V FORMALISATION DE LA NOTION D'OBJET EN BIOLOGIE

Après avoir mis en lumière l'importance des descriptions de spécimens, nous proposons maintenant une formalisation mathématique des objets biologiques[1] à analyser en suivant le cheminement d'élaboration du modèle descriptif. Nous ne formalisons pas ici les traitements à effectuer sur ces objets (induction et raisonnement par cas) qui seront étudiés au chapitre 7.

Nous présentons d'abord les assertions composites permettant le processus de décomposition de la description de l'entité du domaine. Puis les hordes composites sont utilisées pour définir les objets multi-instanciés. La description de l'individu est enfin exprimée à l'aide d'un objet de synthèse comprenant des objets "assertion" et "horde" composites décrivant ses différentes parties. Des propriétés sont ajoutées à l'ensemble de ces objets par des règles sous forme de contraintes. De plus, une connaissance supplémentaire peut être introduite sous la forme d'un ordre hiérarchique sur ces objets. Ce formalisme reprend celui des objets symboliques booléens introduits par [Diday, 1987] et est adapté dans le cadre de l'apprentissage d'objets biologiques sur des individus complexes.

5.1 Les assertions composites

5.1.1 Rappel sur les assertions (symboliques)

Soient $\Omega = \{w_1,\ldots,w_n\}$, l'ensemble des entités concrètes ou individus observés,

$\Pi\ 2\ \Omega$, l'ensemble de tous les individus observables ou abstraits,

$Y = \{y_1,\ldots,y_p\}$, l'ensemble des variables observées du domaine,

$\Lambda\ 2\ Y$, l'ensemble de toutes les variables observables.

Pour chaque variable observée y_i Y, nous pouvons associer un espace d'observation O_i qui est l'ensemble des valeurs possibles de y_i (ce sont des valeurs observables abstraites). On définit alors y_i comme une application de Π dans O_i avec O_i $\{O_1,\ldots,O_p\}$ $\{?\}$.

[1] Comme nous l'avons déjà dit au § 3.6, l'objet biologique peut prendre la signification d'un individu ou d'un composant d'un individu selon le point de vue.

Enfin, soit V_i O_i, l'ensemble des valeurs observées de y_i (valeurs concrètes) avec V_i $\{V_1,...,V_q\}$.

L'assertion symbolique $a_s = {}_i [y_i = V_i]$ exprime que “La variable y_i prend des valeurs dans V_i”. Elle est définie par l'application a_s :

$$a_s : \Omega \rightarrow \{vrai, faux\} \;/\; a_s(w) = vrai \;\; ssi \quad i = 1,...,p \;\; \text{on a } y_i (w) \quad V_i$$

L'extension de a_s notée $| a_s |_\Omega$ est l'ensemble des individus w Ω pour lesquels $a_s(w)$ = vrai.

<u>Exemple</u> :

Soient $\Omega = \{w_1,...,w_n\}$, l'ensemble des descriptions observées de têtes de mammifères,
Λ, l'ensemble de toutes les variables observables sur une tête de mammifère,

Soient $Y = \{ y_1 = [taille(tête)], y_2 = [longueur(nez)], y_3 = [couleur(yeux)]\}$, l’ensemble des variables observées,
O_1 l'ensemble des tailles observables en cm pour la tête,
O_2 l'ensemble des longueurs possibles en cm pour le nez,
O_3 l'ensemble des couleurs possibles pour les yeux.

En considérant l'assertion a_s suivante :

[[taille(tête)] = [100 , 200] = V_1]
[[longueur(nez)] = [60 , 200] = V_2]
[[couleur(yeux)] = {brun, marron} = V_3],

l'extension de a_s notée $| a_s |_\Omega$ est l'ensemble des descriptions des têtes de mammifères qui vérifient l'assertion :

$$| a_s |_\Omega = \{w \quad \Omega \;/\; a_s(w) = vrai \Leftrightarrow y_1(w) \quad V_1 \quad y_2(w) \quad V_2 \quad y_3(w) \quad V_3\}$$

<u>Remarque</u> : Dans cette définition de l'assertion, 'a_s' est à la fois la **notation** (une conjonction d'évènements) et une **fonction** indiquant la méthode de calcul de son extension. Pour nos applications biologiques, nous souhaitons bien séparer les deux afin de <u>représenter</u> d'une part <u>les connaissances de l'expert</u> c'est-à-dire les **descriptions d'objets** (l'ensemble de départ de la fonction) et d'autre part <u>les traitements à effectuer sur ces objets</u> (induction et raisonnement par cas) qui font l’objet du chapitre 7.

5.1.2 Proposition : les objets assertions

Un objet assertion $a = \ _i \ [y_i = V_i]$ est défini par l'application a :
$a : \Omega \rightarrow \{vrai, faux\}$ / $a(w) = vrai$ ssi $i = 1,...,p$ on a $y_i (w) \quad O_i$,

c'est-à-dire que les objets ont des valeurs observées comprises dans le domaine de définition prédéfini des valeurs observables du modèle descriptif.

Le calcul de l'extension de a n'a pas grand intérêt pour nous puisqu'il est l'ensemble des individus (les clônes) qui ont la même description a.

Dans cette définition des objets assertions, l'utilisateur a la possibilité d'indiquer qu'il ne connaît pas la valeur de $y_i (w)$. A cet instant, la réponse “?” signifie l'indécision totale, c'est-à-dire la disjonction de toutes les valeurs possibles de la variable y_i.

On peut aussi définir un objet assertion $a = \ _i \ [X_i = V_i]$ comme une application
$a : \Omega \rightarrow \{vrai, faux\}$ / $a(x) = vrai$ ssi $i = 1,\dots,p$ on a $x_i \quad V_i$, avec la propriété suivante dans le cas ou Y est une bijection : $a_s = a \circ y$

<u>preuve</u> : Y est bijective : $x_i \quad O_i$, $w \quad \Omega$ / $y_i (w) = x_i$
$a_s(w) = \ _i (y_i (w) \quad V_i) = \ _i (x_i \quad V_i) = a (x) = a (y (w)) = (a \circ y) (w)$

5.1.3 Définition des assertions composites

➢ Soit l'espace des parties d'individus observables $\Pi = \Pi_1 \times \Pi_2 \times \dots \times \Pi_{m'}$
Π_i est l'ensemble de toutes les parties élémentaires observables d'une entité w.

Soit l'espace de parties observées d'individus : $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_m$
m étant le nombre de parties de w, on définit $\Omega_i = \{p_i\}$, l'ensemble singleton d'une partie élémentaire observée de l'entité w.
On a $\Omega_i \quad \Pi_i$

<u>Exemple</u> : $\Omega = \{têtes\}$, $\Omega_1 = \{nez\}$, $\Omega_2 = \{yeux\}$

➢ Soit l'espace des variables observables du domaine $\Lambda = \Lambda_1 \times \Lambda_2 \times \dots \times \Lambda_{p'}$
p' étant le nombre d'ensembles de variables des parties de w, Λ_i est l'ensemble de toutes les variables observables d'une partie p_i de Ω_i.

Soit l'espace des variables observées du domaine $Y = Y_1 \times Y_2 \times \dots \times Y_p$

Sur chaque partie de l'entité, on peut définir l'ensemble des variables observées $Y_i = \{y_1^i, .., y_j^i, .., y_q^i\}$, q étant le nombre de variables d'une partie de w, y_j^i représentant la $j^{ème}$ variable de la partie i de w.
On a $Y_i \quad \Lambda_i$

Soit $Q_i = \{q_1^i, .., q_j^i, .., q_q^i\}$, l'ensemble des qualités ou caractères observés d'une partie p_i de Ω_i.

Soit $N_i = \{n_i\}$, l'ensemble singleton comportant le nom de la partie p_i de Ω_i, on a $Y_i = Q_i \circ N_i$.

Exemple : Q_1 = {taille}, N_1 = {tête}, Y_1 = {[taille(tête)]}
Q_2 = {couleur}, N_2 = {yeux}, Y_2 = {[couleur(yeux)]}

➢ Soit l'espace d'observation du domaine $O = O_1 \times O_2 \times \ldots \times O_n$
n est le nombre d'ensembles d'observation des parties de w, O_i est l'ensemble d'observation d'une partie p_i de Ω_i.
Y_i est une fonction de $\Pi_i \rightarrow O_i$.

On peut définir $O_j^i \quad \{O_1^i, \ldots, O_r^i\}$ l'ensemble d'observation où la variable y_j^i prend ses valeurs, r étant le nombre d'ensembles de valeurs observables d'une partie p_i de Ω_i

Définition :

Soit A l'ensemble des assertions du domaine, une assertion composite $a_i \quad$ A est une fonction $\Omega_i \rightarrow$ {vrai, faux} / $a_i(w)$ = vrai ssi

$$i = 1,\ldots,p \quad , \quad j = 1,\ldots,q \quad \text{on a } y_j^i(w) \quad O_j^i$$

avec la propriété suivante :
a_i est définie par la conjonction d'évènements $[y_j^i = V_j^i]$ dont une valeur au moins de y_j^i est une assertion composite a_j définie sur Ω_j :

$$y_j^i \text{ est une fonction de } \Pi_i \rightarrow O_j$$

$$a_i = \wedge_i \; [y_j^i = V_j^i] = \wedge_i \; [[q_j^i(n_i)] = V_j^i] \quad avec \; \exists\, v \in V_j^j \; / \; v \in A_j$$

On peut illustrer la définition précédente par le schéma de la figure 5.1 ou h et i sont des parties de w et sont représentées par des descriptions sous forme de vecteurs des variables y_j^h et y_j^i :

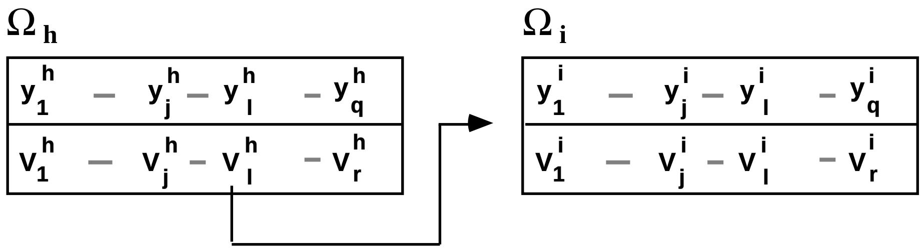

Fig. 5.1 : Illustration d'une assertion composite a_h

5.1.4 Exemples

Une assertion composite “mammifère” peut s'écrire :

a_x = [[sexe(mammifère)] = V_1^x]] [[age(mammifère)] = V_2^x]
[[poids(mammifère)] = V_3^x]
[[subparts(mammifère)] = {tête, tronc, membres}]

Les valeurs de la variable y_4^x = [subparts(mammifère)] sont aussi des assertions composites :

a_y = tête = [[forme(tête)] = V_1^y]] [[taille(tête)] = V_2^y]
[[subparts(tête)] = {oreilles, yeux, bouche, nez}]

De manière analogue pour notre application sur les éponges marines, on définit une assertion composite “corps” de l'éponge :

a_6 = [[forme(corps)] = V_1^6]] [[taille(corps)] = V_2^6]
[[consistance(corps)] = V_3^6] [[couleur(corps)] = V_4^6]
[[subparts(corps)] = {macro-constituants, micro-elements}]

Les valeurs de la variable y_5^6 = [subparts(corps)] sont des assertions composites :

a_8 = [[forme(face-exhalante)] = V_1^8]]
[[subparts(face-exhalante)] = {orifices, cône-central, membrane-criblée}]

(voir a_6 et a_8 sur le schéma du modèle descriptif de la figure 5.2 : on remarquera la correspondance entre un objet physique “face exhalante” et l'objet contextuel “macro constituants”. Ce dernier est un label de description, une étiquette indiquant qu'il faut observer à une échelle macroscopique alors que les autres objets du corps de l'éponge s'observent dans un contexte microscopique) :

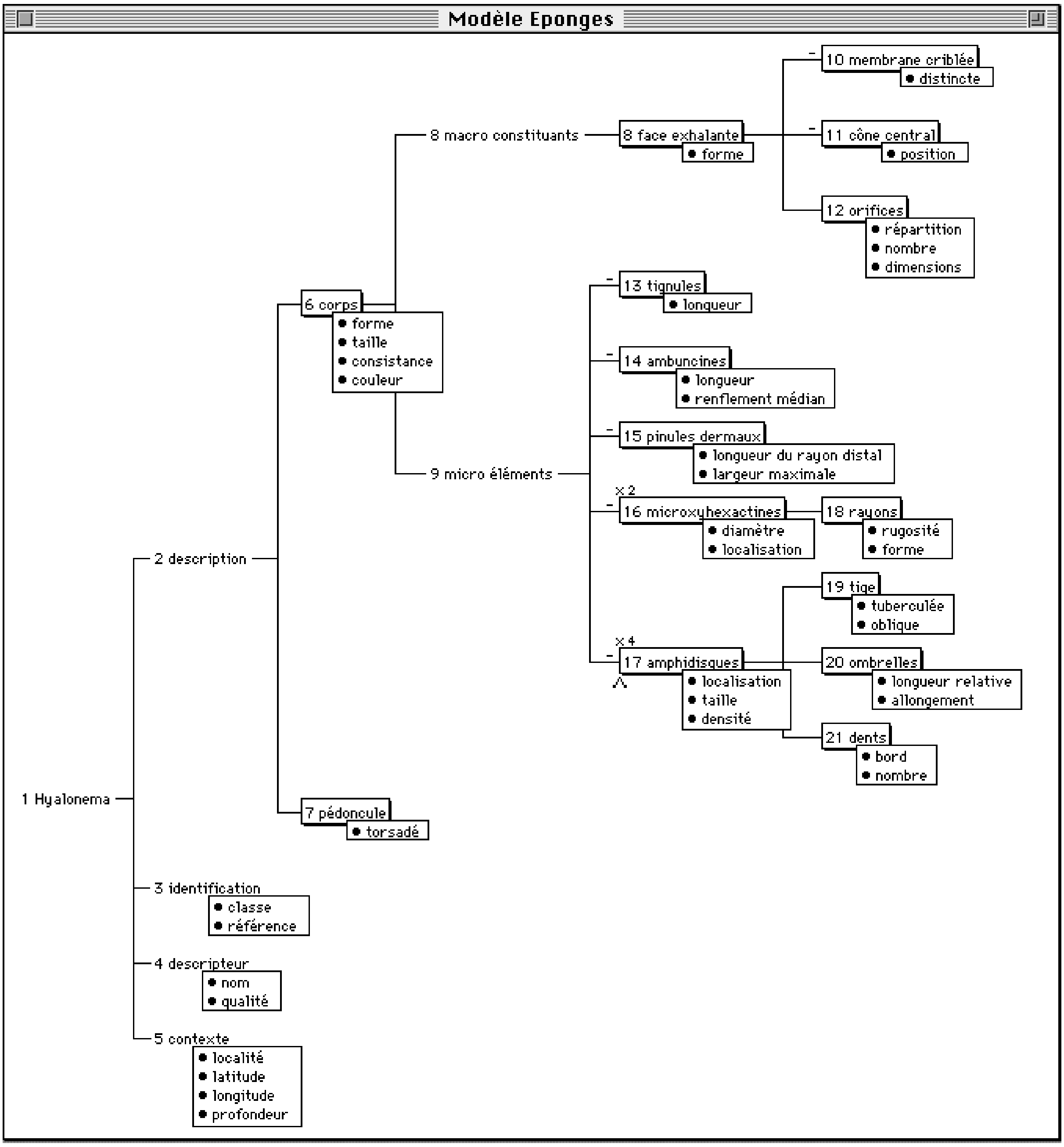

Fig. 5.2 : Schéma de la structure du modèle descriptif

5.2 Les hordes composites

5.2.1 Rappel sur les hordes (symboliques)

La horde symbolique $h_s = {}_{i}\ [y_i(u_i) = V_i]$ exprime que “La variable y_i de l'individu i prend des valeurs dans V_i ”. Elle est définie par l'application h_s.

$h_s : \Omega^n \rightarrow \{vrai, faux\}$ / $u = (u_1,\ldots,u_n) \quad \Omega^n$, $h_s(u) = vrai$ ssi
$i = 1,\ldots,p$ on a $y_i (u_i) \quad V_i$.

Dans la notation précédente, les u_i représentent différentes instances d'individus non nécessairement identiques.

L'extension de h_s notée $| h_s |_\Omega$ est l'ensemble des individus $u \quad \Omega^n$ pour lesquels $h_s(u)$ = vrai. Lorsque tous les u_i sont identiques, l'objet horde se réduit à un objet assertion (on a H A, l'ensemble des hordes du domaine).

5.2.2 proposition : les objets hordes

Un objet horde est défini par la fonction h.
$h : \Omega^n \rightarrow \{vrai, faux\}$ / h(u) = vrai ssi $i = 1,\ldots,p$ on a $y_i (u_i) \quad O_i$.

<u>Exemple</u> :

Soit $\Omega^{32} = \{w_1,\ldots,w_{32}\}$, l'ensemble des descriptions observées de dents chez un humain.

Soit Λ l'ensemble de toutes les variables observables sur une dent d'humain,
Soit Y = { y_1 = [position(dents)], y_2 = [face-occlusale(dents)]}, l'ensemble des variables observées.

Soit O_1 l'ensemble de toutes les positions observables pour y_1.
Soit O_2 l'ensemble des aspects de la face occlusale possibles pour y_2. Il s'agit de la partie mordante de la dent.

Soit l'objet horde décrivant les quatre sortes de dents que l'on rencontre chez l'humain :

h = [[position(dents)] (u_1) = en-arriere = V_1] (molaires)
[[face-occlusale(dents)] (u_1) = 4-pointes = V'_1]

[[position(dents)] (u_2) = au-milieu = V_2] (prémolaires)
[[face-occlusale(dents)] (u_2) = 2-pointes = V'_2]

[[position(dents)] (u_3) = devant = V_3] (canines)
[[face-occlusale(dents)] (u_3) = 1-pointe = V'_3]

[[position(dents)] (u_4) = devant = V_4] (incisives)
[[face-occlusale(dents)] (u_4) = arrête = V'_4]

Dans cet exemple, on caractérise sans les nommer les quatre sortes de dents chez les humains (les Incisives, les Canines, les Prémolaires et les Molaires). On aurait pu très bien décrire chaque classe de dent par une assertion (respectivement chaque u_i) à condition de la nommer. Utiliser les hordes se justifie lorsque l'on désire constituer une base de cas portant sur différentes sortes d'objets d'un même type (les dents) sans que l'utilisateur soit contraint d'en connaître le nom pour les spécialiser.

Par la suite, à partir de ces cas, on pourra effectuer une classification locale (une généralisation) de ces différentes sortes d'objets. Cela permet d'extraire pour chaque sorte une description de Classe avec ses critères représentatifs, ainsi que l'intervalle de variation des valeurs.

L'appartenance d'une nouvelle dent à l'une des Classes peut alors être testée par description puis comparaison avec la représentation en intension des Classes.

5.2.3 Définition des hordes composites

➢ Soit l'espace de description $\Omega = [\Omega_1]^{n1} \times [\Omega_2]^{n2} \times \ldots \times [\Omega_m]^{nm}$
m étant le nombre de parties d'une entité w, on définit l'ensemble des parties élémentaires observées de w : $[\Omega_i]^{ni} = \{p^i_1, \ldots, p^i_k, \ldots, p^i_n\}$, n étant le nombre de parties élémentaires instanciées observées d'une partie p_i de Ω_i , et k étant la $k^{ème}$ instance observée, p^i_k est donc la $k^{ème}$ instance de la partie i de l'entité w.

➢ Soient Π_i , Λ_i , Y_i , O_i les ensembles définis au paragraphe 5.1.1.

<u>Définition</u> :

Soit H l'ensemble des hordes du domaine, une horde composite h_i H est une fonction $[\Omega_i]^{ni} \rightarrow$ {vrai, faux} / $h_i(u_j^i)$ = vrai ssi

$$i = 1,\ldots,p \qquad j = 1,\ldots,q \text{ on a } y_j^i(u_j^i) \quad O_j^i$$

avec la propriété suivante :
h_i est définie par la conjonction d'évènements $[y_j^i(u_k^i) = V_k^{ij}]$ dont les k objets instanciés u_k^i d'une partie élémentaire p_k^i de w ne sont pas nécessairement identiques et dont une valeur au moins de y_j^i est une assertion a_i ou une horde h_i définie sur Ω_j :

$$y_j^i \textit{ est une fonction de } [\Pi_i]^{ni} \rightarrow O_j$$
$$h_i = \wedge_i \; [y_j^i(u_k^i) = V_k^{ij}] \quad \textit{avec } \exists\, v \in V_k^{ij} \;/\; v \in A_j \textit{ ou } v \in H_j$$

De même que pour les assertions composites, on peut illustrer la définition précédente par le schéma de la figure 5.3 ou i est une partie de w représentée par une description sous forme de tableau des variables y_j^i avec chaque ligne correspondant à une instance de la partie i, et j est par exemple une autre partie représentée par une assertion composite :

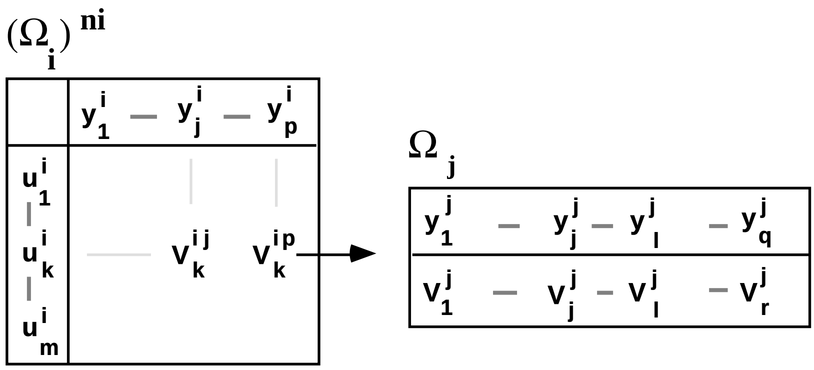

Fig. 5.3 : Illustration d'une horde composite h_i

5.2.4 *Exemples*

Une horde composite “membres” de mammifère s'écrit :

h_x = $[[\text{longueur(membres)}]\ (u_1^x) = V_1^{x,1}]]$

$[[\text{position(membres)}]\ (u_1^x) = V_1^{x,2} = \{\text{avant}\}]$

$[[\text{subparts(membres)}]\ (u_1^x) = \{\text{pieds}(u_1^x), \text{genoux}(u_1^x), \text{jambes}(u_1^x)\}]$

$[[\text{longueur(membres)}]\ (u_2^x) = V_1^{x,2}]]$

$[[\text{position(membres)}]\ (u_2^x) = \{\text{arrière}\}]$

$[[\text{subparts(membres)}]\ (u_2^x) = \{\text{pieds}(u_2^x), \text{genoux}(u_2^x), \text{jambes}(u_2^x)\}]$

Les valeurs $V_1^{x,3}$, $V_2^{x,3}$, $V_3^{x,3}$ de la variables y_3^x = [subparts(membres)] sont ici des assertions :

a_y = pieds(u_1^x) = $[[\text{forme(pieds)}]\ (u_1^x) = V_1^{y,1}]]$ $[[\text{taille(pieds)}]\ (u_1^x) = V_2^{y,2}]$

Dans cet exemple, l'utilisateur qui ne sait pas distinguer les membres postérieurs des membres antérieurs d'un mammifère peut néanmoins décrire deux sortes de membres sans avoir la connaissance de spécialisation nécessaire. Il indiquera seulement la position des membres : à l'avant ou bien à l'arrière.

Une horde composite "microxyhexactines" de l'éponge s'écrit :

h_{16} = $[[\text{diametre(microxyhexactines)}]\ (u_1^{16}) = V_1^{16,1}]$

$[[\text{localisation(microxyhexactines)}]\ (u_1^{16}) = V_1^{16,2}]$

$[[\text{subparts(microxyhexactines)}]\ (u_1^{16}) = \{\text{rayon}(u_1^{16})\}]$

$[[\text{diametre(microxyhexactines)}]\ (u_2^{16}) = V_2^{16,1}]$

$[[\text{localisation(microxyhexactines)}]\ (u_2^{16}) = V_2^{16,2}]$

$[[\text{subparts(microxyhexactines)}]\ (u_2^{16}) = \{\text{rayon}(u_2^{16})\}]$

5.3 Les objets de synthèse

5.3.1 Rappel sur les objets de synthèse (symboliques)

➢ Soient Ω_1 ,…, Ω_m m ensembles de parties élémentaires de w caractérisées respectivement par m ensembles de variables $Y_1, \ldots, Y_m$.

H_i est l'ensemble de toutes les hordes symboliques définies sur $[\Omega_i]^{ni}$.

Un objet de synthèse symbolique s_s est la conjonction des m objets horde $h_1,\dots, h_m$ définis sur $H_1,\dots, H_m$: $s_s = h_1 \quad \dots \quad h_m$ avec $h_i \quad H_i$

➢ Soit l'ensemble U des objets instanciés sur $[\Omega_m]^{nm}$: $U = (U_1, \dots, U_m)$

Soient $U_i = (u_1^{ij}, \dots, u_k^{ij}, \dots, u_n^{ij}) \quad [\Omega_i]^{ni}$, l'ensemble des instances de l'objet i,
$Y_i \quad (Y_1,\dots, Y_m)$, l'ensemble des variables observables de l'objet i,
$V_i \quad (V_1,\dots, V_m)$, l'ensemble des variables observées de l'objet i,
$s(U) = \ _i\ [Y_i(U_i) = V_i]$.

L'objet de synthèse symbolique s_s est défini par l'application s_s :

$s_s : \Omega = [\Omega_1]^{n1} \times [\Omega_2]^{n2} \times \dots \times [\Omega_m]^{nm} \rightarrow \{vrai, faux\}$ /
$s_s(U) = vrai$ ssi $i = 1,\dots,m$
$j = 1,\dots,nm$
$k = 1,\dots,n$ on a $y_k^{ij}(u_k^{ij}) \quad V_k^{ij}$.

L'extension est $|s_s|_\Omega = \{w = (w_1^1, \dots, w_1^i, \dots, w_k^i, \dots, w_{nk}^p) \quad \Omega \ / \ y_k^{ij}(u_k^{ij}) \quad V_k^{ij}$ }.

5.3.2 *Proposition : les objets de synthèse*

L'objet de synthèse s, qui est la description d'un individu ou d'un spécimen (une unité), est défini par l'application s :

$s : \Omega = [\Omega_1]^{n1} \times [\Omega_2]^{n2} \times \dots \times [\Omega_m]^{nm} \rightarrow \{vrai, faux\}$ /
$s(U) = vrai$ ssi $i = 1,\dots,m$
$j = 1,\dots,nm$
$k = 1,\dots,n$ on a $y_k^{ij}(u_k^{ij}) \quad O_j^i$.

5.3.3 *Exemple*

Une éponge dans notre application est une entité individuelle complexe représentée par l'objet de synthèse s qui est la conjonction de 2 objets hordes composites (les amphidisques et les microxyhexactines) et de 20 objets assertions composites. L'objet horde composite "amphidisques" est instancié au maximum 4 fois alors que l'objet horde composite "microxyhexactines" est instancié 2 fois :

$s : \Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_{15} \times [\Omega_{16}^1 \times \Omega_{16}^2] \times [\Omega_{17}^1 \times \Omega_{17}^2 \times \Omega_{17}^3 \times \Omega_{17}^4] \times \Omega_{18} \times \dots \quad \dots \times \Omega_{22} \rightarrow \{vrai, faux\}$

5.4 Les objets munis de méthodes et de propriétés

Afin d'exprimer les liens entre les objets et les variables, on peut être amené à écrire des règles au sein d'un objet assertion ou horde composites. Il suffit pour cela d'ajouter par conjonction des évènements élémentaires définissant par exemple une contrainte sur l'existence d'une variable ou d'un objet.

Les règles peuvent s'exprimer sous la forme (si $y_i = V_i$ alors $y_j = V_j$) s'il y a dépendance entre un sous-ensemble de valeurs $V_i \quad O_i$ et un sous-ensemble de valeurs $V_j \quad O_j$:

Par exemple dans l'assertion "culture" attachée à une parcelle de plantes maraîchères, une règle sur le mode de culture permet de restreindre l'intervalle des valeurs possibles à une valeur pour le type de culture :

a = [[stade(culture)] = V_1] [si [[mode(culture)] = plein-champ] alors
[[type(culture)] = en-sol]].

Ces règles peuvent aussi permettre de restreindre l'espace des variables observables $\Lambda_i^<$ ainsi que l'espace des objets observables $\Pi_i^<$ d'un objet i.

Soit $\Pi_i^< = \{ \Pi_k , k \quad \{1,\dots,m\}$ / "l'objet k est une sous-partie de l'objet i"}, l'ensemble des objets observables sous-parties de l'objet i.

Soit $\Lambda_i^< = \{ \Lambda_k , k \quad \{1,\dots,m\}$ / "l'objet k est une sous-partie de l'objet i"}, l'ensemble des ensembles de variables observables des objets observables sous-parties de l'objet i.

5.4.1 Cas des variables :

S'il y a dépendance entre un sous-ensemble de valeurs $V_j \quad O_j$ et un sous-ensemble de variables $\Lambda_k \quad \Lambda_j$, une règle sera du type :
si $y_i = V_i$ alors $\Lambda_k \quad \Lambda_i^< \; / \quad \Lambda_j = \Lambda_j \setminus \Lambda_k$ (\ est le symbole d'exception)

Exemple :
Prenons l'assertion "orifices" de la face exhalante (partie n° 9 du schéma de la figure 5.2) :

a_{e12} = [[nombre(orifices)] = V_1^{12}]
[si [[nombre(orifices)] {unique, ?}] alors
[$\Lambda_{12} = \Lambda_{12} \setminus$ {[dimensions(orifices)] , [repartition(orifices)]}]]

Cette assertion exprime le fait que si on a qu'un seul orifice à la face exhalante, alors on ne doit pas s'intéresser à la description des variables observables dimension et répartition des orifices (la 1ère parce qu'elle n'est plus pertinente, la seconde parce qu'elle n'est pas observable dans ce contexte).

5.4.2 *Cas des objets :*

S'il y a dépendance entre un sous-ensemble de valeurs V_j O_j et un sous-ensemble d'objets Π_k Π , on peut avoir une règle du type :

si $y_i = V_i$ alors Π_k $\Pi_i^{<}$ / $\Pi = \Pi \setminus \Pi_k$

Exemple :

Prenons la horde composite “amphidisques” :

h_{17} = **[**si [[type(amphidisques)] (u_i^{17}) = micramphidisque] alors

[$\Lambda_{20} = \Lambda_{20}$ \ {[allongement(ombrelles)] (u_i^{17})}]

[$\Pi = \Pi$ \ {dents(u_i^{17})}]]]**]**

[[localisation(amphidisques)] (u_1^{17}) = $V_1^{17,1}$]

[[taille(amphidisques)] (u_1^{17}) = $V_2^{17,1}$]

[subpart(amphidisques)](u_1^{17}) = {tige(u_1^{17}),ombrelles(u_1^{17}),dents(u_1^{17})}]

[[localisation(amphidisques)] (u_2^{17}) = $V_1^{17,2}$]

[[taille(amphidisques)] (u_2^{17}) = $V_2^{17,2}$]

[…]

Dans cette horde, la règle indique que si le type de l'amphidisque est un micramphidisque (quelle que soit l’instance d’amphidisque concernée), alors on ne doit pas s'intéresser à la description de la variable allongement des ombrelles (partie n° 20 de l'éponge) de l'amphidisque (partie n° 17), mais aussi que les dents (partie n° 21) de ces micramphidisques ne sont plus observables dans ce contexte.

Plus généralement, on peut énoncer des règles implicites très fortes s'appliquant à chaque objet de la description dont la valeur attachée à la variable y_i = “sous-partie-de(objet $_i$)” est un objet non présent (¬) ou inconnu (?) :

v V_i , si $y_i = \neg v$ ou $y_i = ?v$ alors

Π_k $\Pi_i^{<}$, Λ_k $\Lambda_i^{<}$

$\Pi = \Pi \setminus \Pi_k$ et $\Lambda_j = \Lambda_j \setminus \Lambda_k$

Exemples :

1) [si [[sous-partie-de(champignon$_1$)] = ?chapeau] alors
[$\Lambda_1 = \Lambda_1$ \ {[couleur(chapeau)] , [forme(chapeau)]}]]

2) [si [[subparts(micro-elements)] = ¬microxyhexactines(u_i)] alors
[$\Lambda_{16} = \Lambda_{16}$ \ {[diametre(microxyhexactines)](u_i) ,
[localisation(microxyhexactines)](u_i)}]
[$\Lambda_{15} = \Lambda_{15}$ \ {[rugosite(rayons)](u_i) , [forme(rayons)](u_i)}]
[Π = Π{rayons(u_i)}]]

La règle 2 signifie que lorsqu'il n'y a pas d'objet microxyhexactines dans la description (qu'il soit multi-instancié par u_i ou non), alors une telle description constituerait une incohérence : il est inutile de décrire le diamètre et la localisation de cet objet ainsi que de renseigner l'objet rayons qui n'existe pas dans ce contexte.

5.5 Les objets classifiés

On peut définir un ordre partiel sur des objets p_i et p'_i en fonction de leurs extensions $|p_i|$ et $|p'_i|$ permettant de formaliser les notions d'héritage et de généralisation [Brito, 1991] :

p_i , p'_i Π_i , on dit que $p_i \leq p'_i$ ssi $|p_i|$ $|p'_i|$

On dira que p_i hérite de p'_i si $p_i \leq p'_i$ et que p'_i est plus général que p_i

Exemple :

En considérant l'ensemble des baleines et l'ensemble des mammifères, la partie p_i = "nageoires latérales" des baleines est une spécialisation de la partie p'_i = "membres-antérieurs" des mammifères.

Les nageoires héritent des propriétés des membres antérieurs tout en ayant des caractéristiques propres (comme celle par exemple de ne pas avoir de sous-parties bras et mains).

5.6 Les exemples d'apprentissage

Un cas ou un exemple d'apprentissage dans notre application est un objet de synthèse. C’est une conjonction $w \quad \Omega$ d'objets instanciés sur la base du modèle descriptif pour lesquels s(w) = vrai :

A titre d'exemple, nous présentons le cas w_{16} pour lequel s(w_{16}) = vrai :

w_{16} = {[[subparts(eponge)] = {description,identification,contexte}]
[[subparts(description)] = corps]

[[classe(identification)] = Prionema]
[[reference(identification)] = “Spinosum Lendenfeld 1915"]

[[localite(contexte)] = “Pacifique est"]
[[latitude(contexte)] = “0°4'N"]
[[longitude(contexte)] = “117°15'W"]
[[profondeur(contexte)] = “4243m"]

[[subparts(corps)] = micro-elements]
[[taille(corps)] = 47 mm]
[[forme(corps)] = aplatie(en-galette)]

[[subparts(micro-elements)] = {amphidisques(v_1), amphidisques(v_2),
amphidisques(v_3), microxyhexactines(u_1),
¬microxyhexactines(u_2), pinules-dermaux}]

[[type(amphidisques)](v_1) = macramphidisque]
[[localisation(amphidisques)](v_1) = disperce-partout]
[[taille(amphidisques)](v_1) = [180 , 299] mus]
[[subparts(amphidisques)](v_1) = {tige(v_1), ombrelles(v_1), dents(v_1)}]

[[type(amphidisques)](v_2) = mesamphidisque]
[[taille(amphidisques)](v_2) = [45 , 127] mus]
[[subparts(amphidisques)](v_2) = {tige(v_2), ombrelles(v_2), dents(v_2)}]

[[type(amphidisques)](v_3) = micramphidisque]
[[localisation(amphidisques)](v_3) = disperse-partout]
[[taille(amphidisques)](v_3) = [13 , 29] mus]

[[subparts(microxyhexactines)] (u_1) = rayons]
[[diametre(microxyhexactines)] (u_1) = [108 , 179] mus]
[[localisation(microxyhexactines)] (u_1) = en-paquets]

[[longueur-du-rayon-distal(pinules-dermaux)] = [100 154] mus] [[largeur-maximale(pinules-dermaux)] = [10 , 17] mus]

[[tuberculee(tige)](v_1) = oui]
[[allongement(ombrelles)](v_1) = [1.00 , 1.63]]
[[longueur-relative(ombrelles)](v_1) = [0.21 , 0.38]]
[[bord(dents)](v_1) = lisse]

[[tuberculee(tige)](v_2) = oui]
[[allongement(ombrelles)](v_2) = [1.08 , 1.87]]
[[longueur-relative(ombrelles)](v_2) = [0.28 , 0.44]]
[[bord(dents)](v_2) = lisse]

[[forme(rayons)] = droits]   [[rugosite(rayons)] = epineux] }

La description de tous les objets de synthèse de *Hyalonema* constitue l'ensemble $\Omega = \{w_1,...,w_n\}$ des entités concrètes ou individus observés.

On appellera aussi cet ensemble la *base de cas ou base d'exemples* du domaine. Dans cette description d'éponge, on remarque que toute l'information est présente, alors que la non-information est absente : l'inconnu comme valeur d'une variable n'est pas une information, pas plus que la présence inconnue d'un objet. On met en évidence l'existence d'une variable qualitative particulière nommée “classe(identification)” qui sera la variable à expliquer ou le concept à apprendre dans le système d'apprentissage. Cette variable possède 12 valeurs ou modalités dans la base de cas.

Par définition, un exemple où ne figure pas cette variable à expliquer sera appelée une <u>observation</u> du domaine. Enfin, on peut aussi noter la variable [forme(corps)] qui prend comme état une hiérarchie de valeurs.

Le problème posé dans la pratique est celui de discriminer efficacement les différents objets de synthèse instanciés dans la base d'exemples afin de fabriquer un système expert d'identification pour le Genre *Hyalonema*. Nous employons pour cela une méthode inductive décrite dans le chapitre 7. Les observations nouvelles serviront à consulter le système expert et eventuellement à constituer de nouveaux cas.

5.7 Conclusion

Le schéma de la figure 5.4 synthétise ce qui vient d'être dit dans ce chapitre. Il montre les relations existant entre les différents espaces qui permettent la description d'un individu. Les parties grisées symbolisent les espaces observables ∏, Λ et O (ce qui est possible d'observer pour chaque individu à décrire) alors que les parties vides illustrent les ensembles observés Ω, Y et V qui composent les descriptions dans la base de cas :

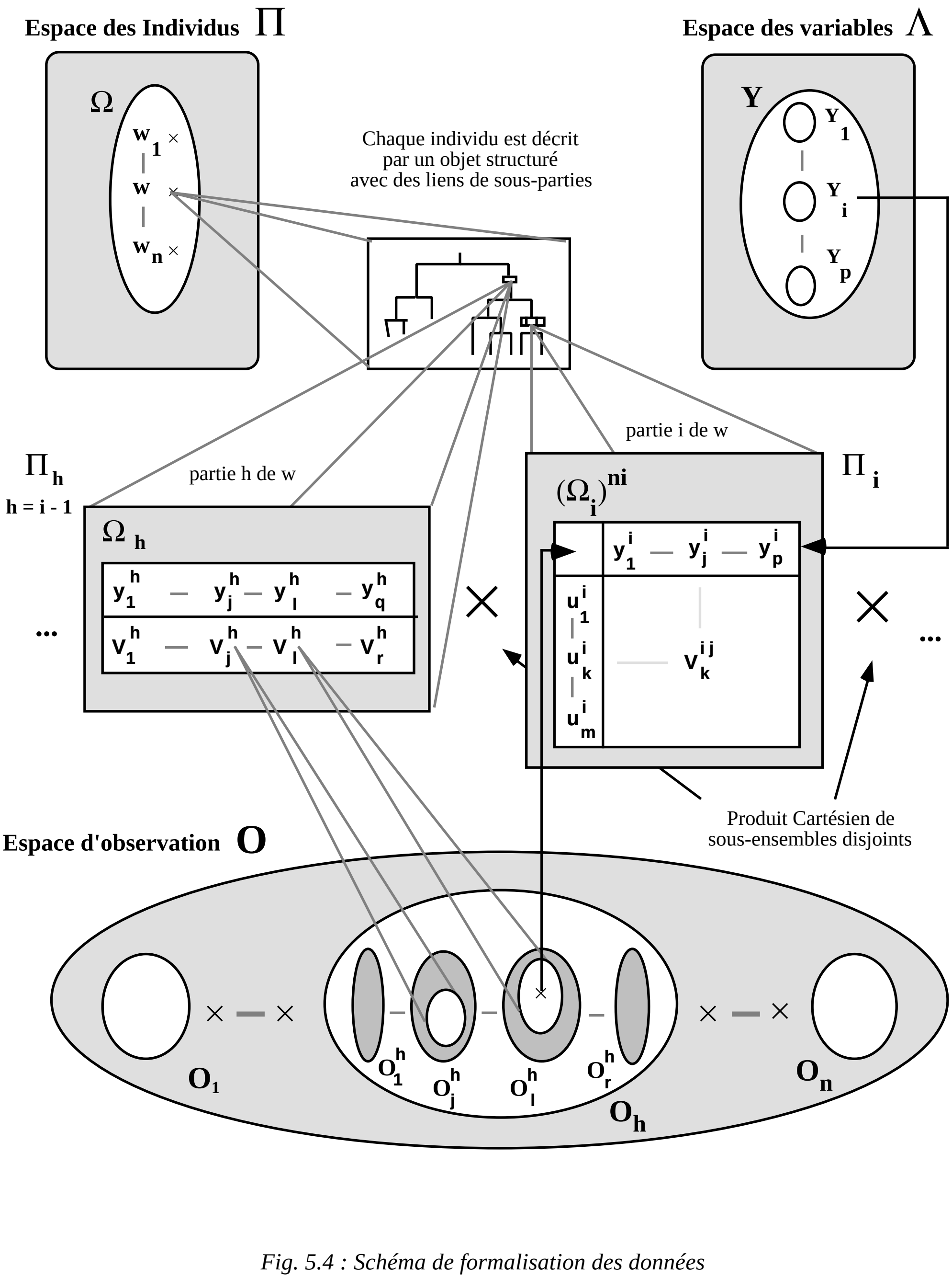

Fig. 5.4 : Schéma de formalisation des données

En reprenant la description d'un individu, on met en évidence qu'elle est structurée selon la relation de composition entre les différents objets. Chaque objet représente une partie de la descrition sans qu'il y ait de recouvrement entre eux (la description est formée du produit cartésien de sous-ensembles disjoints). Cette présentation permet de bien faire la différence entre les objets qui sont décrits une seule fois et ceux qui sont décrits plusieurs fois au sein d'une même description. Un zoom sur chacun de ces objets permet d'apprécier leur description locale. Les premiers sont des objets assertions composites symbolisés sur la figure à gauche par un vecteur de valeurs d'attributs (les variables étant indiquées dans la première ligne). Les seconds sont des objets hordes composites décrits par le tableau de droite avec chaque instance de la horde occupant une ligne du tableau. Pour chaque description locale d'un objet, le schéma montre la différence entre les variables relationnelles et les variables terminales. Les premières ont un espace d'observation dont les valeurs sont d'autres objets observés contrairement aux seconds.

Le schéma précédent n'indique pas néanmoins la description des objets munis de méthodes et propriétés ainsi que les objets classifiés (cf. § 5.4 et 5.5). Ces caractéristiques sont des raffinements de la description d'un individu qui permettent de préciser les conditions d'applicabilité des variables (objets et attributs) en fonction d'autres variables, et de spécifier les variables elles mêmes en fonction du contexte (la spécialisation). Comme ces caractéristiques dépendent du domaine considéré, elles n'ont pas été introduites dans la figure pour ne pas la compliquer d'avantage. Néanmoins, ces connaissances supplémentaires doivent être explicitées dans la représentation des connaissances car elles expriment souvent le “bon sens” qu'il faut donner aux descriptions.