expliquer certaines observations (analyses de sol, traitements préventifs pouvant influencer l'extériorisation des symptômes en pathologie végétale). Les êtres vivants évoluent dans le temps ce qui provoque le plus souvent des phénomènes de dégradation préjudiciables pour leur description. C'est ainsi le cas des échantillons de plantes dont la fraîcheur n'est pas conservée très longtemps ou bien encore des plantes qui ne sont pas prélevées en début d'attaque et qui manifestent des symptômes secondaires (de faiblesse, saprophytes) masquant la cause primaire de la maladie [Conruyt & Piaton, 1987], [Blancard, 1988].

Ensuite, le spécimen devra être identifié, puis conservé dans une collection de manière à pouvoir s'y référer dans l'avenir. Souvent, l'échantillon est incomplet ou n'existe qu'en morceaux du fait des expérimentations qu'il a subies. Des normes de conservation dans les muséums sont établies pour bien l'entretenir et perdre le moins d'information possible. Néanmoins, pour les descriptions de maladies, les échantillons ne peuvent pas être gardés car les symptômes se dégradent avec le temps et sont plutôt fugitifs.



Enfin, pour les descriptions ayant pour origine d'autres descriptions dans des livres, on doit se contenter de ce qui existe :

des descriptions de spécimens,  
des "descriptions" synthétiques de plusieurs spécimens,  
des descriptions incomplètes, etc..

L'expert est à ce moment obligé de réinterpréter des descriptions plus ou moins anciennes. S'il n'a plus accès au spécimen d'origine alors que des moyens nouveaux d'investigation sont possibles (par exemple une expérimentation sur le génome qui ne pouvait être prise en compte lors de la première description), la description s'arrête là. Dans notre application sur les *Hyalonema*, l'expert a surtout travaillé à partir de descriptions et de dessins dans des livres anciens, les spécimens n'étant plus disponibles. Toutefois, les parties du spécimen qui ont servi à établir les classifications sont d'ordre microscopique : ce sont les spicules du squelette de ces éponges siliceuses. L'expert M. Levi dispose encore de certaines de ces préparations.

**3)** La qualité des appareils de mesure joue sur la précision de chaque description. Par exemple, en pathologie végétale, si on demande à l'agriculteur de faire une coupe transversale tout le long de la tige d'un plant de tomate pour regarder la couleur des vaisseaux, et qu'il ne dispose pas de couteau, soit il ne répond pas à la question, soit il fait une réponse approximative après avoir ouvert manuellement avec l'ongle la tige à différents niveaux. Ce bruit dans la phase de description n'est pas toujours contrôlé à l'arrivée (au moment du traitement) : il se peut que l'observation comme quoi les vaisseaux sont bruns est vérifiée au bas de la tige mais pas en haut (il ne pouvait pas le faire sans couteau).

En systématique, le problème est différent : les biologistes travaillent dans un environnement propice aux descriptions précises avec des appareils plus ou moins sophistiqués. Par rapport à l'objectif de classification, les descriptions effectuées à une date donnée sont supposées complètes, c'est-à-dire avec le niveau de précision suffisant pour l'objectif fixé. Ce niveau de précision des descriptions dépend néanmoins de l'évolution des appareils de mesure. Ces derniers permettent d'affiner certains détails qui peuvent se révéler en contradiction avec les observations précédentes sur un autre matériel de mesure. Par exemple, en observant la protoconque (coquille embryonnaire) d'un mollusque, le biologiste aperçoit des ponctuations au microscope optique lui faisant penser à des trous d'épingle et il décrit la protoconque comme ponctuée. S'il observe cette coquille au microscope électronique à balayage, ces "trous" apparaissent alors comme des bosses et la description se transforme alors en protoconque pustulée. La description de la protoconque est devenue plus juste en utilisant un matériel de mesure plus précis : l'aspect concave des ponctuations

(comme des trous) a été remplacé par un aspect convexe dû à des pustules sur la coquille.

4) La qualité des observations est un leitmotiv dans la bouche des experts pour pouvoir effectuer des déterminations correctes. Cela dépend de la perception du descripteur. Par exemple en pathologie végétale, des **descripteurs naïfs par rapport à l'observation** auront tendance à décrire toute anomalie apparente sur la plante comme un symptôme potentiel d'une maladie. Ces descripteurs n'ont pas la capacité de sélectionner les symptômes pertinents par rapport aux maladies répertoriées par l'expert : ainsi en est-il des taches de cuivre bleutées qui résultent de l'application d'un traitement à la bouillie bordelaise contre l'oïdium ou encore du jaunissement dû à la sénescence des feuilles du bas ou à l'ombrage du feuillage : ce sont des descriptions de faux symptômes !

Inversement, sur une même plante, l'expert a tendance à ne voir que les symptômes caractéristiques des maladies les plus importantes et à laisser de côté tous les symptômes annexes sans les désigner : il fait un tri des symptômes de manière inconsciente.

Cet exemple pris en pathologie végétale montre les écarts importants qui existent dans les interprétations de l'observation entre différents utilisateurs. La naïveté des observations de l'utilisateur est un critère à prendre en compte pour évaluer la robustesse des systèmes de détermination. Dans la mesure où l'on souhaite que le système soit utilisé dans les conditions de la réalité, c'est à l'expert et au cognicien de s'adapter aux utilisateurs et non l'inverse.

5) Cette adaptation passe par la confection d'un modèle descriptif et d'un questionnaire permettant une aide à l'observation. Le modèle descriptif constitue la trame du questionnaire calqué sur son architecture afin de fabriquer un véritable guide d'observation comme nous le verrons au chapitre 6.